Numéro National de Thèse : 2017LYSEN009

**THÈSE de DOCTORAT DE L'UNIVERSITÉ DE LYON**

opérée par

**l'École Normale Supérieure de Lyon**

**École Doctorale** N° 340

**Biologie Moléculaire et Intégrative de la Cellule (BMIC)**

**Discipline** : Sciences de la vie

Soutenue publiquement le 02/02/2017 par :

**Coraline PETIT**

---

# Évolution et Développement d'un organe sériel : la molaire.

# Transcriptomique comparée des bourgeons de molaires chez les rongeurs

---

Devant le jury composé de :

| | | |
|---|---|---|
| Mélanie Debiais-Thibaud | Université de Montpellier | rapporteur |
| Hugues Roest Crollius | ENS Paris | rapporteur |
| Michael Schubert | Villefranche-sur-Mer | examinateur |
| Ollivier Gandrillon | ENS de Lyon | examinateur |
| Marie Sémon | ENS de Lyon | directeur de thèse |
| Sophie Pantalacci | ENS de Lyon | co-encadrante de thèse |

# Remerciements

Tout d'abord je tiens à remercier mes directrices de thèse **Marie Sémon** et **Sophie Pantalacci**. Merci d'avoir permis à cette aventure d'avoir lieu. Tout a commencé pendant mon stage de master 1, où vous m'avez accueillie (au LR6 d'ailleurs, vous êtes maintenant revenues mais cette fois avec votre propre équipe!) et pendant lequel vous m'avez transmis la passion des dents et de la bioinformatique! Merci d'avoir prolongé cette aventure avec un second stage en master 2, suivi de ces trois années de thèse. Je sais que ça n'a pas été toujours facile, notamment parce que je ne suis pas très communicative et que je suis un petit lapin dans les phares. Malgré les difficultés que l'on a pu rencontrer, je suis assez fier du travail que j'ai accompli en votre compagnie.

Je remercie mes rapporteurs, **Mélanie Debiais-Thibaud** et **Hugues Roest-Crollius** pour leurs conseils et pour avoir accepté d'évaluer ces travaux.
Je tiens aussi à remercier les membres de mon jury : **Mélanie Debiais-Thibaud**, **Hugues Roest-Crollius**, **Olivier Gandrillon** et **Michael Schubert** qui ont accepté de juger ces travaux ainsi que d'être présents le jour de la soutenance.

Merci à **Daniel Khan**, l'organisateur de l'école de printemps à Aussois CompSysBio 2015, où j'ai pu découvrir le domaine de la biologie des systèmes tout en allant skier pendant nos pauses de l'après midi. Cette école m'a appris beaucoup et m'a permis de faire de belles rencontres.

**Patrick**, nous nous sommes rencontrés à Aussois et notre rencontre m'a beaucoup apportée. Merci, de m'avoir écoutée et d'avoir essayé de me donner confiance en moi. Tu as répondu présent à chaque fois que je t'ai demandé conseils. J'ai beaucoup apprécié nos discussions, ta philosophie de la vie et ta façon d'aborder les choses. Tu m'as rappelé qu'il n'y avait pas de questions bêtes et tu m'as aussi transmis ta curiosité des choses. Une fois tout ça fini, je compte bien venir faire du côté de la Belgique.

Merci aux anciens membres de l'équipe de **Vincent Laudet** à l'IGFL : d'abord merci à **Vincent** lui même mais aussi à **Laure** la poissonnière et à **Anne** la gardienne de souris. Merci à **Alexa**, notamment pour ce super week-end de ski à Cham, ta patience pour nous (Guillaume et moi) avoir appris ou ré-appris à skier. Ces paysages sont à jamais gravés dans ma mémoire. Merci de nous avoir fait rencontré Dr House! Tu es celle qui m'a redonné le goût au ski! Quand j'ai commencé ma thèse, tu finissais la tienne, tu étais et tu es toujours un modèle pour moi.

Merci à mon co-bureau à l'IGFL, **Guillaume** avec qui j'ai partagé pas mal de moments épiques. Merci d'avoir supporté ma folie! Merci pour les roulades de krav maga, les jeux de mots pourris, qui raisonnaient en écho dans le bureau et ta rigueur allemande ou alsacienne (tu peux choisir) qui était souvent mise à rude épreuve à cause des petit lutins dérangeurs.

Pour continuer avec l'équipe, je voudrais aussi remercier **Manon** et **Luke**, quand je pense à vous deux je pense à la chanson "narwhals" que Luke a eu la gentillesse de nous apprendre afin que nous puissions scander à tue tête les paroles exactes. **Manon**, je n'oublierai pas nos grands moments où nous nous exprimions en chansons, et où nous tapions un peu sur le système de **Guillaume**, je crois même qu'une fois nous avions réussi à le faire chanter ! **Luke**, ou devrais-je dire wikiluke, tu es venu te joindre à nous dans la course au jeux de mots pourris, je te félicite pour en trouver de si bon en français, qui n'est pourtant pas ta langue d'origine mais que tu maitrises admirablement. Merci aussi pour ton rire qui fait qu'on ne passe pas inaperçu quand on te raconte une blague, même si c'est un peu l'hôpital qui se moque de la charité par ce que je n'ai pas un rire très discret. Je tiens d'ailleurs à m'excuser au première étage de l'igfl, pour les crises de rire que l'on prenait, parce que toi et moi en écho nous n'étions pas très discrets ! Enfin je souhaite bon courage à **Carine** pour sa thèse, je sais que tu vas tout déchirer ! Aucun read RNA-seq et aucune séquence ne pourront te résister. *PS :* j'ai la preuve photographique que tu vas quand même dans le labo des fois.

Merci à **Alain**, tu étais mon bisousnours ! Nos pauses café/chocolat chaud et nos moments de craquage dans le bureau sont inoubliables, reste un pirate, mais fait attention à ton foie, et la prochaine fois tu me diras "Á lundi !" (désolée elle était trop tentante celle là).

Je voudrais encore remercier quelques personne du premier étage. Tic et Tac, **Hugo** notre hippie refoulé préféré et **Loan** notre golden retriever récemment reconverti en magnifique licorne, qu'en pense papa Gilles ? Enfin, il y a **Mélanie** et **Amélie**, les filles vous êtes un peu dingues quand même, merci. Vous étiez mes fournisseurs officiels de flocons de neige pour mes moments "let it go...". **Amélie**, ma petite méduse dans le couloir, **Mélanie** ma sœur licorne, il y en tellement je ne peux pas faire honneur à tous. Merci d'avoir veillé sur ma folie. Que de bons moments le soir au labo, ou autour d'une bière au Ninka, ou encore au ski ! Merci au padawan d'Amélie, **Jonathan** notre miss tupperware, pour m'avoir trouver un magnifique fond d'écran homme licorne ! Et **Amélie**, je n'ai qu'une chose à te dire : "Et moi pendant ce temps là, je tournais la manivelle et moi pendant ce temps là, je chantais dan les bois..."

Pour le soutien informatique, et aussi pour les oreilles qui saignent, je voudrais remercier le trio magique : **Loïs**, **Emmanuel** (Manu quoi) et **Thomas**. Les gars, merci pour votre patience, pour avoir pris le temps de discuter informatique et de m'expliquer les concepts de bases. Je vous dois aussi le soutien logistique qui m'ont permis de réaliser mes analyses. Vous avez su réveiller la geek qui est en moi. Ah tiens et pour la route un petit capitaine Flam, en l'honneur notre Galaxy chéri. La ligne de commande il n'y a que ça de vrai. Linux Power !

Je remercie mon ami **Thomas** pour son soutien psychologique, sans lequel j'aurai été découragée bien des fois.... Je te dédie à hauteur de 5% ce manuscrit (si si tu vois je l'ai écrit), quoique 5% ce n'est pas encore assez pour tout le soutien que tu m'as apporté. Tu as toujours été là dans les moments de doutes, aussi bien professionnels que personnels. Tu as toujours une oreille très attentive et de précieux conseils. Merci aussi à toi pour m'avoir transmis le virus de la moto, ce qui nous a permis par la suite de faire pas mal de belles balades et de bien sillonner les routes. Merci aussi pour tes précieux conseils de pilotages, sans lesquels je ne prendrai pas autant de plaisir sur mon destrier orange !

Je remercie **Fouad** pour m'avoir enseigné les rudiments du maniement de ce fameux destrier motorisé. Oui il n'y a pas d'or par terre, il faut arrêter de chercher et il n'y a pas de feuilles au bout des tiges de bambous.., et enfin tout est une question de volonté quand on y croit on peut. Merci, cette nouvelle

passion me permet de garder la tête sur les épaules.

Ce qui m'amène à ma deuxième passion, l'équitation. Entre les chevaux et moi c'est une longue histoire, puisque qu'elle a commencé quand j'avais 8 ans. Merci **Maman** de m'avoir permis de réaliser cette passion et de m'avoir accompagné dans les boxes au début quand j'avais peur ! Cependant, j'ai (re)découvert le plaisir de monter à cheval il y a deux ans lors du salon EquitaLyon, en découvrant la monte Western. Je tiens d'ailleurs à remercier **Antoine** et **Josianne** (et leurs chevaux) pour leur cours super, leur pédagogie et leur patience, grâce à vous, malgré mes 10 ans d'équitation, j'ai appris à monter à cheval, à comprendre et à sentir cet animal fascinant. Le cheval est ma seconde bouffée d'oxygène.

Merci à mon **Papa**, nous ne sommes pas de grands bavards mais tu as su être là pour moi. Je crois que tu es le premier à avoir susciter mon intérêt pour l'informatique. Merci à **Martine**, nos discussions m'ont toujours enthousiasmées, et je te dois ma passion pour les oiseaux !

Merci aussi à **Jérémy**, nos vies ont pris des chemins différents, mais tu as su me soutenir quand j'en avais besoin. Tu m'as transmis ta passion de la randonnée et m'as rendu accro à la montagne. Nous avons partager la passion des oiseaux et tu m'as aussi transmis ta passion pour la photo et les fleurs.
Merci aux potes aussi pour les virées à moto et les soirées, entre autre à **Mathieu**, **Anaïs** et **Antho**, merci pour les conseils d'entretien moto et pour les itinéraires sinueux. L'année prochaine je viens au Bol avec vous, ou alors cette fois on monte au Mans ensemble. Laurent, Eric, et bien sur mon papa d'adoption, **Alexandre**, merci pour ce beau périple et ce super week-end au Mans (600km en une journée ça reste mon record !) sans vous je n'aurais jamais découvert le doux bruit des moteurs au rupteur qui nous bercent la nuit ainsi que la douce odeur de la gomme pendant les burns !

Merci à la famille Alonso et Guillerd et aux amis proches de la famille. **Caro**, **Didier** et **Mari-Jo** merci pour votre accueil et votre bonne humeur. **Marie-Hélène**, la jumelle country de ma Maman et **Franck**, le papa ours sur son oural, merci pour tes conseils sur ma conduite. Merci les Guillerd (**Fabrice**, **Sylviane**, **Léa**, **Flora**, **Pépé** et **Mamie** didile), vous avez supportez mes discours sur les yeux des araignées et essayer de comprendre ma théorie sur l'hérédité de la calvitie (pourquoi les gars de notre famille sont chauves quoi). **Fabrice**, parce qu'elle ressort à chaque repas de famille je tiens à l'écrire noir sur blanc dans les remerciements de cette thèse : "Mamie, parrain il a jeté la papillote dans le feu !" (un noël de l'année 1995 ou 96).

Merci à mon frère **Thibault** qui lui aussi me supporte depuis toutes ces années. Je suis fière de toi et de tout ce que tu fais. Merci de nous faire partager tes talents de guitariste à la maison et bientôt nous pourrons aller sillonner les routes à moto ensemble ! Il ne te manque plus que la moto maintenant.

Enfin merci à mes parents. Je n'ai pas toujours été très communicative ni très reconnaissante mais vous avez et vous êtes toujours là pour moi. Même si je ne le dis surement pas assez votre soutien est précieux ! Merci **Maman** et **Alexandre**.

*Lyon, 24 janvier 2017*

Dans mon monde à moi...

Coraline

# Abstract

Developmental programs are the result of the coordinate expression of thousands of genes which is nowadays accessible through RNA-sequencing. Which differences in expression levels underlie differences in developmental programs, (i) between similar organs of the same species? (ii) between homologous organs in different species? While others studies have been focusing on master regulatory genes, I concentrated on transcriptome as a whole to untangle the link between differences in developmental programs and differences in final morphologies. Our model of study is the development of rodent upper and lower first molars, for which we obtained transcriptomes for either whole molar germs at different stages of development or molar germs cut in half. The main variation in these data is a temporal signal which is present in whole organs, as in biological replicates and also in germ halves. The second pattern of variation is the difference between upper and lower molars. We evidenced that transcriptome reflects the relative tissue proportions composing the organ, thus it informed us on differences in cellular proportion between each molars. Thus mouse first molars differ in their relative proportion in mesenchyme and cusp tissue. Then, we showed that specificities of the upper/lower molar developmental programs are conserved between mouse and hamster. However, transcriptomic differences between species are not correlated to the morphological differences, even when the final morphology is similar (eg. the lower molar). This rapid evolution of expression profiles between species is consistent with a phenomenon known as Developmental System Drift (DSD). At last, I was interested in the identity of the buccal and lingual side (BL) of mouse molars, because the supplementary cusps of the mouse first upper molar are formed lingually. We evidenced that the lingual side has an identity of its own and that the differences of expression between buccal and lingual side are increased in mouse upper molar. Finally, the prospects for this work will be to understand the changes of the developmental program that differentiate mouse upper molar from hamster ones, relying more these BL data, to understand how the mouse first upper molar developed formed two supplementary lingual cusps.

Key words : Comparative transcriptomic; transcriptomics signature; temporal dynamics of gene expression, Developmental Dystem Drift (DSD); Evolution of gene expression; Molar development

# Résumé

Les programmes de développement font appel à l'expression coordonnée de milliers de gènes, à laquelle le RNA-seq nous donne maintenant accès. Quelles différences d'expression sous-tendent les différences de programmes de développement, (i) entre organes similaires d'une même espèce ? (ii) entres organes homologues dans des espèces différentes ? Alors que d'autres études s'intéressent à des gènes maître régulateurs, je me suis intéressée au transcriptome pris dans son ensemble, dans ce qu'il peut nous apporter pour comprendre les différences de programme de développement en relation avec la morphologie finale. Notre modèle est le développement des molaires inférieure et supérieure chez les rongeurs, pour lesquelles nous avons obtenus des séries transcriptomiques de germes dentaires complets ou de germe coupés en deux. La majeure partie de la variation dans ces données correspond à un signal temporel qui est observé pour organes complets au cours du développement, mais aussi plus finement entre réplicats et enfin au niveau de dents coupées en deux. Le deuxième patron de variation est la différence entre la molaire inférieure et supérieure. Nous avons mis en évidence que le transcriptome reflète les proportions relatives de tissus qui le composent et nous renseigne sur des différences de population de cellules constituant chaque molaire. Ainsi les molaires de souris diffèrent entre elles par leur composition relative en mésenchyme et en tissu de cuspides. Puis nous avons montré que les spécificités des programmes inf/sup sont conservées chez la souris et le hamster. Cependant, les différences transcriptomiques entre ces deux espèces ne corrèlent pas avec les différences morphologiques, même dans le cas où la morphologie est similaire. Cette évolution rapide des profils temporels d'expression est compatible avec un phénomène de dérive développementale. Enfin, je me suis intéressée à l'identité bucco-lingual (BL) des molaires, car les cuspides supplémentaires de la molaire sup de souris se forment côté lingual. Nous avons montré que le coté lingual a sa propre identité et que les différences d'expression BL sont plus fortes dans la molaire sup de souris. Enfin, les perspectives de ces travaux seront de comprendre les modifications du programme développemental qui différencient les molaires sup de souris et des hamster, en s'appuyant en plus sur les données BL, afin de comprendre comment la molaire sup de souris développe 2 cuspides linguales supplémentaires.

Mots clefs : Transcriptomique comparée ; signature transcriptomique ; dynamique temporelle de l'expression, évolution du niveau d'expression, dérive du système développemental ; développement des molaires

# Table des matières

**Remerciements** **i**

**Table des matières** **viii**

**Liste des figures** **xiii**

**I Introduction** **1**

**1 Evo-devo : comprendre la diversité du monde vivant** **5**

1.1 Histoire de l'évo-dévo . . . . . 5

1.1.1 Les débuts de la discipline au $XIX^{ème}$ siècle . . . . . 5

1.1.2 La ré-unification de la biologie du développement et de la biologie évolutive 6

1.1.2.1 Évolutionnistes . . . . . 6

1.1.2.2 Développementalistes . . . . . 7

1.1.2.3 La découverte des gènes Hox . . . . . 7

1.1.2.4 Les questions de l'évo-dévo . . . . . 8

1.1.3 La notion d'homologie, un concept central de l'évo-dévo . . . . . 9

1.1.3.1 La deep homology, exemple de l'évolution des yeux . . . . . 10

1.1.3.2 L'homologie sérielle, exemple de l'évolution des membres . . . . . 10

1.2 Évolution génétique du développement avec évolution phénotypique . . . . . 11

1.2.1 Évolution de la boite à outils génétique . . . . . 11

1.2.2 Pléiotropie et mutations dans les régions *cis*-régulatrices . . . . . 12

1.2.3 Réseaux génétiques et co-option . . . . . 12

1.2.4 La modularité du développement . . . . . 14

1.2.5 Le paysage épigénétique de Waddington, un modèle pour penser le développement et l'évolution génétique . . . . . 14

1.3 Evolution génétique du développement sans évolution phénotypique . . . . . 15

1.3.1 Évolution cryptique . . . . . 15

1.3.2 La dérive développementale . . . . . 15

1.4 Patrons généraux dans l'évolution du développement . . . . . 18

1.4.1 Contraintes versus flexibilité dans l'évolution du développement . . . . . 18
1.4.2 Les additions terminales, plus grande flexibilité du développement? . . . . 18
1.4.2.1 Les modèles sablier vs. entonnoir . . . . . 18
1.4.3 L'hétérochronie . . . . . 20

**2 L'outil transcriptomique : comprendre l'évolution de l'expression et l'évolution du développement 21**
2.1 Évolution de l'expression avant l'ère génomique . . . . . 21
2.2 Évolution de l'expression des gènes à l'ère génomique . . . . . 22
2.2.1 Les données transcriptomiques à large échelle . . . . . 22
2.2.1.1 Quels sont les modes d'évolution de l'expression? . . . . . 23
2.2.1.2 Le niveau d'expression est-il sous sélection? Les différents approches pour répondre à cette question . . . . . 24
2.2.1.2.1 Données intraspécifiques. . . . . 24
2.2.1.2.2 Données pluri-spécifiques. . . . . 24
2.2.2 Les différents patrons d'évolutions de l'expression . . . . . 27
2.2.2.1 L'effet espèce est-il plus fort que l'effet organe? . . . . . 27
2.2.2.2 Les différentes vitesses d'évolution . . . . . 29
2.3 La transcriptome apporte un nouvel éclairage aux questions de l'évo-dévo . . . . . 29
2.3.1 L'identité entre organes . . . . . 29
2.3.2 Quand l'expression est-elle contrainte au cours du développement? . . . . 30

**3 Le modèle dent 35**
3.1 Les dents et leur évolution . . . . . 35
3.1.1 Qu'est-ce qu'une dent? . . . . . 35
3.1.2 Les dents sont une innovation morphologique des vertébrés à mâchoires . 36
3.1.2.1 La morphologie des dents . . . . . 37
3.1.3 L'évolution des molaires . . . . . 38
3.1.3.1 La molaire tribosphénique . . . . . 39
3.1.3.2 Le plan murin . . . . . 39
3.2 Le développement des dents . . . . . 40
3.2.1 Les dents sont des organes sériels dont l'identité est définie par un code homéotique . . . . . 40
3.2.1.1 Les dents sont des organes sériels . . . . . 40
3.2.1.2 Identité des dents et code homéotique . . . . . 41
3.2.2 Généralités sur le développement de la molaire de souris . . . . . 41
3.2.3 Le nœud d'émail : centre de signalisation responsable de la formation des cuspides . . . . . 42
3.2.4 Un modèle *in silico* du développement de la dent . . . . . 44
3.3 Les voies de signalisation majeures du développement de la molaire . . . . . 45

3.3.1 La voie de signalisation Hedgehog est responsable de la polarisation du germe dans l'axe bucco-lingual . . . 45
3.3.2 La voie de signalisation des BMPs, une voie centrale pour la transition au stade coiffe . . . 46
3.3.3 La voie de signalisation Wnt régule le patterning du germe . . . 47
3.3.4 La voie de signalisation des FGFs régule la croissance de l'épithélium . . . 49
3.4 Comment augmenter le nombre de cuspides? . . . 52
3.4.1 Spécificités morphologiques des premières molaires de souris . . . 52
3.4.2 Modulation de la formation des cuspides au cours du développement de la première molaire de souris . . . 52
3.5 L'asymétrie bucco-linguale au cours du développement des dents . . . 54
3.5.1 Le développement des dents de remplacement est lingual . . . 54
3.5.2 L'asymétrie bucco-linguale dans le développement des dents de souris . . . 54
3.5.2.1 Asymétrie bucco-linguale du développement de la première molaire de souris . . . 55
3.6 Études transcriptomiques du développement des dents de souris . . . 58

**II Matériels et Méthodes 61**

**4 Analyse des données RNA-seq 63**
4.1 Présentation des jeux de données . . . 63
4.2 Extraction d'ARN . . . 64
4.3 Estimation du niveau d'expression à partir de données RNA-seq . . . 64
4.3.1 Comparaison des niveaux d'expression entre espèces . . . 67
4.3.1.1 Alignement des lectures sur un génome de référence . . . 67
4.3.1.2 Assemblage *de novo* de transcriptomes avec *Trinity* . . . 67
4.3.1.3 Comparaison des deux méthodes et stratégie adoptée . . . 68
4.4 Analyses multivariées . . . 68
4.4.1 Les différentes analyses multivariées . . . 68
4.4.2 Identifier les axes majeurs de variations et caractériser la participation de chaque tissu à cette variation . . . 70
4.4.3 Analyses *between* : identifier la variation associée à un facteur . . . 70
4.4.4 Distance BL sur PCA et Bootstrapping . . . 71
4.4.4.1 La distance BL est-elle la même entre la molaire inférieure et la molaire supérieure? . . . 71
4.4.4.2 L'excès de différence BL est-il dû à une différence de progression du programme développemental? . . . 71
4.5 Analyse des gènes différentiellement exprimés (DE) . . . 72
4.5.1 Estimation de "l'âge" des échantillons BL . . . 72

4.5.2 Estimation du stade de développement des échantillons du jeu de données petite échelle . . . 73
4.6 Corrélations entre jeux de données grande et petite échelle temporelle de souris . 74
4.7 Exclure les artéfacts statistiques des données bucco-linguales . . . 75
4.7.1 Ré-échantillonnage des librairies bucco-linguales . . . 75
4.7.2 Dispersion des librairies RNA-seq . . . 75
4.8 Analyse GO de listes de gènes . . . 76
4.9 Estimation des proportions de tissus par déconvolution . . . 77
4.9.1 Le principe . . . 77
4.9.2 Définition des marqueurs . . . 78
4.9.2.1 Remarques sur la méthode de déconvolution . . . 79
4.9.2.2 Ré-échantillonnage des marqueurs par jacknife . . . 79
4.10 Définir un seuil d'émaillitude d'un gènes buccal et identifier des nouveaux marqueurs potentiels du nœud d'émail . . . 79

**5 Microscopie 3D** **83**

**6 *Bmper* : gène candidat de l'asymétrie bucco-linguale** **85**
6.1 Hybridation *in situ* : caractérisation du patron d'expression de *Bmper* . . . 85
6.2 Mutants *bmper* et coupes histologiques . . . 85

**III Résultats** **87**

**7 Transcriptomic signature of serial organs : the mouse first upper and lower molar** **89**

**8 Temporal signal in transcriptome of developing organs** **93**

**9 Comparative transcriptomics of rodent molars** **109**

**10 Characterizing bucco-lingual asymmetry during mouse molar development** **145**

**11 *Bmper* : a candidate for the bucco-lingual asymmetry of tooth development** **173**

**IV Discussion** **181**

**12 Discussion** **183**
12.1 Conclusion . . . 183
12.2 Approche transcriptomique comparée de l'étude des programmes de développement 184
12.2.1 Les difficultés de la comparaison des niveaux d'expression entre espèces . . 184
12.2.2 Caractérisation des signatures transcriptomique au cours du développement des molaires inférieures et supérieures . . . 185
12.2.2.1 Le signal temporel est une signature transcriptomique majoritaire 185

12.2.2.2 Utilisation du signal temporel pour rendre homologues des stades de développement entre différentes espèces et comparer leurs profils d'expression . . . 186
12.2.2.3 L'identité de l'organe est la seconde signature transcriptomique . . 188
12.2.2.3.1 Signature transcriptomique de l'identité de l'organe au niveau de l'organe entier . . . 188
12.2.2.3.2 Signature transcriptomique de l'identité de l'organe au niveau des tissus composants l'organe . . . 188
12.2.2.3.3 Signature transcriptomique de l'identité de l'organe au niveau des cellules même de l'organe . . . 189
12.2.3 Application de l'approche transcriptomique à l'étude du développement bucco-lingual . . . 190
12.2.3.1 Découverte d'une forte identité buccale et linguale . . . 191
12.3 Évolution du développement . . . 192
12.3.1 Conservation d'une différence entre programmes développementaux de molaires inférieures et supérieures . . . 193
12.3.2 Une dynamique temporelle commune entre le hamster et la souris . . . 193
12.3.2.1 Conservation d'un excès de différences entre les programmes développementaux des molaires supérieures et inférieures . . . 193
12.3.3 L'évolution du programme développemental des molaires n'est pas corrélée avec leur évolution morphologique . . . 194
12.3.3.1 L'identité entre espèces est plus forte que l'identité entre molaires 194
12.3.3.2 Dérive du programme développemental . . . 195
12.3.4 Perspectives sur l'origine des molaires tribosphéniques . . . 196
12.3.5 Perspectives sur l'origine du plan murin : comment la souris met en place deux cuspides linguales supplémentaires? . . . 198
12.3.5.1 Les spécificités de la molaire supérieure de souris . . . 198
12.3.5.2 La différence bucco-linguale est plus forte dans la molaire supérieure de souris . . . 199
12.3.5.3 Combinaison des analyses bucco-linguale chez la souris et inférieure-supérieure chez la souris et le hamster . . . 200
12.3.5.4 *Bmper*, un gène candidat du développement des cuspides linguales supplémentaires chez la souris . . . 201
12.3.6 Comment se traduit la convergence évolutive dans les transcriptomes? . . . 202

**V Appendix** **205**

**A Transcriptomic signature of serial organs : the mouse first upper and lower molar** **207**

**Bibliography** **265**

# Table des figures


1.1 Planche des embryons observés par Haeckel . . . . . 6
1.2 Questions de l'évo-dévo . . . . . 9
1.3 Évolution de l'oeil . . . . . 11
1.4 Réseaux génétiques et évolution de la dentition des gnathostomes . . . . . 13
1.5 Paysage épigénétique de Waddington . . . . . 16
1.6 Chemins développemenaux et morphologies . . . . . 17
1.7 Les contraintes au cours du développement . . . . . 19

2.1 Données transcriptomiques à large échelle . . . . . 22
2.2 Patrons d'expression attendus en fonction des forces évolutives en jeu . . . . . 23
2.3 Principe de la PCA . . . . . 26
2.4 PCA de transcriptomes d'organes adultes chez différents mammifères . . . . . 28
2.5 Identité des doigts de poulet . . . . . 32
2.6 Contrainte type "hourgalass" chez les embryons de drosophile . . . . . 33

3.1 Morphologie de la dent . . . . . 36
3.2 Diversité des dents . . . . . 37
3.3 Arbre phylogénétique des gnathostomes et apparition des dents . . . . . 38
3.4 Évolution morphologique des molaires supériures et inférieures . . . . . 40
3.5 Développment des molaires . . . . . 43
3.6 La transition bourgeon-coiffe . . . . . 48
3.7 La morphogenèse de la molaire . . . . . 51
3.8 Développement des dents de remplacement . . . . . 55
3.9 Asymétrie bucco-liguale du développmet de la première molaire de souris . . . . . 57

4.1 Comptes Kallisto vs. HTSeq . . . . . 65
4.2 Principe de l'analyse multivariée *between-group* . . . . . 71
4.3 Principe de l'estimation de l'âge des échantillons bucco-lingual . . . . . 73
4.4 Visualisation des résultas de l'analyse GO . . . . . 77
4.5 Estimer des proportions tissulaires . . . . . 81

5.1 Principe de la reconstruction 3D de la forme de la molaire . . . . . 84

8.1 Temporal ordering of timeseries transcriptomes . . . . . 94

9.1 How to estimate comparable expression levels between species? . . . . . 111
9.2 Schematic representation of variations between mouse and hamster molars at the morphological and transcriptomic level . . . . . 112
9.3 A shared peak of upper-lower differences . . . . . 143
9.4 A shared increased in genes differentially expressed at *mid*-cuspidogenesis . . . . 144

11.1 Enrichment of *Bmper* in upper molars of mouse and hamster . . . . . 175
11.2 *Bmper* expression profile in mouse and hamster . . . . . 176
11.3 *Bmper* expression in mouse molars at ED15.5 . . . . . 178
11.4 Molar morphology of *bmper* homozygous mutant mice at staged ED18.5 . . . . . 180

12.1 Comment les différences de niveau d'expression se traduisent-elles dans un organe?186
12.2 Différents mécanismes de croissance entre les côtés buccaux et linguaux . . . . . 192
12.3 Identité des boucles épithéliales de la molaire de souris au cours du développement197
12.4 Convergence de l'acquisition des cuspides linguales supplémentaires chez la souris et la souris épineuse . . . . . 204

# Liste des tableaux

3.1 Gènes biaisés bucco-lingualement au cours du développement de la première molaire de souris. . . . . 58

4.1 Données réplicats biologiques RNA-seq de bourgeons dentaires de souris au stade ED15.5. *(M= millions de lectures)* . . . . . 66
4.2 Librairies bucco-linguales . . . . . 66
4.3 Librairies tissus purs. . . . . 66

# Introduction Part I

# À propos

Ce travail de thèse a pour objectif de caractériser, à l'échelle global du transcriptome, les signatures transcriptomiques associées à un organe sériel : les premières molaires inférieures et supérieures de souris, qui de plus présentent des morphologies différentes. Cette caractérisation s'inscrit dans une démarche évo-dévo, elle a pour but de comprendre comment de nouvelles formes, contraintes par le programme développemental, peuvent se mettre en place au sein d'une espèce. De plus, l'analyse comparée du programme développemental de la première molaire chez la souris et le hamster va nous permettre d'interroger l'évolution du programme développemental. Est-ce qu'une conservation de la morphologie finale est couplée avec une conservation du programme de développement ? Dans cette introduction, je vais dans un premier temps présenter le champ disciplinaire de l'évo-dévo et définir des concepts qui seront utiles pour comprendre et mettre en perspective les résultats de ma thèse. Dans un deuxième chapitre, je parlerai d'abord de l'évolution du niveau d'expression dans les organes adultes de différentes espèces. Ensuite, j'aborderai la transcriptomique comme outil apportant un éclairage nouveau sur des questions anciennes de l'évo-dévo. Enfin dans le troisième et dernier chapitre de cette introduction je présenterai mon modèle d'étude : la dent. En abordant l'évolution des dents (apparition et morphologie), en insistant sur l'évolution morphologique des molaires ainsi que ce qui est connu sur leur développement.

# 1 Evo-devo : comprendre la diversité du monde vivant

"Nothing in evolution makes sense except in the light of processes, starting with genomes, that affect developmental operations and ultimately spread through a population." [Dover, 1992]

## 1.1 Histoire de l'évo-dévo

### 1.1.1 Les débuts de la discipline au $XIX^{ème}$ siècle

Les origines de l'évo-dévo se trouvent au début du $XIX^{ème}$ siècle à l'âge d'or de l'embryologie comparative. Il est assez vite apparu un besoin de comprendre et d'expliquer les similarités morphologiques des espèces, que ce soit au cours du développement ou au stade adulte. Les travaux de Johann Meckel et d'Étienne Serres au début des années 1800, sont à l'origine de la loi Meckel-Serres ou théorie du parallélisme qui postule que lors de l'embryogenèse, un animal "supérieur" passe par toutes les étapes caractéristiques des organismes "inférieurs" moins complexes. Louis Agassiz établit un parallèle entre le développement de l'embryon, les classifications et les séries fossiles qui vont tous les trois dans les sens d'une complexité croissante. En 1866, Ernst Haeckel reprend la théorie du parallélisme en l'associant avec la théorie de l'évolution proposée par Charles Darwin [1859] et Alfred Russel Wallace, donnant naissance à la théorie de la récapitulation aussi connue sous le nom de loi biogénétique. Plus connue sous l'adage :"l'ontogénie récapitule la phylogénie", la loi biogénétique de Haeckel postule que chaque organisme se développe en passant par les étapes des organismes ancestraux [Richardson *et al.* , 1997; Richardson & Keuck, 2002]. Les fameux dessins des embryons sur lesquels Haeckel appuie sa loi biogénétique sont représentés figure 1.1. La théorie de la récapitulation de Haeckel est basée sur l'observation que les embryons des différents organismes de vertébrés partagent une organisation commune et ont souvent un aspect similaire (fig. 1.1), alors que les différences entre adultes d'espèces différentes deviennent d'autant plus visible que le processus de développement est avancé. C'est aussi en partant de ces observations que Karl Ernst von Baer propose ses "lois de

l'embryologie" en 1828. Les lois de von Baer sont basées sur les quatre postulats suivants : (i) les caractères généraux d'un groupe d'animaux apparaissent en premier, suivi des caractères plus spécifiques, (ii) les caractères spécifiques se forment à partir de caractères généraux, (iii) chaque embryon, au lieu de passer par une forme d'un organisme ancestral, a une forme qui s'écarte de plus en plus de celle des autres organismes, (iv) l'embryon d'un animal "supérieur" ne ressemble jamais à l'adulte d'une espèce "inférieure" mais seulement à son embryon (d'après une revue d' Abzhanov [2013]).

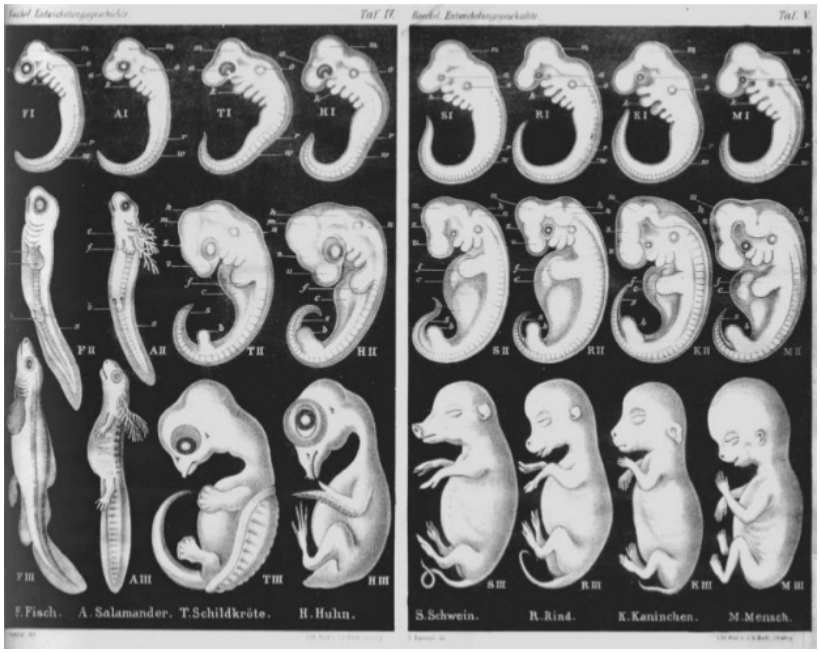



FIGURE 1.1 – **Loi biogénétique : l'ontogénie récapitule la phylogénie**.
Dessins des observations de Haeckel (1874) de l'embryogenèse comparée chez les vertébrés. La première ligne (stages précoces de l'embryogenèse) montre une période pendant laquelle, selon Haeckel, tous les embryons sont semblables. De gauche à droite : poisson, salamandre, tortue, poulet, cochon, vache, lapin, homme. (tiré de Richardson *et al.* [1997])

### 1.1.2 La ré-unification de la biologie du développement et de la biologie évolutive

#### 1.1.2.1 D'une part, les évolutionnistes

Les bases de la théorie de l'évolution sont posées par Darwin [1859] dans son livre *On the origin of species*. Il part de l'observation que les animaux domestiques tels les pigeons, présentent des variations de traits morphologiques, tel l'aspect des plumes, qui peuvent être sélectionnés par les éleveurs en faisant se reproduire entre eux les spécimens qui présentent ces caractères. Darwin en arrive à la conclusion que si l'Homme peut sélectionner spécifiquement des traits morphologiques alors ce processus à aussi lieu dans la nature : c'est ce qu'il nomme la sélection naturelle. Ici sont listés les principes généraux de la théorie de l'évolution de Darwin : (i) l'existence de

variations au sein d'une population, (ii) la compétition pour les ressources limitées entraine une "lutte" pour la survie, (iii) la reproduction des individus les mieux adaptés à la survie dans un environnement donne naissance à une descendance présentant les traits adaptatifs des parents, (iv) de la sorte il peut y avoir apparition de nouvelles espèces.

Thomas Hunt Morgan [1932] (contributions majeures à la génétique, on lui doit notamment la découverte du rôle des chromosomes dans l'hérédité et la découverte des liaisons génétiques) contraste l'approche génétique de l'évolution avec l'approche traditionnelle de l'étude de la morphologie et de l'anatomie comparative. En 1937, Theodosius Dobzhansky, élève de Morgan, redéfinit l'évolution comme étant des changements de fréquences alléliques. Ainsi l'étude de la dynamique des populations devient la discipline majeur de l'étude de l'évolution. La théorie synthétique de l'évolution proposée par Ernst Mayr, Theodosius Dobzhansky et Julian Huxley signe le découplage entre l'étude du développement et de l'évolution. Gilbert *et al.* [1996] analysent comment la génétique et la génétique des populations ont chassé l'embryologie dans l'étude de l'évolution.

#### 1.1.2.2 De l'autre part, les développementalistes

Dans les années 1900, le champ de recherche de l'embryologie comparée se transforme en devenant moins comparative et plus expérimentale et prend le nom de biologie du développement. Les biologistes du développement s'intéressent plus précisément aux processus développementaux tel la mise en place de l'axe antéro-postérieur. Par exemple, Spemann & Mangold découvrent l'organisateur qui porte le même nom et qui est responsable de la formation de l'axe antéro-postérieur, revue dans Gilbert *et al.* [1996]. Les embryologistes s'intéressent à l'organisation embryonnaire tel que la gastrulation, la neurulation, la mise en place de centres organisateurs, ainsi que la différenciation cellulaire. Une des questions de la biologie du développement est de savoir comment l'organe qui se développe est capable de contrôler l'arrangement spatial et l'identité des cellules qui le composent. Cette question de l'identité est d'autant plus intrigante que toutes les cellules descendent de la même cellule œuf. Les travaux de la biologie du développement couplés à une approche plus génétique que comparée, ont permis de mettre en évidence beaucoup de voies de signalisations tels que *BMP*, *Fgf*, *Shh* [Johnston, 2015]. Les modèles d'études sont majoritairement les embryons de grenouille, souris, mouche du vinaigre (drosophile) et le nématode *C. elegans*.

#### 1.1.2.3 La découverte des gènes Hox

L'émergence de la discipline "évo-dévo" se trouve dans les années 1980, avec la découverte des gènes à homéoboite et qui signe la réunification entre la biologie du développement et la biologie évolutive. Dans un premier temps, Scott & Weiner [1984] ont découvert que les gènes clés qui régulent le développement de la drosophile [Lewis, 1978; Lewis *et al.* , 1980] sont situés sur deux clusters qui ont une forte similarité de séquences et qui tous deux contiennent des gènes codant

pour des protéines à homéoboite. Puis, McGinnis *et al.* [1984] montrent que les séquences de ces deux clusters sont aussi conservées chez d'autres métazoaires. Afin d'expliquer la conservation de ces séquences, McGinnis *et al.* [1984] proposent donc que ces gènes du développement de la drosophile seraient eux aussi impliqués dans le développement des métazoaires et pourraient réguler des processus développementaux analogues. En effet, les relations homologues entre les clusters Hox de la drosophile et les clusters Hox des vertébrés (plus précisément la souris et l'Homme) ainsi que la conservation de la colinéarité d'expression au cours du développement (revue dans Krumlauf [1992]; Duboule & Morata [1994]) indique des mécanismes de développement partagés entre espèces. Cette découverte obligent les développementalistes à se confronter à des questions d'évolution, des origines des mécanismes de développement et comment ils ont varié/évolué entre espèces. Elle démontre aussi que les gènes jouent un rôle primordial dans le développement. Ainsi les évolutionnistes se trouvent confrontés à une nouvelle source de variation génétique : celle qui influence le cours du développement. Il devient donc nécessaire de penser un nouveau cadre conceptuel permettant l'étude de l'évolution des mécanismes développementaux afin de comprendre l'évolution morphologique des espèces. C'est ainsi que voit le jour l'évo-dévo [Raff, 2000; Wallace, 2002; Gilbert, 2003; Müller, 2007].

#### 1.1.2.4 Les questions de l'évo-dévo

D'un point de vue évolutif, la découverte des gènes Hox et leur conservation entre phyla, démontre l'existence de mécanismes généraux du développement qui sont conservés entre des organismes morphologiquement divergents. Se pose alors la question de l'émergence de la diversité entre espèces dont, au moins une partie du développement est conservée. Ainsi pour comprendre le processus évolutif il est important de comprendre les mécanismes du développement qui sont les processus qui façonnent l'adulte, donc ce qui sera "vue" par la sélection. Les questions de l'évo-dévo sont nombreuses et vastes et elles reflètent l'histoire de la discipline avec des questions soient d'ordre de développement comparé soient plutôt évolutives. Ainsi, avec une approche plus dévo, les questions posées tourneront plus autour de comment évolue le développement. Une approche comparée des processus développementaux pourra ainsi répondre à ces questions. Un éclairage plus évolutionniste, se demandera comment les processus d'évolution sont contraints par les processus développementaux. Les questions auxquelles ce travail de thèse tente d'apporter des réponses sont listées dans la box 1.2. Une liste plus détaillée de ces questionnements peut être trouvée dans les papiers suivants : Raff [2000]; Wallace [2002]; Gilbert [2003]; Müller [2007].

**BOX | Questions de l'évo-dévo**

- **Quelles sont les grandes règles de l'évolution du développement ?**
  - ➵ Le développement est-il contraint ?
    Si oui, quels moments du développement sont contraints ?
    Quels moments du développement sont plus flexibles ?
  - ➵ Changements du développement quand il y a innovation phénotypique :
    - ∗ Comment les processus développementaux peuvent-ils être modifiés pour générer des morphologies différentes ?
  - ➵ Changement du développement quand il y a conservation du phénotype sur de longues périodes :
    - ∗ Est-ce qu'une morphologie similaire dans deux espèces implique des programmes développementaux similaires ?
- **Comment le développement limite l'évolution ?**
  - ➵ Est-ce que tous les changements génétiques sont possibles ou certains changements sont-ils privilégiés ?
  - ➵ Comment la modularité du développement et les réseaux génétiques contraignent le développement et son évolution ?

FIGURE 1.2 – Box | Questions de l'évo-dévo.

### 1.1.3 La notion d'homologie, un concept central de l'évo-dévo

Afin de comparer des mécanismes développementaux entre espèces, il faut être capable d'identifier des traits et/ou des mécanismes qui partagent la même histoire évolutive. Ainsi l'étude de caractères homologues chez différentes espèces a pour but de reconstruire l'état ancestral de ce caractère. A partir de ce caractère ancestral, l'évolution dans chaque lignée peut alors être analysée. D'où la nécessité de définir des homologies. On doit l'une des premières définitions de l'homologie à Richard Owen en 1843. Il définit des homologues comme étant "le même organe dans différents animaux sous toutes ses déclinaisons de forme et de fonction. C'est encore aujourd'hui une définition assez utilisée pour ce qu'on appelle l'homologie structurale. L'exemple typique de caractères homologues sont les membres. Cependant, la définition du concept même d'homologie est très contestée [de Pinna, 1991; Hall, 2003; Young & HallgrÍmsson, 2005; Wagner, 2007]. En effet, cela vient certainement du fait que l'homologie peut être envisagée à différents niveaux. L'homologie structurale étant le niveau le plus "haut"et le plus visible. Je vais détailler ici, un peu plus spécifiquement deux types d'homologie qui sont importantes pour la mise en perspective de ce travail de thèse. Il s'agit de l'homologie profonde (deep homology) et de l'homologie sérielle (serial homology). Ces deux types d'homologie sont utiles pour comprendre

comment évolue le programme de développement des dents.

#### 1.1.3.1 La deep homology, exemple de l'évolution des yeux

La deep homology, que l'on peut traduire par homologie profonde, est un type d'homologie qui caractérise des réseaux génétiques. Pour mieux comprendre ce concept je propose de le développer à travers l'exemple de l'évolution des yeux. Les yeux des métazoaires ne sont pas des structures homologues, ils seraient apparus de manière indépendante au minimum 40 fois [Land & Fernald, 1992] (voir figure 1.3 pour un exemple de diversité morphologique des yeux). Cependant, Halder *et al.* [1995a] montrent que le gène *Pax-6* de souris induit la formation d'ommatidies chez *Drosophila melanogaster*. Ces résultats suggèrent que le rôle de *Pax-6* ait été conservé au cours de l'évolution des métazoaires. De là, comment réconcilier la non-homologie des structures et l'utilisation commune du gène "maître" *Pax-6*? Dans sa revue sur *Pax-6*, Halder *et al.* [1995b] proposent que *Pax-6* ait permis, dans l'ancêtre commun urbilateria, la mise en place d'un organe photorécepteur, et qu'à partir là, cet organe a pu évoluer et être "amélioré" de bien des façons différentes, d'où l'immense diversité des yeux observées aujourd'hui (figure 1.3). Vopalensky *et al.* [2012] montrent que la mise en place précoce des yeux est gouverné par les mêmes facteurs de transcriptions chez l'amphioxus chez les vertébrés. Ainsi, les yeux seraient des homologues profonds [Shubin *et al.* , 2009], car chez l'ancêtre commun était déjà présent les réseaux génétiques permettant de la mise en place des cellules photos réceptrices

#### 1.1.3.2 L'homologie sérielle, exemple de l'évolution des membres

L'homologie sérielle est la répétition d'une structure (ou d'un module développemental) au sein d'un organisme, tels les vertèbres, les dents au sein de la mâchoire ou encore les membres antérieurs et postérieurs [Wagner *et al.* , 2007]. Je prends le temps de bien définir ce concept, car les dents, qui sont le modèle d'étude présenté dans cette thèse, sont des organes sériels. Cependant, le cas des membres antérieurs/postérieurs est celui qui a reçu le plus d'attention dans la littérature Young & HallgrÍmsson [2005]; Ruvinsky & Gibson-Brown [2000].

Plus que des homologues sérielles, les membres sont aussi des homologues structuraux. Ainsi, les membres antérieurs des tétrapodes tels que l'aile de la chauve souris, la patte antérieure du chat, l'aile du poulet ou encore la patte antérieur du cheval, sont des homologues structuraux. C'est même l'exemple type pour illustrer la notion d'homologie. Enfin, les membres antérieurs/-postérieurs sont aussi homologues profonds car ils partagent des réseaux de gènes qui étaient présents chez l'ancêtre commun des tétrapodes [Shubin *et al.* , 2009]. L'exemple de membres, montre bien que l'homologie est un concept à plusieurs niveaux et qu'il est difficile de le définir clairement. C'est pourquoi lorsque l'on parle d'homologie il est nécessaire de définir de quel type d'homologie (de quel niveau) il s'agit. Cependant, l'homologie, quel que soit le niveau traité, reste un concept essentiel pour comprendre l'évolution des organismes.

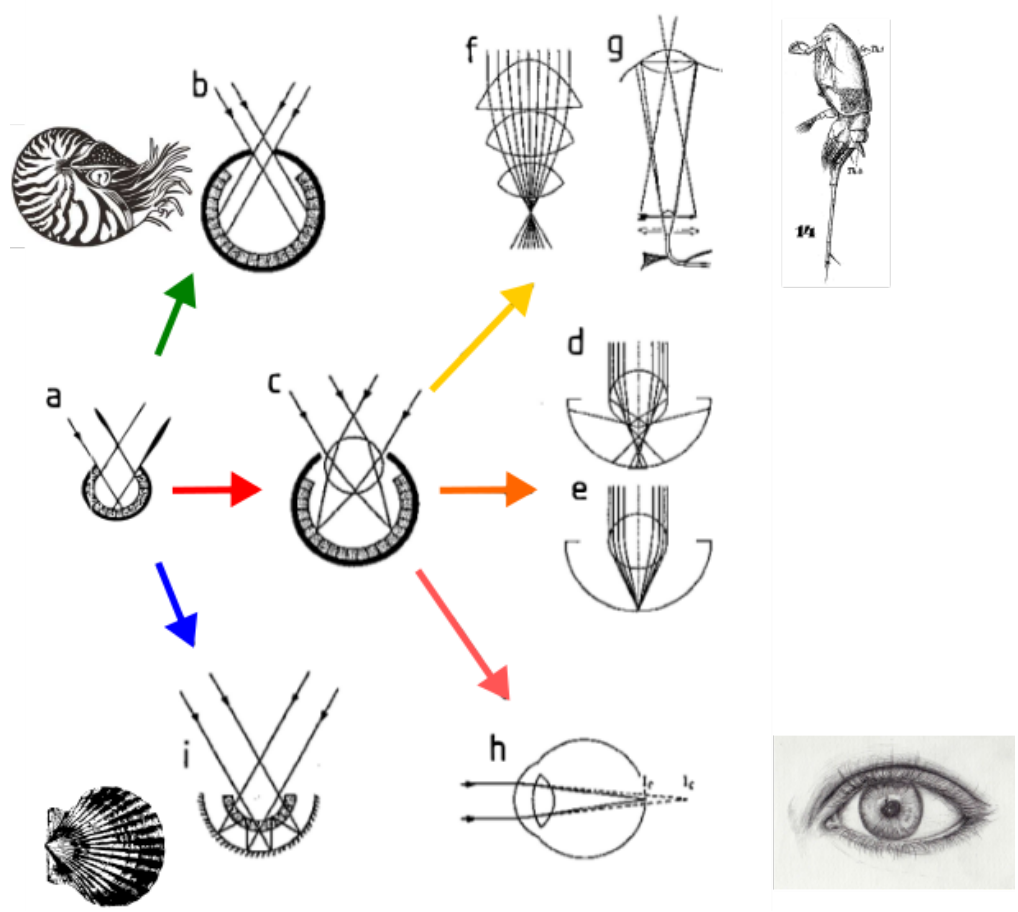



FIGURE 1.3 – **Évolution de l'oeil à chambre unique**.
(a) Oeil à fosse ("pit eye"), commun aux phyla inférieurs. (b) Oeil en trou d'épingle type nautille *Nautilus*ou *Haliotis*. *(c)* Oeil avec lentille. *(d)* Lentille homogène. *(e)* Lentille non-homogène "Matthiessen" *(f)* Lentille multiple du mâle *Pontella*. *(g)* Lentille double de *Copilia*; la flèche pleine représente la position de l'image et la flèche vide le mouvement de la seconde lentille. *(h)* Oeil d'Homme avec cornée et lentille. Ie, image formée par la cornée seule ; Ir, image finale sur la rétine. *(i)* Oeil "mirroir" de la coquille St Jacque (*Pecten*) (Adapté de Land & Fernald [1992])

## 1.2 Évolution génétique du développement couplée à une évolution phénotypique

### 1.2.1 Évolution de la boite à outils génétique

La "boîte à outils" développementale, ou genetic toolkit, se compose d'un nombre limité de familles de gènes et de voies de signalisations qui sont communes à toutes les espèces et qui sont utilisées au cours du développement, comme par exemple les gènes Hox. Les familles sont certes peu nombreuses, mais au sein de ces familles les pertes et les gains de gènes sont très fréquents [Robertis, 2008]. C'est notamment cette grande diversité de la composition des

ces quelques familles qui est à relier à la grande diversité morphologique observée [Bertrand *et al.* , 2004]. Nichols *et al.* [2006] démontrent que les voies Wnt, TGF$\beta$, Notch, Hedgehog, Jak/Stat et RTK (récepteurs à tyrosine kinase), soit 6 voies de signalisation des bilatériens, sont présentes chez l'éponge basale *Oscarella carmela*. De plus, Kusserow *et al.* [2005] montrent que 11 des 12 sous-familles de *Wnt* sont présentes et exprimées chez le cnidaire *Nematostella vectensis* et seraient impliquées dans la gastrulation de l'anémone de mer. Cette conservation va bien au delà de la simple conservation de séquence, dans certains cas il y a aussi conservation de l'activité biochimique et donc du rôle biologique dans le processus développemental. Un exemple emblématique est celui du gène *Pax-6* et de son orthologue chez la drosophile, *Eyeless*. Ainsi il a été montré que le gène *Pax-6* est impliqué dans le développement des yeux de diverses métazoaires (revue dans Halder *et al.* [1995b], voir partie 1.1.3). De même, la dent, qui est un appendice ectodermique, réutilise les mêmes réseaux génétiques que les plumes, les glandes mammaires, les poils et autres appendices ectodermiques [Pispa & Thesleff, 2003].

### 1.2.2 Pléiotropie et mutations dans les régions *cis*-régulatrices

Les gènes pléiotropes sont des gènes qui sont exprimés à plusieurs endroits de l'organismes, dans plusieurs organes. Les gènes du développement sont des gènes largement pléiotropes [cite]. Dans le cas particulier du développement, la pléiotropie signifie que les mêmes gènes vont participer à la mise en place de plusieurs traits. Ceci va contraindre l'évolution des processus développementaux, car il y a donc une forte interconnexion entre les traits. Les régions *cis*-régulatrices du gène, peuvent impacter la transcription, le traitement post-transcriptionnel, la traduction ou encore la dégradation de la protéine ou du transcrit. Les mutations dans les régions *cis*-régulatrice vont pouvoir impacter sur où et/ou quand le gène est exprimé. Ainsi des modifications dans la région *cis*-régulatrices permettent de s'affranchir de la contrainte de la pléiotropie. Il a d'ailleurs été proposé que les mutations dans les régions *cis*-régulatrices seraient plus fréquentes [Wray, 2007; Carroll, 2008] que les mutations dans les régions codantes. Bien qu'Hoekstra & Coyne [2007] ne soient pas d'accord avec cette hypothèse.

### 1.2.3 Réseaux génétiques et co-option

Les réseaux de régulation génétique, ou GRNs pour Genes Regulatory Networks, sont des ensembles de gènes qui sont co-exprimés ou exprimés en cascades et qui vont permettre la mise en place d'une structure quasi-autonome ou modules (voir la partie 1.2.4). Les gènes du développement sont assemblés en réseaux [Davidson, 2010]. La co-option est la ré-utilisation d'un réseau génétique dans un autre organe, ou dans un autre endroit du corps. Pour illustrer ce concept, prenons l'exemple des dents chez les poissons téléostes et plus particulièrement les cichlides du lac Malawi qui présentent à la fois des dents orales et des dents pharyngiennes. Fraser *et al.* [2009] montrent que le nombre de dents orales et le nombre de dents pharyngiennes sont corrélées bien que les deux caractères soient évolutivement découplés. Les auteurs proposent

donc l'existence d'un réseau de régulation du développement de la dent ancestral qui aurait été co-opté dans la mâchoire orale et la mâchoire pharyngienne pour mettre en place les dents. Fraser *et al.* [2009] identifient les gènes clés qui sont utilisés aussi bien pour les dents orales que pharyngiennes et entre différents vertébrés (téléostes, tétrapodes), leurs résultats sur l'évolution des dents chez les vertébrés à mâchoires et leurs hypothèses sur la co-option du réseau génétique est représenté dans la figure 1.4B. Je reviendrai sur l'évolution des dents dans la partie 3.1.2.

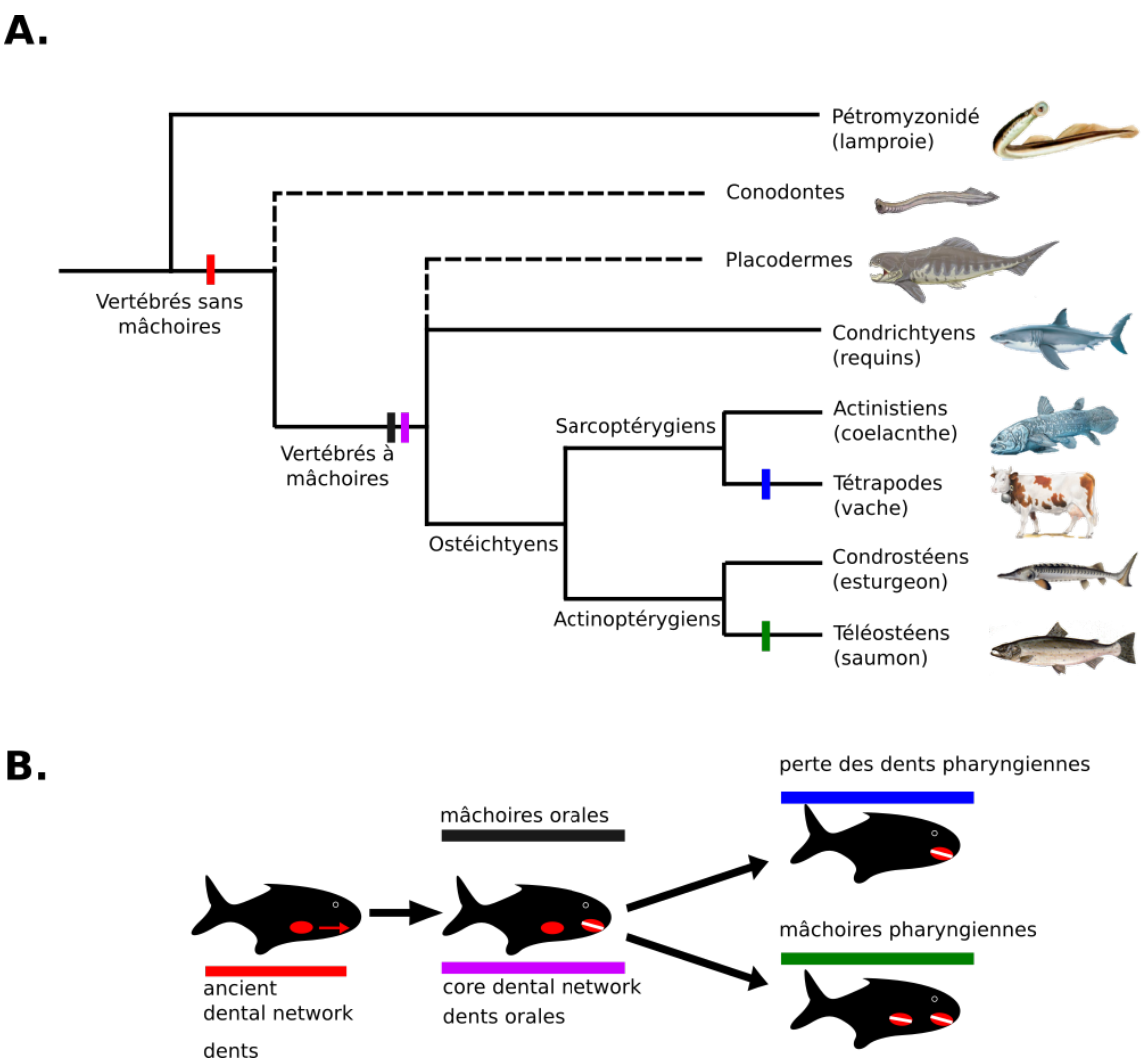



FIGURE 1.4 – **Réseaux génétiques et évolution de la dentition des gnathostomes.**
**(A.)** Arbre phylogénétique des vertébrés à mâchoires. Les traits en pointillés représentent les groupes éteints. La légende des barres de couleurs est explicitée dans le panel B. **(B.)** Représentation schématique de l'apparition des dents et des réseaux de signalisation permettant la mise en place des dents au cours de l'évolution des vertébrés à mâchoires. (Adapté de Panafieu & Gries [2007]; Fraser *et al.* [2009])

### 1.2.4 Les organismes sont des structures modulaires

L'assemblage en réseaux de gènes permet l'émergence de la modularité. Un module est un réseau d'interaction minimum qui possède sa relative autonomie par rapport au reste de l'organisme [Wagner, 1996; Wagner *et al.* , 2007]. Il existe plusieurs niveaux de modularité, comme revue par Wagner *et al.* [2007] : modularité variationnelle, fonctionnelle, développementale. Je vais ici rentrer un peu plus en détail dans le concept de modularité développementale. La modularité développementale peut désigner deux concepts différents, (1) elle peut définir une partie d'un embryon qui a une existence quasi autonome en terme patron de formation et de différentiation (tels que les bourgeons de membre, les dents) ou (2) elle peut désigner une voie de signalisation autonome [Bolker, 2000; Wagner *et al.* , 2007]. Du fait de l'autonomie de certaines parties par rapport aux autres l'évolution peut alors impacter chaque module indépendamment des autres. Les dents sont un exemple de modularité développementale. En effet la dentition est composée de modules plus ou moins autonomes que sont les dents [Stock, 2001]. Cette propriété de la dentition va avoir des effets sur la façon dont vont pouvoir évoluer les dents au sein de la dentition. En effet, chaque module, du fait de sa quasi-autonomie pourra être vu de façon indépendante par rapport à la sélection. Je reviendrai plus en détail sur l'évolution des dents dans le chapitre 3. Un autre exemple de modularité développementale est la largeur et la longueur du bec des pinsons de Darwin. Ainsi, Abzhanov *et al.* [2004, 2006] ont montré que la longueur et la largeur du bec chez les pinsons de Darwin sont deux caractères évolutivement découplés qui sont contrôlés par deux réseaux génétiques bien distincts. Ainsi des différences de niveaux d'expression de *Bmp4* vont modifier la largeur du bec, alors que ce sont différents niveau de *calmoduline* qui sont responsables des variations de longueur du bec. Cependant dans certaines espèces, les deux traits longueur/largeur sont corrélés. Wagner *et al.* [2007], dans leur revue sur la modularité, en concluent que l'indépendance évolutive de deux traits ne garantit pas l'existence de programmes développementaux indépendants.

### 1.2.5 Le paysage épigénétique de Waddington, un modèle pour penser le développement et l'évolution génétique

Le développement est un processus complexe qui permet la mise ne place des caractères qui seront vus par la sélection. Il est donc important d'avoir une vision globale et de définir un cadre conceptuel qui permettent de penser comment évolue et comment est contraint le développement. Dans les années 1940, Waddington propose un cadre conceptuel pour représenter le développement, connu sous le nom de paysage épigénétique. Il utilise la métaphore d'une bille dégringolant dans un paysage vallonné (*fig. 1.5 A*), dont les reliefs sont contrôlés par les gènes (*fig. 1.5 B*). Durant son "voyage" à travers ce paysage la bille va suivre un tel ou tel chemin en fonction des obstacles (les reliefs) qu'elle va rencontrer. Ce qui est intéressant de noter dans ce cadre conceptuel c'est qu'il prend en compte la notion d'historique, puisque les décisions de la bille en aval sont influencées par le chemin qu'elle a suivi en amont. Ce paysage est utilisé

pour représenter métaphoriquement une cellule ou un embryon et son parcours tout au long du développement. Les dessous du paysage se composent de l'information génétique, et donc des réseaux de gènes de l'embryon (*fig. 1.5 B*). L'idée derrière cette représentation est que les gènes ont des effets quantitatifs et que la coordination de leurs actions façonne le paysage. Ce paysage peut être utilisé pour penser les effets des mutations sur le programme développemental. Ainsi, dans ce cadre conceptuel, une mutation ayant des effets sur le phénotype final sera une mutation qui change la topologie du paysage, en cassant un filin par exemple.

## 1.3 Evolution génétique du développement sans évolution phénotypique

### 1.3.1 Évolution cryptique

Une mutation est dit cryptique lorsqu'elle n'a pas d'effet sur le phénotype final. Si l'on se replace dans le paysage épigénétique de Waddington, cela signifie que les mutations cryptiques ne vont pas modifier la topologie du paysage, bien qu'elles puissent jouer sur la tension et/ou la position des filins, comme représenté sur la figure 1.5C (panel du haut). De telle sorte que lorsqu'une mutation non-cryptique va venir perturber le paysage, en provoquant la rupture d'un filin, les conséquences de cette mutation seront différentes en fonction du contexte cryptique dans lequel elle a lieu. Ceci est illustré au niveau de la figure 1.5C (panel du bas), où l'on peut voir que l'effet d'une mutation identique (corde coupée) sera différente en fonction de chaque contexte cryptique [Paaby & Rockman, 2014]. Cet effet contexte ou background est bien connu des généticiens.

### 1.3.2 La dérive développementale

La dérive développementale, ou Developmental System Drift (DSD), est la divergence entre programmes développementaux malgré une forte conservation morphologique d'un trait (figure 1.6) [True & Haag, 2001; Kiontke *et al.* , 2007]. L'hypothèse la plus répandue et implicitement acceptée est qu'une morphologie similaire dans deux espèces différentes sous entend une conservation des programmes développementaux permettant de mettre en place cette structure similaire. Félix [1999] résume dans sa revue sur l'évolution des mécanismes développementaux chez les nématodes que dans 5 espèces du genre nématode, il y a 5 mécanismes développementaux mettant en place les cellules précurseur de la vulve alors que la morphologie finale est similaire. Ces conclusions ont été obtenues par analyse des patrons d'expression des gènes *Lim-39* et *Hom-c* ainsi que l'étude du lignage cellulaire des cellules de la vulve. Kiontke *et al.* [2007] étudient la mise en place de la vulve de nématode de façon plus systématique, dans 51 espèces de *rhabditis* pour une quarantaine de caractéristiques du développement (compétence, induction, identité cellulaire...). Cette étude montre une forte variabilité des mécanismes de développement permettant la mise en place de la vulve. Dans le cadre conceptuel du paysage de Waddington, la

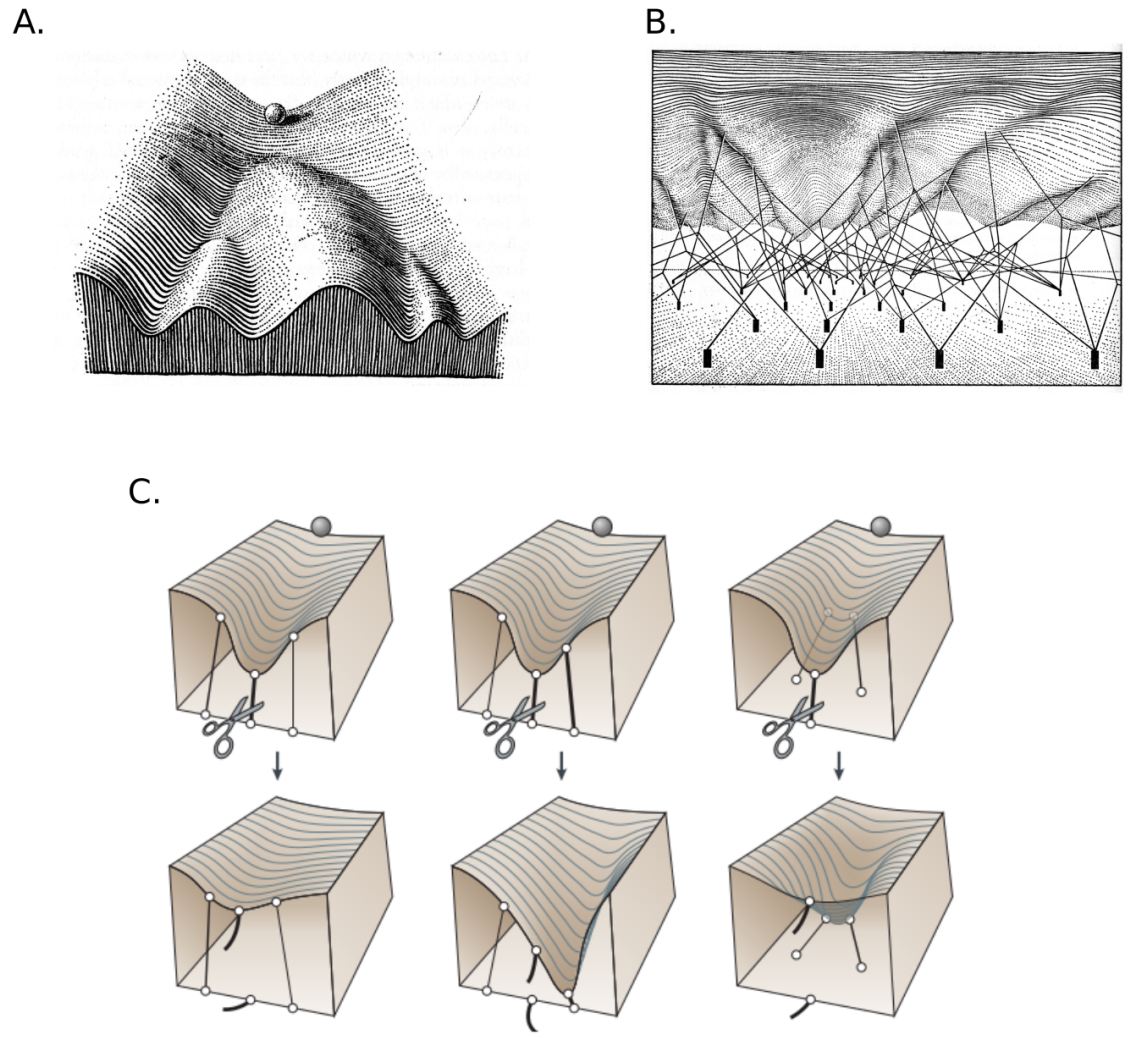



FIGURE 1.5 – **Paysage épigénétique de Waddington**.
**A.** Représentation du paysage épigénétique. La bille représente une cellule ou une partie d'un embryon. Le chemin suivi par la bille lorsqu'elle roule à travers le paysage, correspond à son "histoire" développementale . **B.** Les dessous du paysage de Waddington. Les gènes sont représentés comme des piquets ancrés au sol desquels partent des filins qui façonnent les vallées du paysage. (Adapté de Slack [2002]) **C.** La cassure d'un filin, due à une mutation non cryptique, a une résultante différente sur le paysage adaptatif selon l'environnement cryptique dans lequel elle a lieu. (Tiré de Paaby & Rockman [2014])

dérive développementale est l'existence de plusieurs chemins qui mènent à un même point final. Le DSD est en fait un concept plus intuitif qu'il n'y parait. Rappelons ici que ce qui est vu par la sélection est la morphologie finale, la sélection n'a pas de prise directe sur le programme de développement qui permet la mise en place du trait. C'est pourquoi la dérive développementale peut être observée.

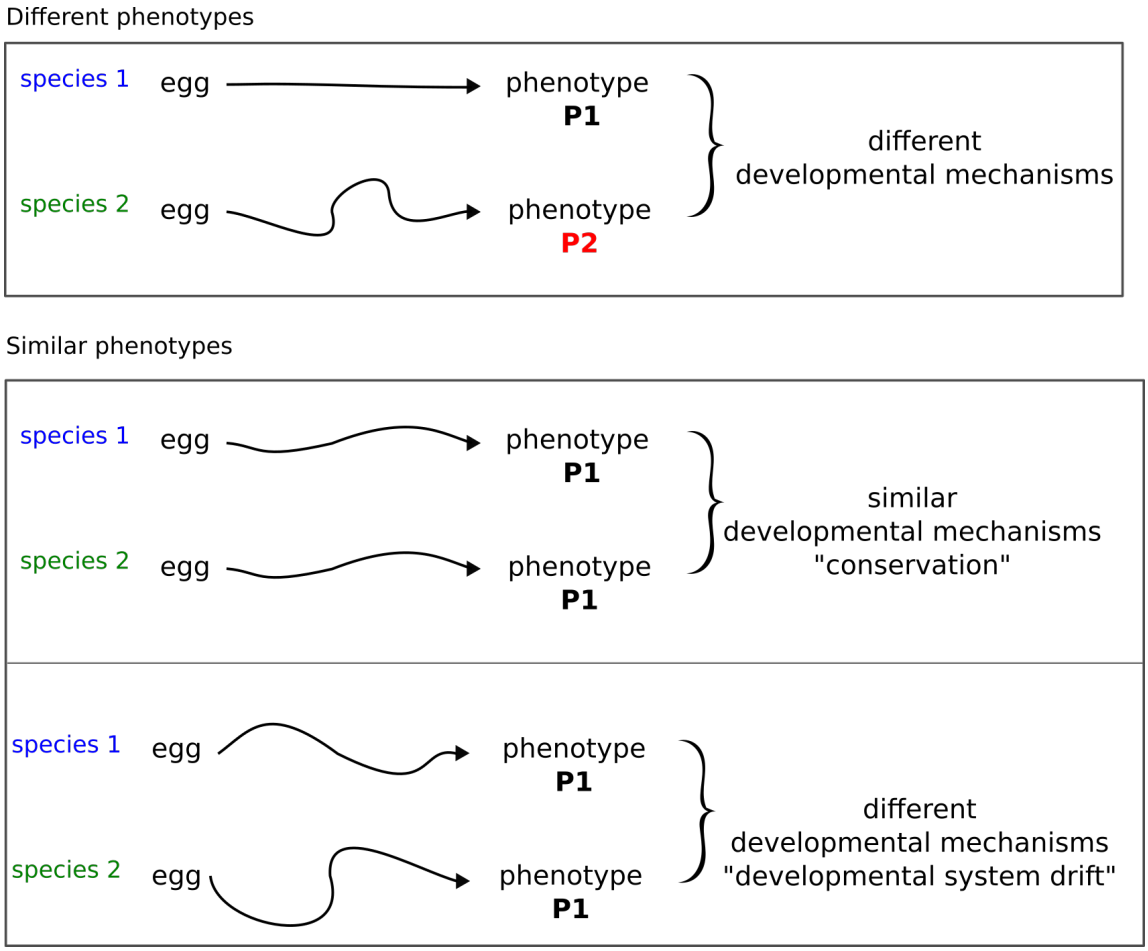



FIGURE 1.6 – **Chemins développementaux et morphologies**
Représentation schématique des programmes développementaux dans deux espèces en fonction de la similarité entre phénotype.

## 1.4 Patrons généraux dans l'évolution du développement

### 1.4.1 Contraintes versus flexibilité dans l'évolution du développement

Dans cette partie, je vais aborder les questions de contraintes au cours du développement. L'évolution du développement est-elle contrainte ? Si oui, comment et par quoi le développement est-il contraint. Certaines périodes sont-elles plus contraintes et d'autres plus flexibles ?

### 1.4.2 Le modèle des additions terminales permet intuitivement une plus grande flexibilité du développement

Comment les innovations morphologiques apparaissent-elles au cours du développement ? La façon la plus intuitive et parce qu'elle semble la plus simple est l'addition terminale. C'est à dire, ajouter un nouveau caractère à la fin du programme de développement, une fois que tous les caractères essentiels ont été mis en place. Cette hypothèse des additions terminales reprend les principes de la loi biogénétique qui sous entend que les innovations morphologiques des groupes "supérieurs" arrivent à la fin du développement. De cette hypothèse, se base l'idée que, du point de vue des contraintes qui ont lieu au cours du développement, cela serait plus facile de rajouter un élément morphologique à la fin du développement. Cela permettrai la mise en place de nouvelles structures sans avoir à toucher au programme de développement qui met en place les structures plus ancestrales.

#### 1.4.2.1 Les modèles du sablier et de l'entonnoir pour comprendre où se retrouvent les contraintes au cours du développement

Cela nous amène à la question des contraintes et de la flexibilité au cours du développement. Les études d'embryologie comparée ont clairement identifié des moments durant l'embryogenèse où les embryons étaient très similaires et partageaient un certain nombre de caractères alors que cette ressemblance n'est plus visible dans les stades ultérieurs voire même les embryons divergent fortement. Cette observation est plus connue sous le nom troisième loi de von Baer (voir la partie 1.1.1). La période pendant laquelle les embryons se ressemblent fortement est nommée stade phylotypique et se retrouve à peu près à mi-embryogenèse. L'hypothèse pour expliquer cette observation est que cette forte similarité morphologique soit causée par de fortes contraintes fonctionnelles au cours du développement. Ces contraintes fonctionnelles limiteraient les possibilités de modifications du processus développemental car la plupart des changements seraient éliminés par sélection négative (purifiante). Parce que les gènes du développement sont fortement pléiotropes, les mutations de ces gènes sont plus susceptibles de causer des phénotypes délétères, contre-sélectionnés. Réciproquement, les divergences ultérieures peuvent résulter soit de la relaxation de la sélection négative, soit alternativement de sélection positive (adaptation). Ce questionnement sur la conservation des stades de développement se résume en fait à deux grands modèles : le modèle de l'entonnoir (funnel) et le modèle du

sablier (hourglass) représentés dans la figure 1.7. Le modèle de l'entonnoir (funnel) se focalise principalement sur les contraintes, qui diminuent au cours du développement. L'hypothèse de ce modèle est que, de par la nature cumulative du développement, les petits changements dans les étapes précoces auront tendance à avoir un impact important sur le phénotype adulte et seront contre-sélectionnés. Le début du développement doit donc être plus contraint que les étapes ultérieures du développement. Le modèle du sablier (hourglass) a été conçu pour s'adapter aux observations morphologiques de von Baer et peut se décomposer en trois étapes. Premièrement, les stades précoces du développement seraient soumis à une forte sélection diversifiante (positive), car ces stades sont fortement dépendants de la niche écologique occupée. Deuxièmement, les contraintes développemementales (sélection négative) atteindraient seulement un maximum au milieu de l'embryogenèse, notamment à cause de la mise en place, par les gènes Hox, du plan d'organisation [Duboule, 1994]. Et enfin, la fin du développement se caractériserait par un relâchement de la sélection négative. Ces trois étapes donnant la forme d'un sablier (figure 1.7). Cependant, d'un point de vue morphologique, l'existence d'un stade phylotypique n'est pas totalement démontré [Richardson *et al.* , 1997]. Les études RNA-seq apportent un nouvel éclairage à cette discussion et sont présentées dans la partie 2.3.2.

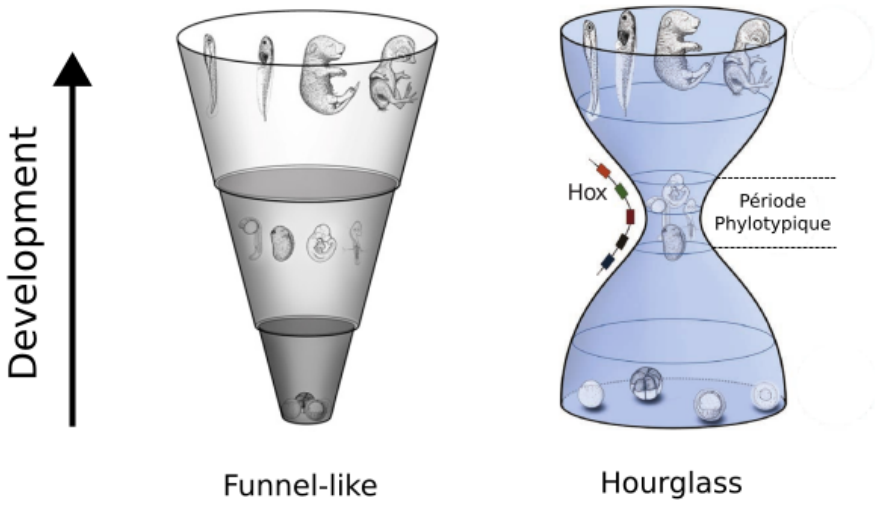



FIGURE 1.7 – **Les modèles de l'évolution des contraintes au cours du développement.**
Présentation des deux modèles de l'évolution des contraintes au cours du développement. Dans le modèle de l'entonnoir (funnel), les contraintes sur le développement sont les plus fortes aux stades précoces de l'embryogenèse et diminuent au fur et à mesure. La forte similarité à mi-embryogenèse serait le résultat des ces contraintes précoces [Roux & Robinson-Rechavi, 2008; Comte *et al.* , 2010]. Dans le modèle du sablier (hourglass), les fortes contraintes sur le développement ont lieu à mi-embryogenèse, durant le stade phylotypique, quand le plan d'organisation est mis en place (notamment par les gènes Hox [Duboule, 1994]). Alors que les stades précoces et tardifs seraient moins contraints (réduction de la sélection purifiante) du fait de l'adaptation aux différentes niches écologiques au début du développement ; quant à la fin du développement il y aurait un relâchement des contraintes. (Adapté de Irie & Kuratani [2011])

### 1.4.3 La notion de bricolage de l'évolution appliquée au développement : l'hétérochronie

L'hétérochronie se définit comme une modification dans la vitesse et la durée d'un processus développemental [Klingenberg, 1998], conduisant à des modifications morphologiques. Le paramètre majeur de la notion d'hétérochronie est le facteur temps. Cette notion peut être récapitulée par l'adage suivant : "faire du neuf avec du vieux en déplaçant les processus dans le temps". Le concept d'hétérochronie a été théorisé par Gould (1977) et son modèle des pendules. Cependant dans ce modèle, Gould sépare deux modules/concepts qui sont en réalité indissociables au cours du développement : la taille et le forme [Klingenberg, 1998]. Cependant, ce cadre théorique permet de penser la flexibilité du développement. Nous avons discuté plus haut que les gènes du développement sont organisés en réseaux, ce qui permet notamment la modularité du développement. Ainsi chaque module peut être déplacé dans le temps ou bien son temps d'action peut être rallongé ou raccourci et ainsi permettre de modifier le phénotype final.

# 2 L'outil transcriptomique : comprendre l'évolution de l'expression et l'évolution du développement

## 2.1 Évolution de l'expression avant l'ère génomique

Dès 1975, il est clair que la "distance génétique" entre l'Homme et le chimpanzé, obtenue par comparaison de séquences (nucléiques ou protéiques), n'était pas suffisante pour expliquer leurs différences phénotypiques. King & Wilson [1975] suggèrent que ce soit des mutations dans les régions régulatrices des gènes, et donc des variations du niveau d'expression, entre Homme et chimpanzé, qui seraient responsables de leurs différences évolutives (voir partie 1.2.2). Cette observation montre bien qu'il est important d'aller étudier les variations du niveau d'expression afin de mieux comprendre l'évolution entre les espèces.

Il existe différentes échelles pour interroger la conservation du niveau d'expression : l'échelle du gène, l'échelle de la famille de gènes et l'échelle globale du transcriptome. Historiquement, les premières études ce sont concentrées sur les études à échelle du gène et de la famille de gènes, par absence de moyens techniques pour poser cette question à l'échelle globale. Ainsi, les premiers outils disponibles étaient l'hybridation *in situ* qui renseigne sur où et quand est exprimé un transcrit. Par cette approche, Averof & Patel [1997] montrent que des modifications du patron d'expression des gènes Hox *Ubx* et *AbdA* corrèlent avec la modification des membres thoraciques en appendices alimentaires. À l'échelle des familles de gènes, l'étude des duplications permet de façon indirecte d'interroger l'évolution du niveau d'expression. Ainsi, les épisodes de duplication peuvent être corrélés à l'apparition d'innovations morphologiques [Sidow, 1996]. En effet, après une duplication trois scénarios sont alors possibles. Dans le premier cas, le ou les gènes alors présents en deux copies peuvent être gardés tel quel, il y aura alors un effet de dosage, le niveau d'expression du gène sera alors potentiellement doublé. Dans le deuxième cas, il peut y avoir sous-fonctionnalisation, c'est à dire que chaque orthologue va assurer une partie de la fonction du gène ancestral, par exemple si le gène ancestral était exprimé dans deux organes différents, chaque orthologue va pouvoir être exprimé dans un des deux organes. Dans le troisième et dernier cas, il peut y avoir néo-fonctionnalisation, c'est à dire que l'un des deux gènes dupliqué peut acquérir

une nouvelle fonction. Par exemple, l'étude de la super-famille des récepteurs nucléaires chez les bilatériens montre qu'après duplication la perte et/ou la rétention de certains de ces gènes est corrélée aux innovations morphologiques chez différents groupes de bilatériens [Bertrand *et al.* , 2004]. Les premières études comparées à large échelle sont basées sur des banques d'ESTs [Ewing *et al.* , 1999; Duret & Mouchiroud, 2000]. Par exemple, Duret & Mouchiroud [2000] montrent que les substitutions non synonymes sont plus fréquentes dans les gènes exprimés spécifiquement dans certains tissus (comme le foie) que dans d'autres (comme le cerveau). D'autre part, les sites non-synonymes des gènes exprimés dans de nombreux tissus (qui sont donc des gènes pléiotropes) évoluent plus lentement que ceux des gènes tissu-spécifiques.

## 2.2 Évolution de l'expression des gènes à l'ère génomique

### 2.2.1 Les données transcriptomiques à large échelle

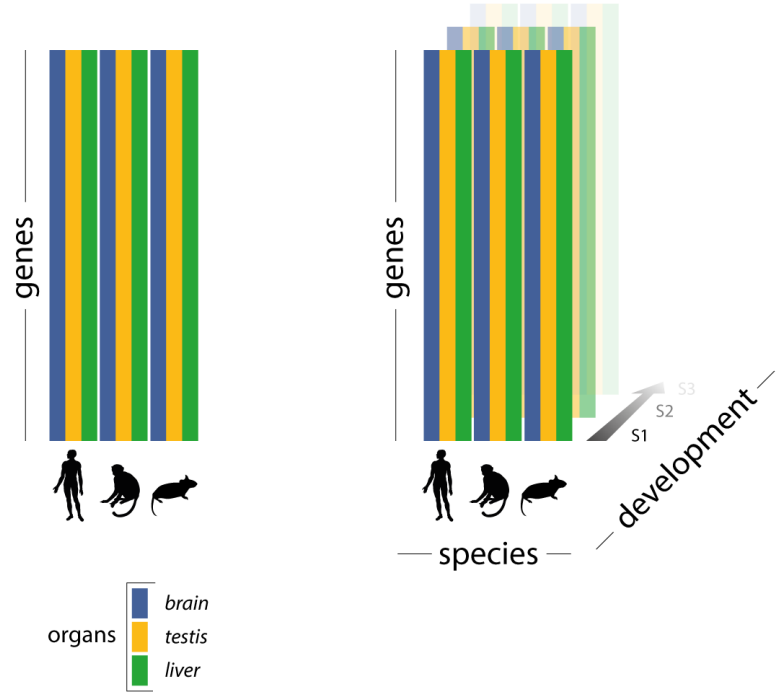



FIGURE 2.1 – **Données transcriptomiques à large échelle**.

La figure 2.1 représente de façon schématique un jeu de données large échelle qui est constitué du niveau d'expression de 10000 à 20000 gènes orthologues dans différents tissus et dans plusieurs espèces [Khaitovich *et al.* , 2005; Brawand *et al.* , 2011]. Dans le cas de l'étude du développement ces données sont multipliées par le nombre de stades de développement étudiés [Levin *et al.* , 2016]. Ce type de jeu de données est très différent des autres jeux de données comparatifs, par exemple le jeux de données (type cancer, comparant tissus sains vs. tissus

malades) car ces données proviennent pour la plupart d'espèces non-modèles, il est donc très difficile de contrôler pour l'âge des échantillons, le sexe, les conditions environnementales, etc. De plus, l'échantillonnage est limité, de ce fait ces études n'ont pas toujours accès à toutes les données nécessaires pour une analyse statistiquement rigoureuse. Enfin, le fait que ces données proviennent de différents tissus et de différentes espèces rend la normalisation de ces données difficiles.
L'assemblage de grands jeu de données comparant l'expression entre de nombreuses espèces est encore aujourd'hui une prouesse technique qui a permis de commencer à répondre à des questions sur le mode d'évolution de l'expression des gènes.

#### 2.2.1.1 Quels sont les modes d'évolution de l'expression ?

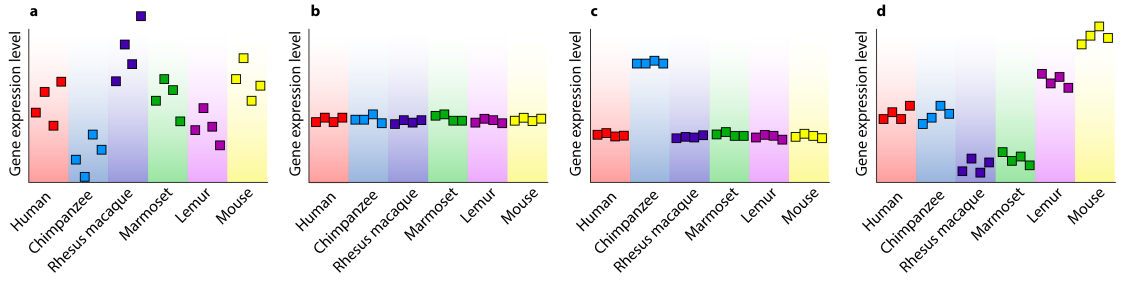



FIGURE 2.2 – **Patrons d'expression attendus en fonction des forces évolutives en jeux**.
**(a.)** Fortes variations du niveau d'expression intra-spécifique et inter-spécifique. Ces variations peuvent être dues à des différences environnementales, ou encore à des différences de stades. **(b.)** Contraintes pour maintenir un niveau d'expression constant. Sélection purifiante/négative. **(c.)** Contraintes pour maintenir un niveau d'expression constant chez les autres espèces, alors que le chimpanzé s'écarte de cette valeur. Un scénario de sélection adaptative est plus probable qu'une relaxation de la sélection purifiante car il n'y a pas de variabilité intra-spécifique. **(d.)** Niveau variable dans les espèces, les différences du niveau d'expression entre espèces sont d'autant plus grandes que les espèces sont éloignées phylogénétiquement, sans tendance général dans la phylogénie (changements d'expressions aléatoires). Neutralité. (D'après Romero *et al.* [2012])

Déterminer le mode d'évolution de l'expression des gènes est un travail actuellement en cours qui suscite de nombreux débats, comme présenté plus bas dans la partie 2.2.1.2). La figure 2.2 représente les patrons d'expressions attendus en fonction des différentes forces évolutives en jeu. Comme représentée sur la figure 2.2d, l'expression varie entre individus et elle varie aussi entre espèces proportionnellement à la distance phylogénétique entre espèces. Cependant, des variations du niveau d'expression entre espèces peuvent aussi être dues à de la sélection adaptative. Théoriquement, une façon de tester la neutralité de l'évolution de l'expression d'un gène serait de déterminer si son niveau d'expression entre espèces varie plus qu'attendu sous le modèle neutre. Si de plus les variations entre individus sont faibles, on pourra conclure si une différence de niveau d'expression entre espèce due à une sélection adaptative (illustré figure 2.2c). Cependant, pour tester l'attendu sous le modèle neutre il faudrait disposer de deux paramètres qui sont le taux de mutations réel ($\mu$) (et non estimé) [Baer *et al.* , 2007] et la taille

effective de la population ($N_e$) [Lemos *et al.* , 2005]. Dans la cas ou le niveau d'expression ne varie pas, ni entre individus, ni entres espèces, comme illustré figure 2.2b, alors la force évolutive en jeu est probablement la sélection purifiante et tous les changements ont été éliminés par la sélection naturelle.

#### 2.2.1.2 Le niveau d'expression est-il sous sélection ? Les différents approches pour répondre à cette question

Il existe deux niveaux de lecture de la variation du niveau d'expression, un niveau populationnel, qui interroge la variabilité entre individus (exemple Enard *et al.* [2002]) ou un niveau pluri-spécifiques, qui interroge la variabilité entre espèces.

**2.2.1.2.1 Données intraspécifiques.** Une première approche consiste à interroger la variabilité du niveau d'expression entre individus d'une même espèce. Lynch & Hill [1986] proposent un modèle de variation mutationelle attendue sous un modèle d'évolution neutre. Deux forces entrent en jeux dans ce modèle : la taille effective de la population ($N_e$) et la variance mutationelle ($V_m$) entre chaque génération. Par exemple, Rifkin *et al.* [2005] déterminent de façon expérimentale le taux de mutation variationnel ($V_m$) entre plusieurs lignées isogéniques de *Drosophila melanogaster*. Ils trouvent que le contexte développemental de l'expression d'un gène va affecter son patron d'évolution, ainsi les gènes fortement exprimés durant le développement ont une expression plus contrainte entre individus, alors que des gènes faiblement exprimés sont plus variables.

**2.2.1.2.2 Données pluri-spécifiques.** Une deuxième approche consiste à interroger la variabilité de l'expression entre espèces. Je vais séparer ce paragraphe en quatre parties, correspondant aux différentes méthodologies pour déterminer le patron d'évolution de l'expression entre espèces.

La première méthode se base sur la comparaison de la conservation entre paires de gènes orthologues et paires de gènes choisies aléatoirement. L'analyse comparée de Chan *et al.* [2009] de trois vertébrés non-mammifères (la grenouille, le poulet, le poisson-globe), de la souris et de l'homme dans une dizaine de tissus montre que plus d'un tiers des orthologues ont un niveau d'expression conservé. Les auteurs concluent alors que le niveau d'expression est contraint et donc soumis à la sélection purifiante. Cependant, le seuil de conservation, qui doit être plus qu'attendu sous le modèle neutre n'est pas connu pour des petites distances phylogénétique. Donc, il est donc difficile de conclure sur la force évolutive en jeu lorsqu'on étudie des espèces phylogénétiquement (trop) proche.

La seconde méthode consiste à comparer le taux de divergence entre espèces en fonction de leur distance phylogénétique. Sous l'hypothèse du modèle neutre, une augmentation linéaire de la divergence est attendue, un peu comme une horloge moléculaire. Khaitovich *et al.* [2004] montrent une augmentation linéaire du la divergence du niveau d'expression en fonction du

temps chez l'homme et le chimpanzé.

La troisième méthode est une approche phylogénétique. Elle consiste à calculer la vraisemblance des données sous un modèle neutre d'évolution de l'expression, souvent un mouvement Brownien. Cette vraisemblance est alors comparée à la vraisemblance obtenue sous modèle d'évolution positive, souvent un modèle OU (Ornstein-Uhlenbeck) qui permet d'introduire un optimum du niveau d'expression. Cependant, ce type d'approche ne peut pas encore être utilisée à son plein potentiel du fait du petit nombre d'espèces comparativement au grand nombre de gènes traités, il n'y a pas assez de puissance statistique pour détecter un effet.

La dernière et quatrième approche consiste à utiliser des approches multivariées type analyse en composante principale (voir box PCA) et clustering. Cette approche permet d'identifier des groupes de transcriptomes qui varient de la même façon. Je développerai cette approche plus bas dans la partie 2.2.2.1, où les grandes signatures transcriptomiques entre espèces et entre tissus sont recherchées.

**BOX | PCA : Analyse en Composante Principale.**

**Le principe**

Le transcriptome peut être vu comme un objet à $n \times p$ dimensions, où $n$ est le nombre de gènes exprimés et $p$ est le nombre d'échantillons (voir figure 2.1). Cependant notre cerveau n'est pas capable de visualiser $n \times p$ dimensions. La PCA, ou ACP pour Analyse en Composante Principale, va réduire le nombre de dimensions tout en conservant la variabilité associée au jeu de données. La réduction du nombre de dimensions s'effectue par l'identification de directions le long desquelles la variation est maximale. Ces directions sont appelées composantes principales (PCs), et elles sont définies de telle sorte qu'elles soient orthogonales entre elles. (Adapté de Ringnér [2008]; Breschi *et al.* [2016]).

**Exemple illustratif**

Si on prend l'exemple d'un nuage de point formant un poisson, le premier axe qui maximise la variation (PC1) suivra la direction la plus allongée, ici de la queue à la tête du poisson. Le second axe (PC2), qui est orthogonal au premier, suivra la seconde direction la plus allongée, ici du ventre vers le dos, et ainsi de suite. Ainsi, la PCA détermine les axes qui expliquent le mieux la variation associée à l'objet (exemple inspiré de [Delafontaine, 2015]. De plus, le pourcentage de variation associé à chaque axe est déterminé. Comme les axes sont définis de façon à maximiser la variation, PC1 est associé au plus fort pourcentage de variation et la variation diminue au cours des axes suivants. La PCA cherche à maximiser la variance, elle va donc donner beaucoup de poids aux points extrêmes, si l'on reprend notre exemple du poisson les points de la têtes et de la queue seront bien identifiés.

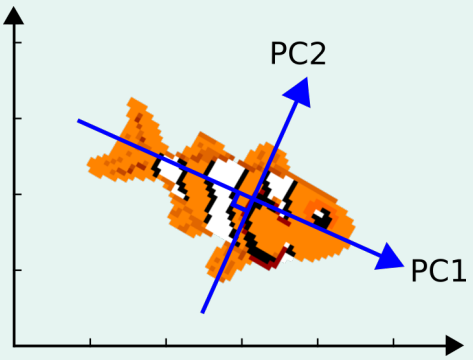



FIGURE 2.3 – Principe de la PCA

**Autres méthodes multivariées**

Il exite d'autres méthodes d'analyses multivariées telles que le multidimensional scaling (MDS), tSNE [Maaten & Hinton, 2008], les «diffusion map» [Coifman *et al.* , 2005; Angerer *et al.* , 2015] (voir chapitre 4 *p. 68*). Les méthodes telles que tSNE et diffusion map sont plus appropriées pour identifier des groupes de données qui partagent un certain nombre de variations continues, par exemple sur de données des cellules en différenciation.

### 2.2.2 Les différents patrons d'évolutions de l'expression

De là ce pose la question des grands patrons d'évolutions du niveau d'expression. La contrainte du niveau d'expression est-elle plus grande entre espèces ou alors entre organes? Le taux d'évolution est-il le même pour tous les organes et pour toutes les espèces? Ce sont les sujets que nous allons aborder dans la suite de cette introduction.

#### 2.2.2.1 L'effet espèce est-il plus fort que l'effet organe?

L'étude à large échelle permet d'interroger le transcriptome de différents tissus dans différentes espèces. La question qui découle alors de ces comparaisons est quelle signature transcriptomique est prépondérante : l'effet espèce est-il plus important que l'effet identité entre organes? La première technique qui a permis d'étudier le transcriptome à large échelle et d'interroger les variations et l'évolution de niveau d'expression à large échelle est la puce à ADN (ou microarray). Yanai *et al.* [2004] étudient le transcriptome de 16 tissus chez la souris et l'homme, grâce aux puces à ADN. Ils montrent que les tissus groupent par espèces, ce qui semble indiquer que l'effet espèce est plus fort que l'identité de l'organe. Cependant, une ré-analyse de ces données par Liao & Zhang [2005] conclut que l'identité entre organes prévaut sur l'identité spécifique (effet espèce). Les auteurs de cette seconde étude arrivent à cette conclusion car ils ont utilisé une méthode différente pour normaliser les puces, qui sont des puces différentes pour chaque espèces [Su *et al.* , 2002].

Une controverse sur le même sujet est apparue 10 ans plus tard, qui met en évidence une difficulté technique de l'analyse des transcriptomes. Cette fois, il s'agit de données RNA-seq et non plus de données de puces. Le RNA-seq [Wang *et al.* , 2009], pour séquençage d'ARN, est une technologie qui permet aussi des études à large échelle (sur la totalité du transcriptome) avec une bon taux de reproductibilité [Marioni *et al.* , 2008], cependant l'avantage de cette méthode c'est qu'elle permet de travailler sur des espèces non-modèles. Ainsi Lin *et al.* [2014], analysent des données du consortium ENCODE, de souris et les comparent a des données transcriptomiques humaines et trouvent une signature de l'espèce. Néanmoins, une ré-analyse de ces données par Gilad & Mizrahi-Man [2015], montre que le fort effet espèce trouvé par Lin *et al.* [2014] est en fait dû à un effet du lot (batch) de séquençage, un artéfact dû à la provenance différente des échantillons et de runs de séquençage différents. Ainsi, après la prise en compte de cet effet, Gilad & Mizrahi-Man [2015] montrent la prévalence de l'identité entre organe sur l'effet espèce. Cette controverse permet donc de mettre en évidence que les études comparées de transcriptomes à large échelle sont sensibles à des effets de lot, tel le run de séquençage mais aussi l'estimation du niveau d'expression biaisée par la disponibilité de génome de référence. Brawand *et al.* [2011] ont étudié, par RNA-seq, les transcriptomes de 6 tissus (cerveau, cervelet, coeur, rein, foie et testicules) dans 10 espèces représentatives des lignées majeures des mammifères (homme, chimpanzé, bonobo, gorille, orang-outan, macaque, souris,opossum, ornithorynque et poulet). Une analyse en composante principale (PCA, voir box PCA) des ces données transcriptomiques

montre une forte identité transcriptomique entre organe. La figure 2.4 montre en effet que les organes sont groupés entre eux sur l'axe 2 de la PCA. Brawand *et al.* [2011] concluent donc que l'identité des organes prévaut sur l'identité spécifique. Récemment, Breschi *et al.* [2016] quantifient la part relative de l'effet espèce et de l'effet organe dans le transcriptome. Ils montrent qu'il y a un continuum dans la variation du niveau d'expression entre des gènes qui varient fortement entre organes et d'autre qui varient fortement entre espèces. Donc la question à poser n'est pas tant quelle signature prévaut dans le transcriptome, mais quels patrons de variation sont responsables de ces signatures transcriptomiques. Une possibilité serait aussi de quantifier la part de chaque variation à la signature transcriptomique.

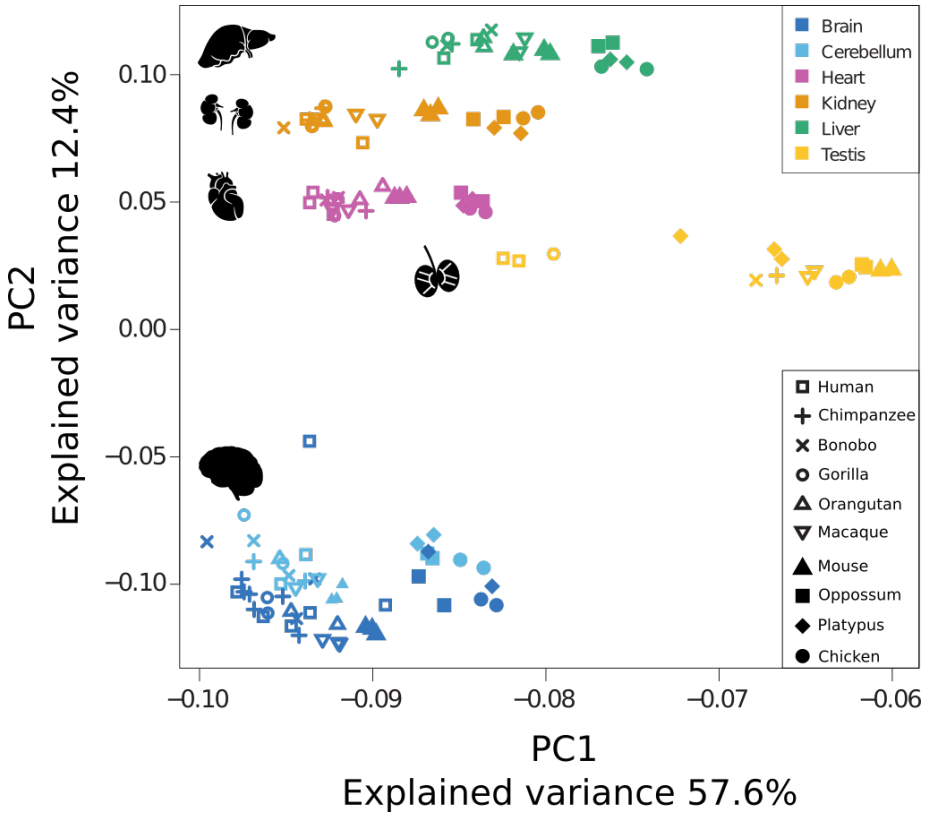



FIGURE 2.4 – **Analyse en composante principale de transcriptomes d'organes adultes chez différents mammifères**.
L'analyse en composante principale (voir fig 2.2.1.2.2) des transcriptomes de 6 organes adultes venant de 10 espèces représentatives des lignées majeures des mammifères groupe clairement les organes sur l'axe 2 (PC2, variance associée 12.4%). L'Axe 1 (PC1) qui représente 57.6% de la variation totale entre les transcriptomes est un axe hybride qui sépare d'abord les testicules des autres organes et ensuite on voit un gradient de séparation d'espèce pour chaque organes. (Adapté de Brawand *et al.* [2011])

#### 2.2.2.2 Les différentes vitesses d'évolution

Les analyses à large échelle des niveaux d'expression, d'abord par microarray puis par RNA-seq, ont permis d'identifier des patrons d'évolutions différents en fonction des organes considérés. Ainsi, le niveau d'expression dans le tissu nerveux évolue moins rapidement, alors que le foie montre moins de conservation du niveau d'expression [Khaitovich *et al.* , 2005, 2006; Chan *et al.* , 2009; Brawand *et al.* , 2011; Necsulea & Kaessmann, 2014]. Pour expliquer ces différences de taux d'évolution entre organe, il faut faire l'hypothèse que chaque tissu est associé à un niveau de contrainte différent. Ainsi le cerveau est l'organe dont les niveaux d'expression sont les plus contraints, alors que le foie est moins contraint. Le faible taux d'évolution du niveaux d'expression dans le tissu nerveux peut être dû à une forte pression de sélection purifiante ou à une action moindre la sélection adaptative comparativement aux autres tissus. Ces analyses comparées ont aussi mis en évidence la spécificité des testicules. Les testicules sont le tissu qui ont le plus fort taux d'évolution du niveau d'expression entre espèces. Cela pourrait être dû à une relaxation de la pression de sélection dans les testicules ou alors à une transcription plus permissive du fait de l'état particulier dans la chromatine dans les testicules, ou encore une forte pression diversifiante relative à la compétition spermatique (revue dans Necsulea & Kaessmann [2014]). Khaitovich *et al.* [2005] montrent que la signature transcriptomique du tissu est portée à la fois par des gènes tissu-spécifique mais aussi par le reste des gènes exprimés dans le tissu. Ainsi l'effet organe ou l'effet espèce ne serait pas une propriété portée par l'ensemble des gènes du transcriptome, mais serait portée par des groupes de gènes distincts. Ainsi, en fonction des organes et/ou des espèces considérés l'un de ces deux groupes prendrait le pas sur l'autre.

Ces études ont aussi permis de mettre en évidence des différences dans les vitesses d'évolution des niveaux d'expression entre espèces [Enard *et al.* , 2002; Khaitovich *et al.* , 2005; Brawand *et al.* , 2011]. Par exemple, le niveau d'expression évolue plus rapidement chez les primates que chez la souris [Brawand *et al.* , 2011; Necsulea & Kaessmann, 2014].

## 2.3 La transcriptome apporte un nouvel éclairage aux questions de l'évo-dévo

### 2.3.1 L'identité entre organes

L'identité d'un caractère est fortement lié au réseau génétique qui permet sa mise en place au cours du développement. C'est pourquoi, l'étude du niveau d'expression des gènes à l'échelle du génome devrait fournir des informations sur l'identité d'un caractère [Arendt, 2008]. Le transcriptome d'un organe à un temps t du développement peut être utilisé comme un phénotype moléculaire [Pantalacci & Sémon, 2014]. L'analyse transcriptomique d'organes au cours du développement est donc un outil précieux pour déterminer des homologies entre caractères. Ainsi, Wang *et al.* [2011] utilisent l'approche transcriptomique pour déterminer l'identité des

3 doigts de l'aile chez les oiseaux car les données embryologiques et paléontologiques sont en désaccord. Cette étude permet de conclure que le doigt le plus antérieur de l'aile est homologue au premier des quatre doigts de la patte chez l'oiseau, bien que le premier doigt de l'aile se développe spatialement à la place du doigt numéro 2 (de la patte ancestrale à 5 doigts des tétrapodes), comme représenté à la figure 2.5. Ce résultat suggère un décalage postérieur ("frame shift" hypothesis) de la position des doigts de l'aile (voir Young *et al.* [2011]; Carkett & Logan [2011] pour une revue sur le sujet). Cependant l'identité transcriptomique des doigts 2 et 3 de l'aile ne correspond à aucun des trois autres doigts de la patte de poulet, suggérant l'acquisition d'une identité propre à ces doigts. Tschopp *et al.* [2014] étudient l'origine des organes génitaux externes chez la souris et l'anole (un lézard). Il a été montré que le bourgeon de membre et le tubercule génitale partagent de similarités moléculaires, patron d'expression commun pour certains gènes (*Hox, Fgfs et Shh*) [Kondo *et al.* , 1997; Yamada *et al.* , 2006; Cohn, 2011; Lin *et al.* , 2013], qui suggère une homologie profonde entre les bourgeons de membre et le tubercule génital. Une analyse multivariée des transcriptomes de bourgeons de membre et de tubercules génitaux de souris et d'anole à deux stades de développement, montre que le premier axe de variation est associé à l'espèce et le deuxième axe de variation est associé à l'identité de l'organe, comme observé par Brawand *et al.* [2011] pour des organes adultes. Ces études montrent donc que la transcriptomique permet d'interroger l'identité des organes en se basant sur la ressemblance de leur profil transcriptomique.

### 2.3.2 Quand l'expression est-elle contrainte au cours du développement ?

L'étude de l'évolution de l'expression au cours du développement à une échelle globale, va permettre l'identification des périodes du développement pendant lesquelles l'expression des gènes est la plus contrainte. Ainsi, les premières études basées sur des ESTs (Expressed Sequenced Tags) et des puces à ADN chez le poisson zèbre et la souris, observent que les gènes exprimés précocement sont plus contraints que les gènes exprimés plus tardivement au cours du développement [Roux & Robinson-Rechavi, 2008; Comte *et al.* , 2010]. Ces études proposent que l'existence d'une période phylotypique (contrainte morphologique) soit due à la contrainte génomique qui a lieu au début du développement.

Dans les études montrant l'existence d'une période de développement plus conservée à mi-embryogenèse, on distingue deux types d'approches. D'une part les études qui estiment l'âge du transcriptome. Irie & Sehara-Fujisawa [2007] définissent un"index ancestral". Plus particulièrement, l'index ancestral des vertébrés, qui est le ratio entre le nombre de gènes non-redondant spécifiques des vertébrés (possédant au moins un homologue chez les protostomes, les urochordés, les téléostes et les amphibiens) exprimés au stade $k$, sur le nombre total de gènes non-redondant exprimés au même stade $k$. Les auteurs font l'hypothèse que le stade phylotypique devrait être le stade qui est enrichi en gènes des vertébrés. Ainsi, ils identifient un stade (le "pharyngula state") a mi-embryogenèse qui possède le plus fort ratio d'index ancestral des vertébrés, supportant

ainsi l'idée d'un stade fortement contraint,- le stade phylotypique. Domazet-Lošo & Tautz [2010] s'inspirent de cet index ancestral et proposent une nouvelle métrique, le TAI (Transcriptome Age Index), qui permet de dater l'âge d'un transcriptome. Ainsi ils identifient chez le poisson zèbre et la drosophile (données de puce à ADN), un stade à mi-embryogenèse où le TAI est le plus faible, c'est à dire que le transcriptome est ancestral. Quint *et al.* [2012] identifient aussi un stade phylotypique chez les plantes en utilisant le TAI. Cependant, Piasecka *et al.* [2013] montrent qu'après re-normalisation de ces données, l'âge du transcriptome ne permet pas de détecter la signature d'un stade fortement conservé à mi-embryogenèse. Bien qu'ils ne trouvent pas non plus ce patron d'hourglass au niveau des séquences ni au niveau de l'expression, ils identifient une forte conservation des séquences régulatrices à mi-embryogenèse. D'autre part, les études qui estiment le taux de divergence des niveaux d'expression entre espèces à différents temps de développement. Ainsi, Kalinka *et al.* [2010] montrent une plus forte conservation de l'expression à mi-embryogenèse entre six espèces de drosophiles (figure 2.6). De même Irie & Kuratani [2011] retrouvent ce patron de conservation entre le poisson zèbre, la grenouille, le poulet et la souris.

Ainsi, l'affrontement entre les modèles entonnoir et sablier (présentés partie 1.4.2.1) se voit offrir de nouveaux arguments grâce aux études large échelle. Cependant, il n'y a pas de consensus clair quant à la prévalence d'un modèle par rapport à autre, bien que la signature du modèle sablier soit celle que les études cherchent à mettre en évidence [Irie & Sehara-Fujisawa, 2007; Irie & Kuratani, 2011; Domazet-Lošo & Tautz, 2010; Kalinka *et al.* , 2010; Prud'homme & Gompel, 2010; Quint *et al.* , 2012; Drost *et al.* , 2015]. Notamment, une des difficultés rencontrée est de rendre homologue des stades de développement entre différentes espèces, afin de pouvoir les comparer et aussi d'obtenir des niveaux d'expression comparables entre espèces (voir chapitre 9).

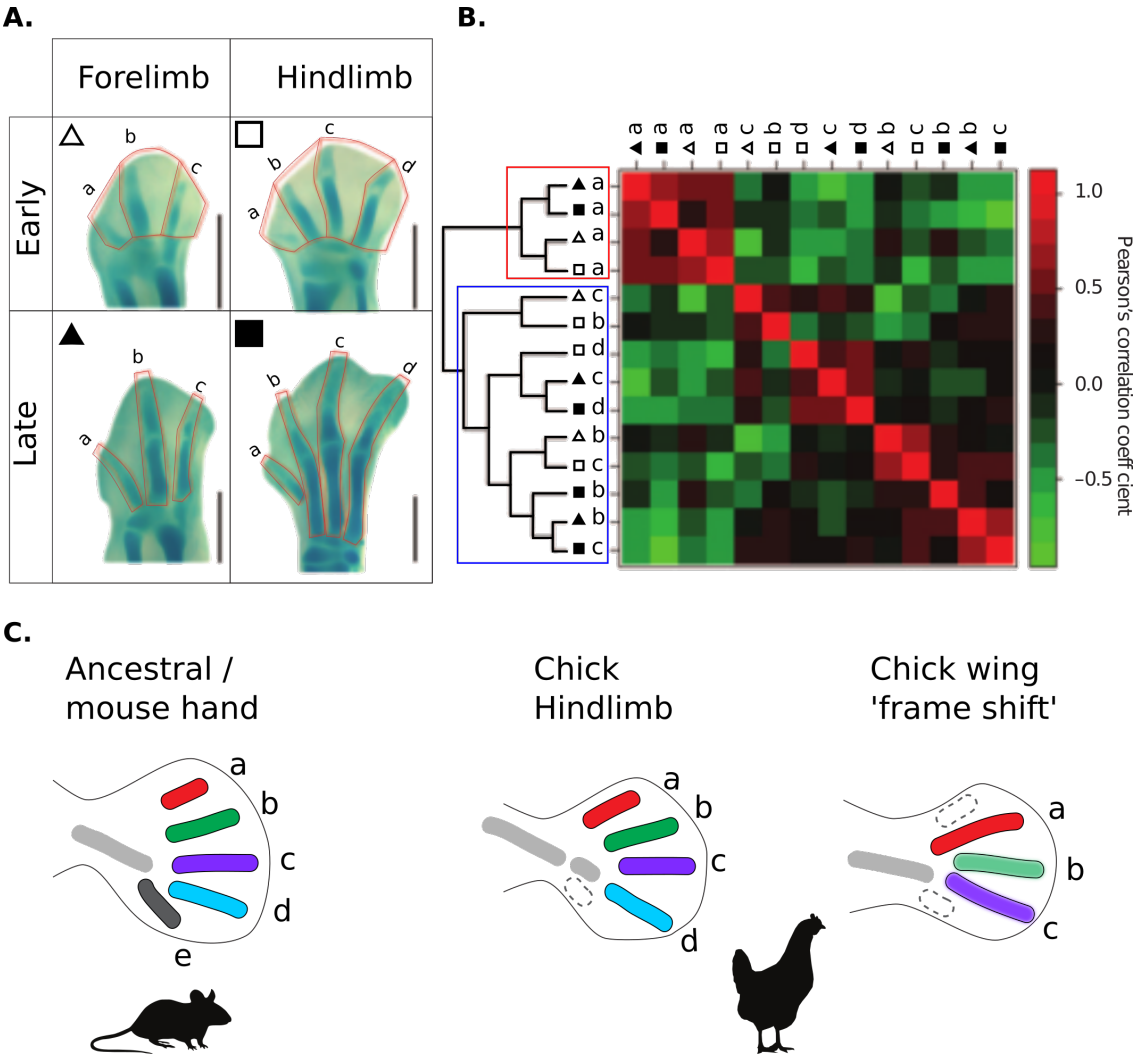



FIGURE 2.5 – **Identité des doigts de poulet**.
**(A.)** Protocole expérimental montrant la dissection des doigts de poulet (rectangles rouge) des membres antérieurs (forelimb) et postérieurs (hindlimb) de poulet à deux stades de développement. Early : stade 28, représenté et Late : stade 31. Le membre antérieur possède 3 doigts notés $a, b$ et $c$. Le membre postérieur possède quatre doigts notés $a, b, c$ et $d$. **(B.)** La matrice des corrélations de Pearson par paire d'échantillon (heatmap), met en évidence la structuration des données en 2 groupes (cadre rouge et bleu). Le premier groupe (cadre rouge) est composé des échantillons antérieurs et postérieurs du doigt $a$. Cela suppose que le doigt $a$ des membres postérieurs et inférieurs sont homologues. Le deuxième groupe (cadre bleu) est un mélange des échantillons des doigts $b, c$ et $d$. Donc il n'y a pas de correspondance claire entre les doigts $b/c$ du membre antérieur et les doigts $b/c/d$ du membre postérieur. Les membres antérieurs sont symbolisés par un triangle ; les membres postérieurs sont symbolisés par un carré. Les symboles blanc représentent le stade *early* et les symboles noir le stade *late*. **(C.)** Représentation schématique du développement des doigts dans les membres de poulet et le membre ancestral (exemple de la souris). La couleur représente l'identité moléculaire. Les données transcriptomiques de Wang *et al.* [2011] suggèrent que le premier doigt de l'aile de poulet, bien qu'il se développe en position 2, est homologue au premier doigt de la patte de poulet, d'où la couleur rouge partagée par tous les doigt $a$. (Adapté de Wang *et al.* [2011]) et Carkett & Logan [2011])

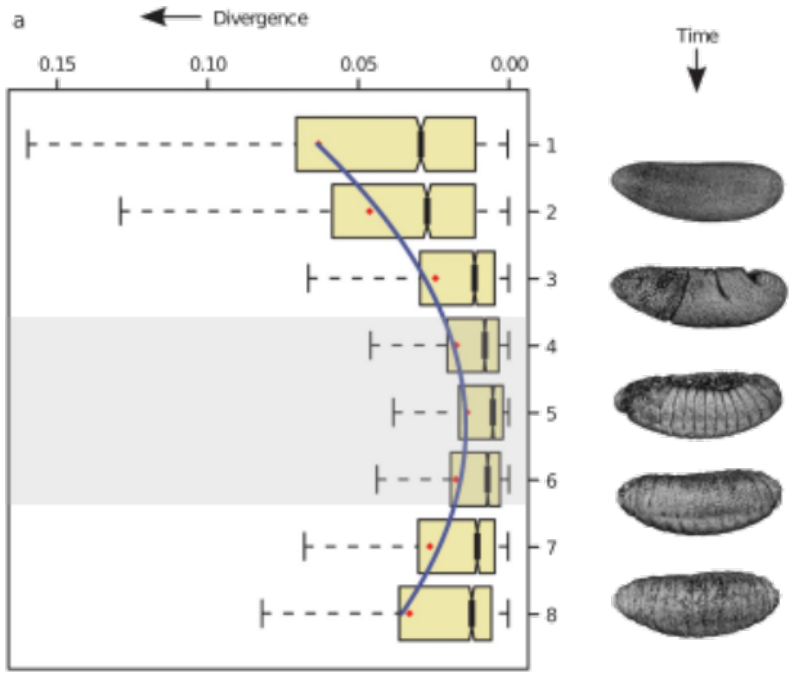



FIGURE 2.6 – **Contrainte type "hourglass" dans les embryons de drosophile**.
Données d'expression pour 6 espèces de drosophile (*D. melanogaster, D. simulans, D. ananassae, D. persimilis, D. pseudoobscura and D. virilis*), dont le temps de divergence le plus important remonte à 40 millions d'années. La zone grisée représente une période du développement où le niveau d'expression est le plus similaire entre espèces. Cette période correspond au stade phylotypique. (D'après Kalinka *et al.* [2010])

# 3 Le modèle dent

Les dents reflètent assez bien l'historique de l'évo-dévo. Il y d'abord eu les études comparées du développement de la molaire, tout en essayant de replacer dans un contexte évolutif [Osborn, 1887]. Les conclusions de ces études du $XIX^{ème}$ siècles sont revues par Butler [1956] . Puis il y eut l'ère de la génétique qui éclipsa les données de morphologie comparées ainsi que le hypothèses sur la mécanistique du développement qui avaient alors été avancées. Les études génétiques du développement de la dent ont permis de mettre en évidence de nombreux gènes et réseaux génétiques (revue dans Tucker & Sharpe [2004]; Jernvall & Thesleff [2012]; Balic & Thesleff [2015]). Elles ont notamment mis en évidence et caractérisé les gènes qui permettent le dialogue entre l'épithélium et le mésenchyme,les deux tissus constitutifs de la dent [Mina & Kollar, 1987; Bei & Maas, 1998; Bei *et al.* , 2000; Munne *et al.* , 2009; Tummers & Thesleff, 2009]. Ces données de développement ont permis l'établissement d'un modèle *in silico* du développement de la dent de mammifère [Salazar-Ciudad & Jernvall, 2002] (que je détaillerai dans la partie 3.2.4). Ce modèle permet de répondre à des questions d'évo-dévo telles que l'émergence de la variabilité morphologique [Salazar-Ciudad & Jernvall, 2010; Harjunmaa *et al.* , 2014].

## 3.1 Les dents et leur évolution

### 3.1.1 Qu'est-ce qu'une dent ?

Les dents sont des organes quasi-minéralisés, constituées exclusivement d'émail et de dentine [Butler, 1956]. De ce fait elles sont extrêmement bien conservées dans les registres fossiles. Ce qui en fait un bon modèle d'évolution pour comprendre l'émergence de la variabilité morphologique [Lazzari *et al.* , 2008]. D'un point du vue morphologique, la dent se compose de racines et d'une couronne, comme représenté sur la figure 3.1. Les dents les plus simples sont coniques, comme les dents de reptiles ou de poissons représentées sur la figure 3.2A ; les dents les plus complexes sont multicuspides, comme les dents des mammifères représentées sur la figure 3.2A. Ce qui est

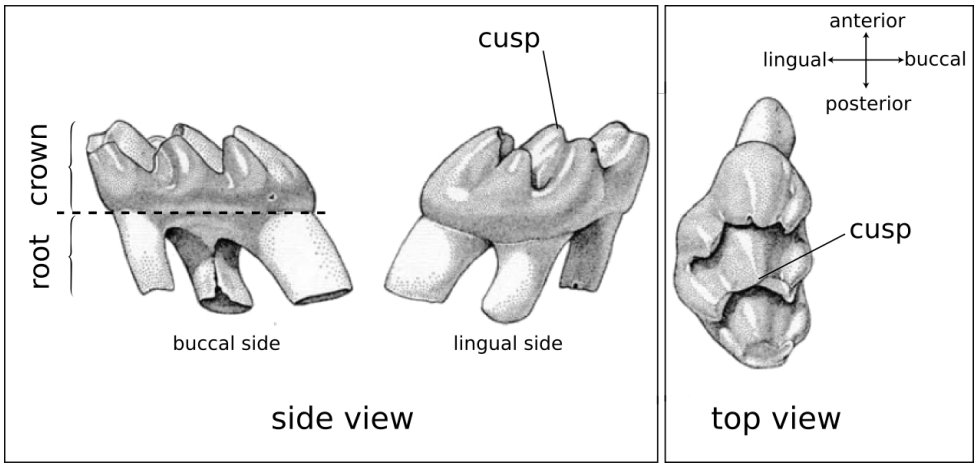



FIGURE 3.1 – **Morphologie de la molaire et terminologie**
Dessin d'une première molaire supérieure de souris (*Mus musculus*) sous différents angles de vue. Les pointes sur les dents sont appelées des cuspides. (Adapté de Grüneberg [1965])

le plus frappant sur cette figure 3.2 c'est la grande diversité morphologique des dents. La forme des dents est fortement corrélée au régime alimentaire [Evans *et al.* , 2006]. La question que je vais me poser ensuite est d'où viennent les dents et quels animaux présentent des dents ?

### 3.1.2 Les dents sont une innovation morphologique des vertébrés à mâchoires

Comme mentionné plus haut, les dents sont des organes servant à l'alimentation, elles sont physiquement et fonctionnellement liées aux mâchoires [McCollum & Sharpe, 2001]. Il existe deux hypothèses majeures pour expliquer l'origine évolutive des dents : (1) l'hypothèse "inside-out", propose que les dents proviennent des denticules pharyngiennes, (2) l'hypothèse "outside-in", couple l'évolution des dents à celles des mâchoires et propose que les dents soient une internalisation des denticules de peau (armure dermique) similaires à celles présentes sur la peau des requins modernes. Pour approfondir ces deux hypothèses, je vous invite à vous référer aux revues suivantes : Huysseune *et al.* [2009]; Fraser *et al.* [2010]. Les dents seraient une innovation morphologique apparue une fois uniquement chez les vertébrés à mâchoires (les gnathostomes). Cependant, l'existence de structures de type "dents" chez les conodontes [Briggs *et al.* , 1983; Sansom *et al.* , 1992] et chez les placodermes [Smith & Johanson, 2003; Smith, 2003] remettent en cause cette hypothèse. Je vous invite à consulter la figure 3.3 qui représente la phylogénie des vertébrés à mâchoires. Elle présente les lignées dans lesquelles les dents "vraies" sont présentent, ainsi que les lignées dans lesquelles des structures types dents ont été retrouvées. La présence de dents "vraies" chez les conodontes impliquerait une apparition des dents découplée de l'apparition de la mâchoire. Cependant, l'homologie entre les éléments des conodontes (voir Goudemand *et al.* [2011] pour une reconstruction fonctionnelle de ces éléments) et les dents "vraies" semble ne pas être avérée [Murdock *et al.* , 2013; Janvier, 2013].

L'homologie entre les denticules retrouvées chez les espèces de placodermes dérivés et les dents n'est pas très claire [Burrow, 2003]. Cependant, Smith & Johanson [2003]; Smith [2003] proposent que les dents soient apparues au moins deux fois indépendamment au cours de l'évolution.

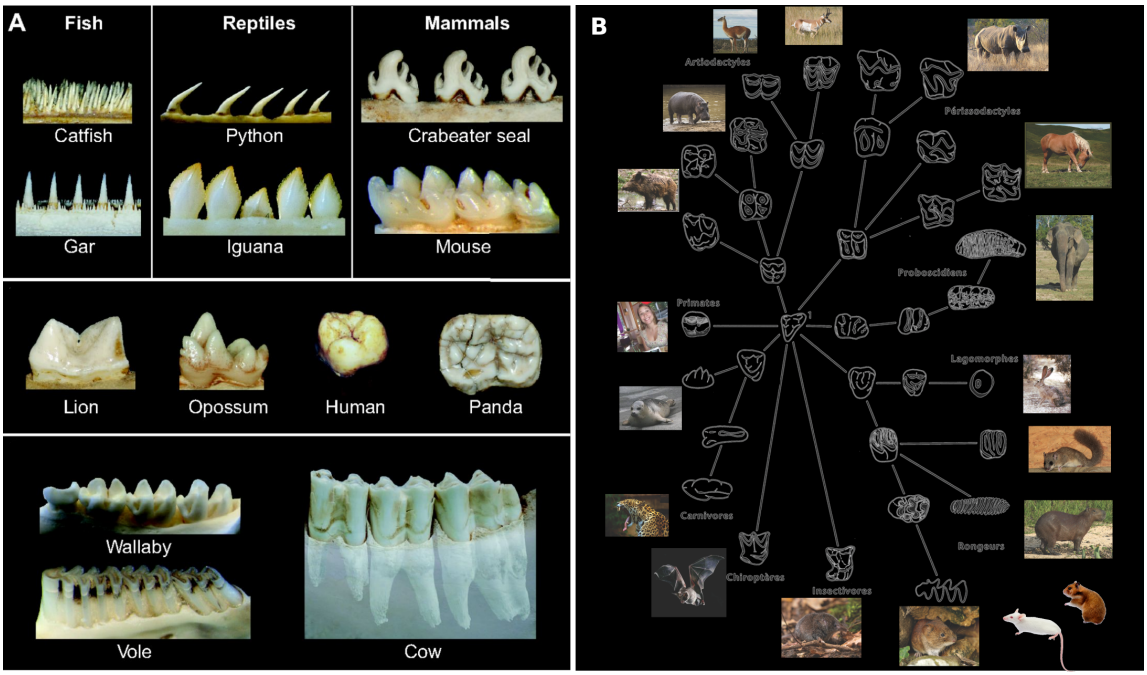



FIGURE 3.2 – **Diversité évolutive de la forme des dents dans des espèces actuelles**
**A.** Diversité des dents chez les vertébrés. (Tiré de Jernvall & Thesleff [2012]) **B.** Diversité morphologique des molaires chez les mammifères.

#### 3.1.2.1 La morphologie des dents

Comme discuté plus haut (partie 3.1.2), les dents "vraies" sont présentes chez les vertébrés à mâchoires. Les dents sont soient présentes dans la cavité orale attachées à la mâchoire, réparties sur une ou plusieurs rangées, ou/et elles sont présentes dans la cavité oro-pharyngienne (figure 3.3B). La forme ancestrale des dents est une forme conique, tels les dents du gar, du poisson chat ou du python que l'on peut voir sur la figure 3.2. De plus, une caractéristique ancestrale des dents est un remplacement continu (comme chez les requins par exemple). Les ornementations ainsi que l'augmentation du nombre de cuspides sont des caractéristiques dérivées des dents. Cette complexification de la forme des dents est souvent associée à une perte du remplacement continu (sauf quelques exceptions [Rodrigues *et al.* , 2011]). La complexification de la forme des dents est souvent associée à une réduction du nombre de dents qui constitue la dentition [Stock *et al.* , 1997]. Enfin, la régionalisation ou hétérodontie [Butler, 1995] est aussi une des caractéristiques dérivées de la dentition. C'est à dire que la dentition est composées de dents qui ont des morphologies différentes et assurent des fonctions différentes au sein de la dentition. Cette hétérondontie est souvent couplée à une une augmentation du nombre de cuspides, et

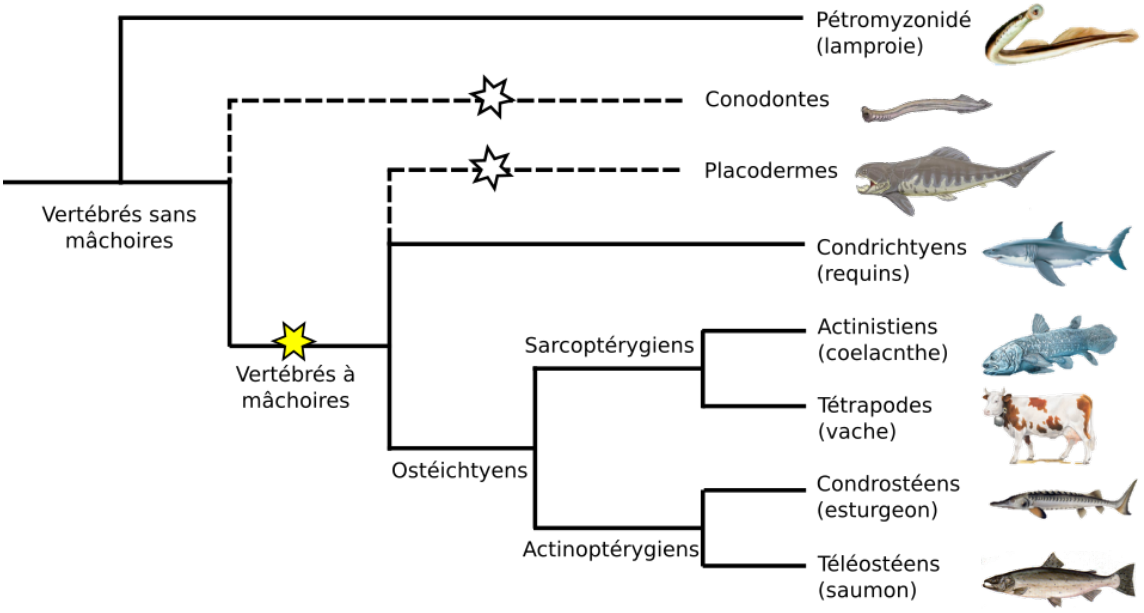



FIGURE 3.3 – **Arbre phylogénétique et apparition des dents**
Arbre phylogénétique des vertébrés à mâchoires. Les traits en pointillés représentent les groupes éteints. Étoile jaune : présence de dents "vraies". Étoile blanche : présence de structures semblables à des dents dont l'homologie est encore discutée [Burrow, 2003; Murdock *et al.* , 2013]. (Adapté de Panafieu & Gries [2007]; Fraser *et al.* [2009])

l'apparition de surfaces de broyages (sur les molaires par exemple) permettent une assimilation plus rapide et plus efficace de la nourriture [Miletich *et al.* , 2005]. Par la suite je vais me concentrer sur une dent en particulier, la molaire, qui est le modèle d'étude de cette thèse.

### 3.1.3 L'évolution des molaires

Les molaires sont une innovation morphologique des mammifères [Tims, 1903; Luo, 2007]. Ce sont des dents complexes qui sont capables de rentrer en occlusion, avec la molaire de la mâchoire opposée, permettant ainsi le broyage des aliments [Reisz, 2006]. La figure 3.2B montre la grande diversité des molaires chez les mammifères actuels, dont on pense qu'une telle diversité est due à de nombreuses radiations évolutives principalement liées au régime alimentaire. Ces radiations sont bien illustrées dans le registre fossile, où des nouveaux plans d'organisation des molaires sont fréquemment associés à d'importantes diversifications des lignées. Dans ce paragraphe je vais m'intéresser à l'évolution des molaires, de leurs origines jusqu'à arriver à mon modèle d'étude : la molaire de rongeur.

#### 3.1.3.1 La molaire tribosphénique

Au Trias (> 200 MYA), chez les reptiles mammaliens, les dents étaient simplement des dents coniques, à une seule cuspide donc. Chez les mammaliformes, des mammifères archaïques, la dent triconodonte [Osborn, 1888] fait son apparition, il s'agit d'une dent à trois cuspides alignées sur l'axe antéro-postérieur formant une seule et même rangée, elle exploite donc un espace à deux dimensions. À la fin du Jurassique / début du Crétacé il y a environ 160-200 millions d'années, apparait la molaire tribosphénique [Butler, 1990] (figure 3.1A) chez les premiers mammifères. En plus de l'axe antéro-postérieur, cette dent exploite l'espace bucco-lingual. Ainsi au lieu d'être réparties sur une seule ligne, les cuspides sont réparties sur 2 lignes. Cette exploitation de la troisième dimension de l'espace offre une surface pour broyer les aliments [Schultz & Martin, 2014] et l'occlusion entre les dents inférieures et supérieures offre surement un avantage à la prise alimentaire. À partir de cette molaire tribosphénique, les mammifères ont évolué une grande diversité morphologique de molaires dont un aperçu est représenté figure 3.2B. On pense que cette innovation a fait le succès évolutif des mammifères.

#### 3.1.3.2 Le plan murin

Une quatrième cuspide ajoutée à cette molaire tribosphénique donne naissance à la molaire quadrituberculaire [Hunter & Jernvall, 1995], qui est un plan dentaire relativement commun (c'est celui de l'homme par exemple) et c'est aussi le plan ancestral des rongeurs. Les rongeurs étudiés dans cette thèse, la souris et le hamster, sont des Muridés. La morphologie ancestrale des molaires de Muridés est caractérisée par l'ajout de 1 à 2 cuspides supplémentaires à l'avant de la dent. Le plan ancestral des Muridés se nomme plan cricétin [Lazzari, 2006], il est représenté sur la figure 3.4, et se caractérise entre autre par la présence de 6 cuspides. Par la suite, je qualifierai d'ancestrale toute molaire ayant 6 cuspides. Chez le hamster, le deuxième rongeur étudié dans ce manuscrit, le plan dentaire est un plan cricétin dérivé (figure 3.4). Le plan dentaire des Murinés (souris et rats) se nomme plan murin, il est caractérisé par une configuration différente des crêtes (les connexions entre les cuspides) pour la molaire inférieur, ces changements de connexions sont associés à un nombre plus élevé de cuspides pour la molaire supérieure [Lazzari, 2006]. La souris présente la particularité de posséder un nombre de cuspides différents entre la molaire supérieure et inférieure. Ainsi, la molaire inférieure conserve un nombre de cuspide ancestral (c=6) réparties sur 2 rangées, alors que la molaire supérieure développe 2 cuspides supplémentaires sur une troisième rangée linguale (figure 3.4). Encore une fois, notez que c'est l'espace linguale qui est exploité pour mettre en place ces cuspides supplémentaires.

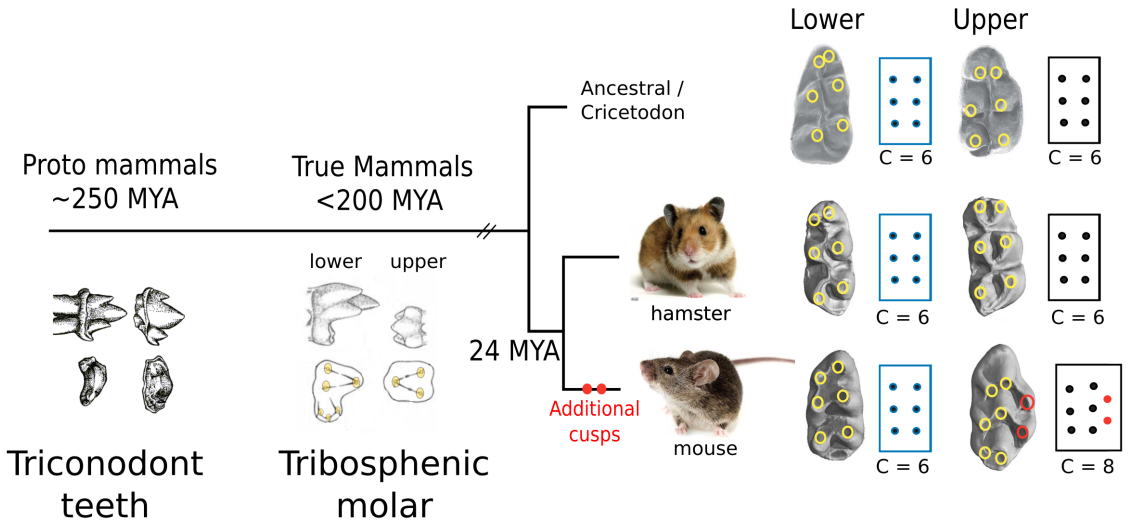



FIGURE 3.4 – **Évolution morphologique des molaires supérieures et inférieures**.
Les molaires supérieure et inférieure ont des morphologies extrêmement différentes depuis leurs origines dans les premiers mammifères il y a ~200 millions d'années (200 MYA). Dans la lignée de la **souris**, la molaire supérieure est drastiquement différente avec l'acquisition de 2 cuspides supplémentaires sur une troisième rangée (cuspides représentées en rouge). Alors que la molaire inférieure possède un nombre ancestral de cuspides (C=6). Le **hamster** conserve le même nombre de cuspides sur ces molaires inférieure et supérieure. C : nombre de cuspides

## 3.2 Le développement des dents

### 3.2.1 Les dents sont des organes sériels dont l'identité est définie par un code homéotique

#### 3.2.1.1 Les dents sont des organes sériels

La dentition est composée de structures répétées : les dents. Ainsi, la dentition peut être décrite comme un organe segmenté où chaque dents sont des homologues sériels. Du point de vue des processus développementaux, la dent peut être vue comme un module [Stock, 2001]. Ainsi le module "dent" et les réseaux génétiques associés à ce module vont être (ré)utilisés à chaque fois qu'une dent de la dentition sera mise en place [Jernvall & Thesleff, 2000, 2012]. Cependant, la place de chaque dent dans la mâchoire est spécifique, ainsi chaque module sera initié avec des paramètres différents, dépendant de "l'environnement de la mâchoire", c'est-à-dire sa place dans la mâchoire, la présence d'autres structures dans le voisinage, et procédera de manière spécifique à son "environnement" (voir ci-dessous le paragraphe sur l'identité des dents et le code homéotique). De la sorte, bien qu'ils partagent une homologie "identité dent", les modules *dent* de la dentition seront partiellement découplés les uns des autres et suivront une évolution relativement indépendante, avec la contrainte forte d'aboutir à la fin du développement à

une dentition fonctionnelle, au sein de la mâchoire et entre mâchoire supérieure et mâchoire inférieure (occlusion). Cette propriété va donc pouvoir impacter l'évolution du programme développemental des dents. Les réseaux de gènes qui mettent en place les cuspides au sein d'une dent sont communs [Keränen *et al.* , 1998; Jernvall & Thesleff, 2000], on peut alors parler d'un module "cuspide" qui est réutilisé pendant la morphogenèse de la couronne lors du développement séquentiel des cuspides. Ainsi, les cuspides d'une même dent peuvent être considérées comme des homologues sériels.

#### 3.2.1.2 Identité des dents et code homéotique

La mâchoire n'exprime pas de gènes Hox, mais elle exprime des gènes à homéoboite, tels les *Dlx, Msx, Lhx*, qui vont spécifier l'identité de la mâchoire inférieure (mandibule) et supérieure (maxillaire). Par exemple, les gènes *Dlx5/6* sont nécessaires pour spécifier l'identité de la mâchoire inférieure, s'ils ne sont pas présents, la mâchoire inférieure se développe avec une identité de mâchoire supérieure [Depew, 2002; Beverdam *et al.* , 2002]. Plus d'information sur le développement craniofacial sont accessibles dans la revue de Minoux & Rijli [2010]. L'identité des dents composant la dentition est définie en ré-utilisant ces gènes à homéoboite. La co-expression spécifique de certains de ces gènes va permettre de régionaliser la mâchoire et de définir des dents avec des identités différentes, c'est ce que l'on appelle le code homéotique odontogénique (revue dans McCollum & Sharpe [2001]; Tucker & Sharpe [2004]; Catón & Tucker [2009]). Par exemple, le gène à homéoboite *Barx1*, exprimé dans le mésenchyme à ED10 chez la souris, est responsable de l'identité spécifique des molaires [Tucker, 1998; Miletich *et al.* , 2005]. Les *Dlx* sont aussi essentiels pour spécifier l'identité des molaires. Si *Dlx1* et *Dlx2* sont mutés en même temps chez la souris, alors la molaire supérieure ne se développera pas (arrêt au stade placode), alors que les incisives et la molaire inférieure se développent normalement [Thomas *et al.* , 1997]. Le développement normal de la molaire inférieure est probablement possible en raison de la compensation par d'autres membres de la famille *Dlx* (comme *Dlx5/6*) qui sont exprimés dans la mâchoire inférieure, mais pas dans la mâchoire supérieure [Qiu *et al.* , 1997; Thomas *et al.* , 1997].

### 3.2.2 Généralités sur le développement de la molaire de souris

La dent est un appendice ectodermique comme les plumes, les ongles/griffes, les glandes mammaires, les poils [Pispa & Thesleff, 2003]. Le programme de développement de la dent partage d'ailleurs de nombreuses voies de signalisation ainsi que de nombreux processus de morphogenèse avec les autres appendices ectodermiques . Notamment l'initiation de la morphogenèse qui passe par un stade placode où il y a un épaississement de l'épithélium, ainsi que le dialogue constant entre l'épithélium et le mésenchyme tout au long du développement. La souris possède une dentition assez simple, ainsi sur chaque quadrant de mâchoire on retrouve une incisive à croissance continue, ainsi que trois molaires. Le diastème, qui est un espace sans dent de la mâchoire, sépare l'incisive des molaires.

La première molaire de souris est un modèle majeur du développement des dents et son développement est relativement bien compris, comme le montrent les revues de Tucker & Sharpe [2004]; Jernvall & Thesleff [2012]. Le développement de la dent est extrêmement dynamique et implique un dialogue constant entre le l'écto-mésenchyme dérivé des cellules des crêtes neurales (que j'appellerai simplement mésenchyme pour le reste de ce manuscrit), et l'épithélium. Le potentiel inducteur dentaire bascule entre l'épithélium et le mésenchyme tout au long du développement de la molaire [Bei *et al.* , 2000]. Le signal d'initiation de développement du bourgeon dentaire, qui a lieu à 10 jours de développement (Embryonic Day 10 ou ED10) est porté par l'épithélium [Mina & Kollar, 1987], il permet l'induction du mésenchyme et la formation de la placode épithéliale [Jussila & Thesleff, 2012]. Cette placode est le premier signe morphologique du développement du bourgeon dentaire, elle apparait à ED11.5. À ED12, le potentiel odontogénique bascule dans le mésenchyme [Mina & Kollar, 1987; Jussila & Thesleff, 2012]. À ED13.5, l'épithélium s'invagine dans le mésenchyme sous-jacent pour former le stade "bourgeon" (ou bud) et le mésenchyme se condense autour de ce bourgeon épithélial. Une demie journée plus tard, à ED14.0, se forme le nœud d'émail primaire (PEK, primary Enamel Knot) [Jernvall *et al.* , 1994, 1998], qui est un centre de signalisation transitoire mis en place au niveau de l'épithélium. Le PEK favoriserait la prolifération de l'épithélium qui serait contraint de s'enrouler autour du mésenchyme sous-jacent, formant ainsi la couronne de la dent. Je reviendrai sur le devenir du pEK dans ma partie résultats dans le chapitre 10. Après la mise en place de ce PEK, débute la morphogenèse de la couronne avec la formation séquentielle des cuspides, c'est cette période du développement qui a été étudiée dans ce manuscrit. À partir du stade "coiffe" tardif (cap stage), autour de ED14.5-ED15, le nœud d'émail primaire (PEK) est remplacé par des nœuds d'émail secondaires (SEKs, secondary Enamel Knots) qui se forment de façon séquentielle au sein de la couronne (illustré figure 3.5). Les sEKs, contrôlent la deuxième phase de prolifération et d'invagination de l'épithélium dans le mésenchyme dentaire condensé et sont les centres de signalisation responsables de la formation des cuspides. Au stade ED18.5, les cellules à l'interface entre l'épithélium et le mésenchyme (IEE, inner enamel epithelium) commencent à se différencier en cellules sécrétrices de dentine ou d'émail fixant ainsi la forme adoptée par le mésenchyme (et induite par les sEK) ou cours la morphogenèse de la couronne. Enfin, les racines se forment plus tardivement, au moment où la dent érupte.

### 3.2.3 Le nœud d'émail : centre de signalisation responsable de la formation des cuspides

Le nœud d'émail, ou enamel knot (EK) est un centre de signalisation épithélial composé d'un groupe compact de cellules épithéliales qui ne se divisent plus [Jernvall *et al.* , 1994]. Le nœud d'émail primaire ou PEK (primary enamel knot) se met en place au sommet du bourgeon au stade ED14 (figure 3.5). La formation du PEK au sommet du bourgeon marque le site d'initiation du repliement de l'épithélium autour du mésenchyme sous-jacent qui se condense. Les cellules du nœud d'émail ne prolifèrent pas alors que la croissance de l'épithélium adjacent est stimulée,

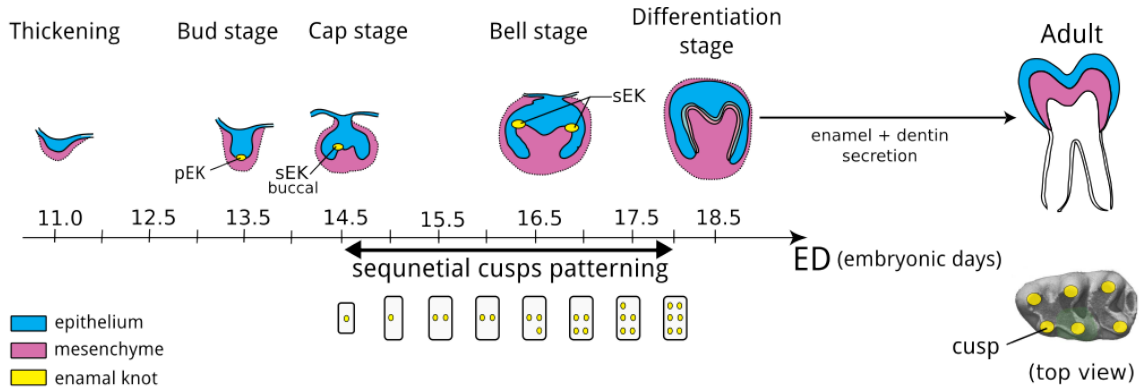



FIGURE 3.5 – **Développement de la première molaire de souris**.
Représentation schématique du développement de la première molaire de souris. À 11 jours de développement (ED11), épaississement de l'épithélium. À ED13.5, le stade bourgeon marque l'invagination de l'épithélium dans le mésenchyme condensé sous-jacent. Au sommet du bourgeon en formation, le pEK (nœud d'émail primaire), qui est un centre de signalisation, se forme. À partir du stade "coiffe" (ED14.5) et jusqu'au stade cloche (ED16.5-ED18), formation séquentielle des sEKs qui sont les centres de signalisation responsables du développement des cuspides. sEK : nœud d'émail secondaire.

ce qui entraîne la formation de boucles cervicales entourant le mésenchyme (figure 3.7). L'arrêt de la prolifération des cellules composant le nœud d'émail (enamel knot ou EK) est causée par l'expression de l'inhibiteur de kinase cycline-dépendant *p21* (*Cdkn1*) qui est induit par l'expression de *Bmp4* dans le mésenchyme sous-jacent [Jernvall *et al.* , 1998] au stade ED14. À partir du stade ED14.5, *Bmp4* et *Shh* sont aussi retrouvés exprimés dans le nœud d'émail [Åberg *et al.* , 1997; Jernvall *et al.* , 1998; Vaahtokari *et al.* , 1996]. Si le PEK ne se forme pas le germe ne peut pas effectuer sa transition au stade coiffe, qui est morphologiquement caractérisé par le début de la croissance des boucles cervicales. Ainsi, la mise en place du PEK marque le début de la cuspidogenèse. Une autre caractéristique du nœud d'émail est la présence d'apoptose [Vaahtokari *et al.* , 1996; Lesot *et al.* , 1996; Viriot *et al.* , 1997; Jernvall *et al.* , 1998; Coin *et al.* , 1999] qui jouerait un rôle dans la mise en place de la morphologie de la couronne. Il a été proposé que le PEK serait résorbé, au moins en partie, par apoptose avant la mise en place du premier SEK [Jernvall *et al.* , 1998; Matalova *et al.* , 2005, 2006]. Cependant, lorsque l'apoptose est inhibée, aucun défaut de morphologie à l'âge adulte n'a été observé [Coin *et al.* , 2000; Matalova *et al.* , 2006; Setkova *et al.* , 2007]. Matalova *et al.* [2006] notent néanmoins des défauts de repliement de l'épithélium au cours du développement des souris hétérozygotes et homozygotes mutantes pour le gène *caspase-3*.

Par la suite, des nœuds d'émail secondaires ou SEKs pour secondary enamel knots sont formés séquentiellement et se mettent en place au sommet des futurs cuspides. Ces nœuds d'émail secondaires vont stimuler la croissance des boucles épithéliales, notamment par l'expression de *Fgf4* [Jernvall *et al.* , 1994; Thesleff *et al.* , 2001], ce qui va permettre de former des vallées. De

plus, les SEKs vont permettre la prolifération des cellules sous-jacente ainsi que la maturation des cuspides [Thesleff *et al.* , 2001].

### 3.2.4 Un modèle *in silico* du développement de la dent

Salazar-Ciudad & Jernvall [2002, 2010] développent un modèle computationnel du développement de la dent capable de reproduire les morphologies dentaires de plusieurs mammifères (souris, campagnol, phoque), mais aussi des morphologies ancestrales telles que la dent triconodonte et la molaire tribosphénique [Osborn, 1888; Butler, 1990] (voir la partie sur l'évolution des molaires *(p. 38)* ). Ce modèle combine des données de la littérature concernant les interactions génétiques et la croissance du bourgeon, pour proposer un modèle morphodynamique dans lequel la morphologie du bourgeon dentaire joue un rôle à part entière dans la mise en place du patron de cuspides. De plus, les auteurs montrent que des morphologies similaires, ici l'obtention d'une dent quadrituberculaire, peuvent être atteintes avec des paramètres différents [Salazar-Ciudad & Jernvall, 2002], ce qui sous entend d'un point de vue de l'évolution des programmes de développement qu'il y a plusieurs programmes développementaux qui permettent d'arriver à une morphologie finale identique.

Je ne vais pas décrire ici le support mathématique du modèle, mais ils sont disponibles dans la revue suivante : Salazar-Ciudad [2008]. Je vais simplement détailler les lignes générales du modèle. Le point de départ du modèle comprend une couche de cellules épithéliales identiques situées au dessus d'une couche de cellules mésenchymateuses identiques. Ces cellules vont répondre à deux molécules diffusibles, un activateur et un inhibiteur qui vont affecter la croissance du bourgeon de façon différentielle. Ainsi l'activateur, qui est un indicateur du nœud d'émail (EK, qui contrôle la mise en place des cuspides), va induire la différenciation cellulaire. Alors que l'inhibiteur, réprime la différenciation induite par le nœud d'émail et promeut la croissance du bourgeon. En plus des paramètres décrivant ces deux molécules (*eg.* vitesse de diffusion, intensité de l'inhibition, concentration des molécules) , il faut ajouter des paramètres décrivant la croissance asymétrique le long des axes antéro-postérieur et bucco-lingual.

Pour illustrer le concept de développement morphodynamique, le modèle montre que la formation d'un nouveau nœud d'émail au temps $t+1$ dépend de la morphologie au temps $t$ car les changements de la géométrie du tissu vont impacter la portée d'activité de l'activateur et de l'inhibiteur. Ainsi la distribution spatiale des nœuds d'émail est affectée par la forme du bourgeon dentaire et à son tour, la forme du bourgeon va être modifiée par la distribution spatiale des nœuds.

Cette propriété morphodynamique du développement de la dent [Salazar-Ciudad, 2012] rend plus complexe la carte génotype-phénotype que sont les liens entre les variations génétiques et les variations phénotypiques. Dans le cas du développement de la dent, il ne s'agit pas de prendre en compte uniquement l'expression (ou non) des gènes, il faut aussi prendre en compte la forme

du bourgeon dentaire qui influence la mise en en place des cuspides, elle même influencée par les cuspides présentes sur le bourgeon.

## 3.3 Les voies de signalisation majeures du développement de la molaire

Les études de cette thèse portent sur la caractérisation du développement des germes molaires inférieures et supérieures chez la souris afin d'identifier les différences entre les molaires qui permettent de mettre en place deux cuspides linguales supplémentaires sur la molaire supérieure. Comme mentionné plus haut, les premières molaires de souris ont un développement asymétrique de leurs boucles cervicales, cependant cette asymétrie se résorbe dans la molaire inférieure alors qu'elle perdure dans la molaire supérieure. C'est pourquoi je vais m'intéresser dans cette partie à introduire l'initiation et la croissance des boucles cervicales. Pour cela, je vais me concentrer sur les voies de signalisation majeures au cours de la morphogenèse de la couronne : la transition du stade bourgeon au stade coiffe car c'est à ce stade que va s'initier la formation et la croissance des boucles cervicales ainsi que le formation des nœuds d'émail (le PEK et les SEKs) qui sont essentiels à la formation des cuspides. Je vais découper cette partie en quatre paragraphes qui vont traiter des quatre voies majoritaires que sont la voie Hh, BMPs, Wnt et FGFs. Je commencerai d'abord par décrire rapidement les acteurs de la voie de signalisation puis je décrirai son rôle lors de la morphogenèse de la couronne.

### 3.3.1 La voie de signalisation Hedgehog est responsable de la polarisation du germe dans l'axe bucco-lingual

Dans les cellules réceptrices, la protéine SHH sécrétée se lie au récepteur Patched1 (PTCH1), activant ainsi Smoothened (SMO), une protéine transmembranaire inhibée par PTCH1 en l'absence de SHH. SMO à son tour empêche le traitement protéolytique du facteur de transcription GLI3 dans sa forme répressive. Ceci favorise la génération des activateurs GLI (1-3) qui régulent la transcription des gènes sous le contrôle de la voie Hedgehog, qui comprend *Ptch1* et *Gli1* [Chen *et al.* , 2007; Ingham *et al.* , 2011]. Le récepteur *Ptch1*, l'activateur *Smo* et l'effecteur transcriptionnel *Gli1* sont tous les trois localisés au niveau du cil primaire, ce qui suggère que la voie Hh est active dans les cils et que les cils sont essentiels à la voie de signalisation Hh et à la transduction du signal [Haycraft *et al.* , 2005; Rohatgi *et al.* , 2007; Sasai & Briscoe, 2012]. Ainsi, Caparros-Martin *et al.* [2012] montrent que *Evc* régule le trafic de la forme active de *Gil3* dans le cil primaire permettant ainsi de moduler la transduction du signal par la voie Hh. De plus, Ohazama *et al.* [2009] montrent qu'une perturbation du bon maintien des cils, par exemple du transporteur intraflagellaire de cils *polaris*, conduit à une plus large expression de *Shh* dans l'épithélium dentaire de souris, causant le développement de dents supplémentaires dans le diastème.

En l'absence d'expression de *Shh* dans l'épithélium, Dassule *et al.* [2000] observent un sous développement (hypoplasie) de la boucle cervicale linguale. Ceci pourrait être dû à un défaut de prolifération du côté lingual, cependant l'analyse des coupes histologiques suggèrent que le côté lingual soit resté dans un état non différencié, car on ne distingue pas la structuration caractéristique de l'épithélium. Réciproquement, Nakatomi *et al.* [2013] observent que les souris homozygotes mutantes pour *Evc*, qui régule la disponibilité en GLI3 [Caparros-Martin *et al.* , 2012] et donc sur active la voie Hh là où elle est déjà activée, présentent une anomalie de développement des boucles cervicales. Ainsi la boucle cervicale buccale est sous développée (ou hypoplasique) alors que la boucle cervicale linguale est fortement développée (ou hyperplasique). Dans le mutant *Evc* l'expression de *Shh* est normale, mais l'expression des gènes cibles de la voie de signalisation Hh, tel *Ptch1* et *Gli1*, est affectée. De plus, dans les mutants *Evc*, les auteurs observent que l'expression de *Shh, Wnt10b* et *Lef1* est étendue buccalement,ce qui suggère que la voie Hh pourrait réguler la voie Wnt.

Ces résultats suggèrent que la voie Hh joue un rôle dans la polarisation du germe dans l'axe bucco-lingual ainsi que dans la croissance de boucles cervicales. De plus, Li *et al.* [2015] montrent que la voie Hh permet le maintien des cellules souches présentes dans les boucles cervicales.

### 3.3.2 La voie de signalisation des BMPs, une voie centrale pour la transition au stade coiffe

Les BMPs, Bone Morphogenic Proteins, sont des morphogènes qui participent à la morphogenèse du tissu au cours du développement. Un dimère de BMP va recruter et se lier à un récepteur sérine-thréonine kinase de type I ou de type II (BMPR-I/-II). Cette liaison avec le récepteur va provoquer la phosporylation des protéines Smads (SMAD1/5/8), ces Smads phosphorylées vont pouvoir interagir avec SMAD4 est être transloquées dans le noyaux, le complexe va aloir pouvoir se lier à l'ADN et recruter des co-répresseurs et des co-activateurs pour réguler la transcription de gènes cibles (revue dans Dutko & Mullins [2011]).

Åberg *et al.* [1997] caractérisent le profils d'expression des BMPs (*Bmp2* à *Bmp7*) à différents stades de développement de la molaires et Nadiri *et al.* [2006] répertorient l'expression de leurs récepteurs (BMPRs). Je vais plus particulièrement détailler l'expression de *Bmp4* car au cours du développement de la molaire, le patron d'expression de *Bmp4* suit et coïncide avec le potentiel odontogénique ce qui suggère l'importance de son rôle au cours du développement du germe dentaire [Vainio *et al.* , 1993; Jernvall *et al.* , 1998; Jernvall & Thesleff, 2000].

D'abord, *Bmp4* est exprimé dans l'épithélium au cours de l'initiation, puis son expression bascule dans le mésenchyme où il va participer à l'induction du PEK dans l'épithélium, puis au stade coiffe (ED14.5) il est exprimé dans le mésenchyme condensé et dans le PEK. BMP4 est un des quatre acteurs, avec MSX1 [Vainio *et al.* , 1993] (dont il induit l'expression) et PAX9 [Peters *et al.* , 1998] et *Runx2* [Wang *et al.* , 2005] jouant un rôle prépondérant dans la transition du stade

bourgeon au stade coiffe (figure 3.6).

Ainsi, dans un germe mutant $Msx1^{-/-}$, il a été montré que l'expression de *Bmp4* dans le mésenchyme est fortement réduite et que la simple addition de *Bmp4* sur ce germe mutant en culture permet la reprise du développement jusqu'au stade cloche tardif [Chen *et al.* , 1996; Bei *et al.* , 2000]. Ces données suggèrent que l'expression de BMP4 dans le mésenchyme, qui est biaisée buccalement, dont l'expression est auto-induite de façon MSX1-dépendante [Chen *et al.* , 1996], permet la transition du stade bourgeon au stade coiffe [Bei *et al.* , 2000; Miletich *et al.* , 2005; O'Connell *et al.* , 2012]. Par ailleurs, il a été précisé que la formation d'un complexe protéique entre PAX9 et MSX1 est nécessaire pour permettre l'expression de *Bmp4* dans le mésenchyme [Ogawa *et al.* , 2006]. Ainsi, Jia *et al.* [2012] montrent qu'en l'absence d'expression de *Bmp4* dans le mésenchyme, la molaire inférieure est stoppée au stade bourgeon. La perte d'expression de *Bmp4* dans l'épithélium des germes de molaires inférieures de souris mutantes $Bmp4^{ncko/ncko}$ suggèrent un défaut de mise en place du PEK. Jia *et al.* [2012] observent que dans des souris dépourvues de *Bmp4* dans le mésenchyme et d'une copie du facteur de transcription *Osr2* l'arrêt au stade bourgeon est levé et la molaire inférieure se développe jusqu'à l'éruption mais la taille et la morphologie sont fortement impactées. En revanche dans ce même mutant $Bmp4^{ncko/ncko}$ le développement de la première molaire de souris n'est stoppé [Jia *et al.* , 2012], ce qui suggère une différence de régulation des voies de signalisation dans les molaires supérieures et inférieures. D'autre part, Jia *et al.* [2016] révèlent que cette différence est probablement due à une réduction de l'activité de la voie de signalisation Wnt dans la première molaire inférieure causée par la plus forte expression de l'inhibiteur *Dkk2* dans le mésenchyme. Je reviendrai sur le résultat concernant Wnt ci-dessous dans le paragraphe réservé à cette voie.

Enfin, la voie des BMPs joue aussi un rôle dans le maintien des boucles cervicales ainsi que la croissance des boucles cervicales. D'une part, Li *et al.* [2015] montrent que la population de cellules souches présente à l'extrémité des boucles cervicales, qui expriment *Sox2*, est maintenue par le réseau de signalisation BMP-Smad4-Shh-Gli1. Les auteurs montrent que la voie BMP-Smad4 permet l'activation de la voie Hh (via SHH) qui permet l'expression de *Gli1* qui lui même activerait l'expression de *Sox2*. D'autre part, comme revue par Rider & Mulloy [2010], la follistatine (*Fst*) antagonise l'action de la voie des BMPs en se liant directement aux BMPs (par exemple à BMP7 dans le muscle [Amthor *et al.* , 2002]) ou en se liant à l'*activine* pour laquelle il a une forte affinité [Thompson *et al.* , 2005]. Les mutants *Fst* (Follistatine) présentent une morphologie aberrante de la boucle cervicale linguale à partir du stade ED16 ainsi que des défauts de repliement de l'épithélium [Wang *et al.* , 2004b]. Ces données suggèrent la voie BMPs, régulé par la follistatine, régule la croissance des boucles épithéliales inter-cuspales.

### 3.3.3 La voie de signalisation Wnt régule le patterning du germe

L'activation de la voie Wnt passe par la liaison de Wnt à l'un son récepteur transmembranaire Frizzled (Fz). La voie la plus connue et la mieux détaillée est la voie dite "canonique", elle implique

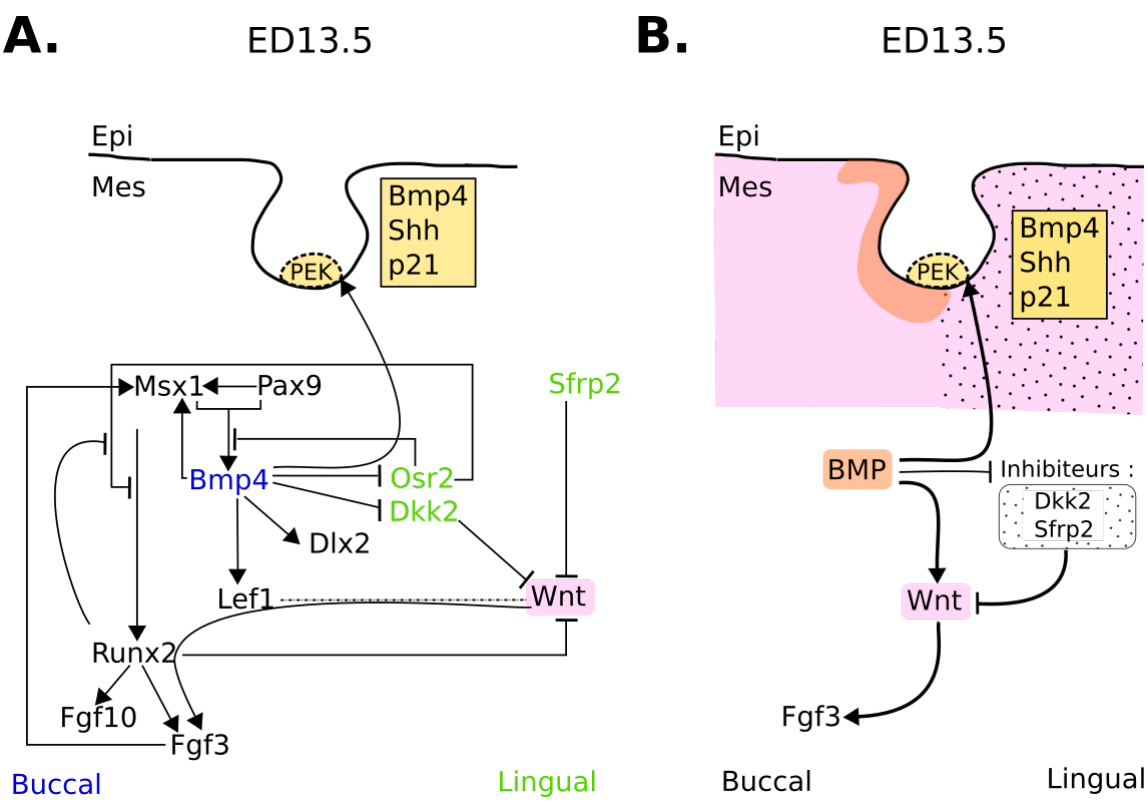



FIGURE 3.6 – **Transition du stade bourgeon au stade coiffe**.
**A.** Réseau d'interactions génétique qui sous tend la transition du stade bourgeon au stade coiffe qui est une étape importante dans le développement du germe de molaire, marquant le début de la morphogenèse de la couronne. L'expression des facteurs de transcriptions ***Msx1*** [Vainio *et al.* , 1993], ***Pax9*** [Peters *et al.* , 1998] (contrôlé par FGFs [Mandler & Neubüser, 2001]) et ***Runx2*** [Wang *et al.* , 2005; Kwon *et al.* , 2015] sont nécessaires pour que cette transition ait lieu. *Runx2* active la voie FGFs (*Fgf3, 10*), probablement de concert avec la voie Wnt [Chen *et al.* , 2009], et pourrait inhiber la voie Wnt (sur-expression de *Wnt5a* dans une mutant $Runx2^{-/-}$) [Åberg *et al.* , 2004b]. ***Bmp4*** est le quatrième acteur nécessaire à cette transition coiffe [Chen *et al.* , 1996; Bei *et al.* , 2000; Miletich *et al.* , 2005; O'Connell *et al.* , 2012]. La formation d'un complexe MSX1-PAX9 est nécessaire pour permettre l'expression de BMP4 dans le mésenchyme [Ogawa *et al.* , 2006]. BMP4 permet l'expression dans le mésenchyme de *Lef1*, *Msx1* et *Dlx2* [Chen *et al.* , 1996; Bei & Maas, 1998] alors qu'il inhibe l'expression de *Osr2* et *Dkk2* [Jia *et al.* , 2012, 2016]. Ce réseau de signalisation abouti à la formation du PEK (nœud d'émail primaire) dans l'épithélium, dans lequel sont exprimés *Bmp4*, *p21* et *Shh* [Jernvall *et al.* , 1998; Kwon *et al.* , 2015]. En bleu sont représentés les gènes dont l'expression est buccale et en vert ceux dont l'expression est linguale. **B.** Bilan des réseaux d'interactions nécessaires au bon déroulement de la transition coiffe. Les voies de signalisation BMP et Wnt sont nécessaires à cette transition. Cependant Wnt doit pouvoir être finement régulé spatialement (inhibé côté lingual) pour assurer le bon déroulement de la transition.

la stabilisation et la translocation de la $\beta$-caténine dans le noyau [Logan & Nusse, 2004; Freese *et al.* , 2010; Clevers & Nusse, 2012]. L'accumulation de $\beta$-caténine dans le noyau convertit les

protéines de la famille TCF/Lef se liant à l'ADN de répresseur en activateur de transcription (revue par Clevers & Nusse [2012]). En l'absence de signalisation Wnt la $\beta$-caténine est phosphorylée par la GSK3$\beta$ et est dégradée. LRP6 et LRP5 sont des co-récepteurs de Wnt dans la voie canonique. D'autre voies Wnt, dites voies non-canoniques ont aussi été mises en évidence (revues dans Semenov *et al.* [2007]; Sebbagh & Borg [2014]).

Le développement du germe dentaire est arrêté au stade bourgeon lorsque la *β-caténine*, donc la voie Wnt, est inactivée dans l'épithélium [Liu *et al.* , 2008] ou le mésenchyme [Chen *et al.* , 2009]. Ainsi, Chen *et al.* [2009] ont montré que la *β-catenine* est requise dans le mésenchyme afin de permettre l'expression mésenchymateuse de *Lef1* et *Fgf3*, de plus elle est nécessaire à la formation de nœud d'émail primaire (PEK). Kratochwil *et al.* [1996] montrent que la perte de l'expression *Lef1* dans le mésenchyme conduit à l'arrêt du développement du germe dentaire au stade bourgeon donc l'expression de *Lef1* est nécessaire pour l'induction du mésenchyme au stade ED13-ED14. Ce qui suggère que l'activation de la voie voie Wnt est nécessaire au développement de la molaire. Cependant, Järvinen *et al.* [2006] montrent que la sur-expression de la *β-caténine* dans les germes de molaires au stade bourgeon, ne permet pas à la transition bourgeon-coiffe d'avoir lieu et entraine la formation / l'initiation de germes multiples. Ces germes multiples vont par la suite se développer et former des dents avec une morphologie simple, bi- ou mono-cuspide. Ces données suggèrent que l'activation de la voie Wnt doit être finement régulée pour permettre le développement de la molaire.

Cette conclusion se précise par les résultats obtenus par Jia *et al.* [2016]. Dans un premier temps, Jia *et al.* [2012] observent qu'en l'absence d'expression de *Bmp4* dans le mésenchyme ($Bmp4^{ncko/ncko}$), la molaire inférieure est stoppée au stade bourgeon alors que la première molaire supérieure continue son développement. Dans un deuxième temps, les auteurs montrent que dans ces mutants $Bmp4^{ncko/ncko}$ la voie Wnt est plus fortement inhibée côté lingual, à cause de la plus forte expression de *Sfrp2* et *Dkk2*, que dans dans la souche sauvage [Jia *et al.* , 2016]. De plus, le sauvetage des germes de molaires inférieures dans les doubles mutants $Bmp4^{ncko/ncko}$ $Osr2^{+/-}$ serait le résultat de la levée de la forte inhibition de la voie Wnt côté lingual, pour revenir à un taux d'inhibition plus faible et plus proche de celui observé dans la souche sauvage, bien que la taille et la morphologie de la molaire inférieure soient impactées. Ces données suggèrent qu'en temps normal, la voie Wnt est activement réprimée côté lingual par la plus forte expression d'inhibiteurs de la voie Wnt. Si cette répression est perturbée côté lingual alors le développement du germe de molaire est affecté. Ceci suggère que la voie Wnt doit être finement régulée au cours du développement de la molaire, aussi bien quantitativement que spatialement.

### 3.3.4 La voie de signalisation des FGFs régule la croissance de l'épithélium

Les ligands FGFs, Fibroblast Growth Factors, sont au nombre de 18 chez les mammifères, auxquels viennent s'ajouter 4 gènes fortement conservés codant pour 4 récepteurs à tyrosine kinase (FGFR1-4) dont le nombre est accru par épissage alternatif [Turner & Grose, 2010; Ornitz & Itoh,

2015]. Après dimérisation des récepteurs liés au ligand, la transduction du signal s'effectue par une cascade de phosphorylation de plusieurs protéines accessoires conduisant à l'activation de l'une des voies suivantes : Ras-MAP Kinases, PI3K/AKT ou encore PLC$\gamma$ [Dailey *et al.* , 2005; Turner & Grose, 2010]. La réponse à l'activation de la voie des FGFs dépend du type cellulaire et/ou de l'état de maturation de la cellule, ainsi elle est peut être impliquée dans la prolifération, la différenciation, l'arrêt de la croissance, et l'apoptose. La voie FGFs est aussi connue pour son rôle dans la ramification des épithélia (branching morphogenesis) [Makarenkova *et al.* , 2009], aussi bien dans les poumons [Abler *et al.* , 2009; Goto *et al.* , 2014] que dans les reins [Qiao *et al.* , 2001; Bates, 2011]. Il a aussi montré l'existence d'un dialogue entre les voies FGFs et Wnt qui passe par la régulation de la $\beta$*-caténine* par FGF. Ainsi dans la dent, la signalisation FGFs, plus particulièrement *Fgf3*, inhibe la signalisation de la $\beta$*-caténine* intracellulaire en activant la voie PI3K/AKT pour réguler la localisation sub-cellulaire de la GSK3$\beta$ dans les cellules mésenchymateuses [Liu *et al.* , 2013]. Les auteurs de cette étude [Liu *et al.* , 2013] montrent que l'inhibition de la voie Wnt par la voie FGF (*Fgf3, 4, 9*), dans le mésenchyme permet de conserver le potentiel odontogénique du mésenchyme au stade ED13.5 afin qu'il ne participe pas au développement osseux qui a lieu tout autour de la dent.

Historiquement, la signalisation FGFs a été mise en évidence dans le développement de la dent par la présence de *Fgf4* dans le nœud d'émail, à la fois le PEK et les SEKs [Niswander & Martin, 1992; Jernvall *et al.* , 1994]. Exclusivement dans le PEK on retrouve *Fgf9* [Kettunen & Thesleff, 1998] et *Fgf15*. *Fgf20* est quant à lui exprimé aussi bien dans le PEK que les SEKs [Porntaveetus *et al.* , 2011]. *Fgf20* est un effecteur en aval de la voie de l'ectodysplasine (*Eda*), il participe au contrôle la taille du germe dentaire dans l'axe antéro-postérieur et bucco-lingual [Haara *et al.* , 2012]. *Fgf3* et *Fgf10* sont exprimés dans le mésenchyme à partir du stade bourgeon tardif. *Fgf3* induit l'expression du facteur de transcription *Msx1* dans le mésenchyme dans le bourgeon (stade ED13.5) qui est nécessaire pour la transition du stade bourgeon au stade coiffe [Bei & Maas, 1998]. En outre, *Fgf3* inhiberait la voit Wnt pour maintenir le potentiel odontogénique du mésenchyme [Liu *et al.* , 2013].

Enfin *Fgf3* joue un rôle dans la croissance du germe de première molaire. En effet, Charles *et al.* [2009a] montrent un effet dose de *Fgf3*, ainsi moins il y de *Fgf3* exprimé plus la dent est petite. De plus les auteurs observent aussi cette effet dose sur la mise en place des deux dernières cuspides à se former dans la première molaire supérieure, cette perturbation de la mise en place des cuspides pourrait être dû au fait que la taille du germe est réduite [Salazar-Ciudad & Jernvall, 2002]. Åberg *et al.* [2004b] montrent que *Runx2* régule l'expression de *Fgf3* dans le mésenchyme. Ainsi, ils observent que des billes de FGFs ne permettent pas l'expression de *Fgf3* dans le mutant *Runx2*, alors qu'elles y arrivent dans le sauvage donc ils concluent que *Runx2* est nécessaire pour la transduction de signal FGFs. Quant à FGF10, il stimule la prolifération cellulaire dans l'épithélium uniquement, il participerait ainsi à la prolifération des boucles cervicales [Kettunen *et al.* , 2000] et des cuspides [Pispa *et al.* , 1999].

La figure 3.7 résume de façon schématique les changements morphologiques pendant la morphogenèse de la couronne et reprend les actions principales des voies de signalisations décrites plus haut.

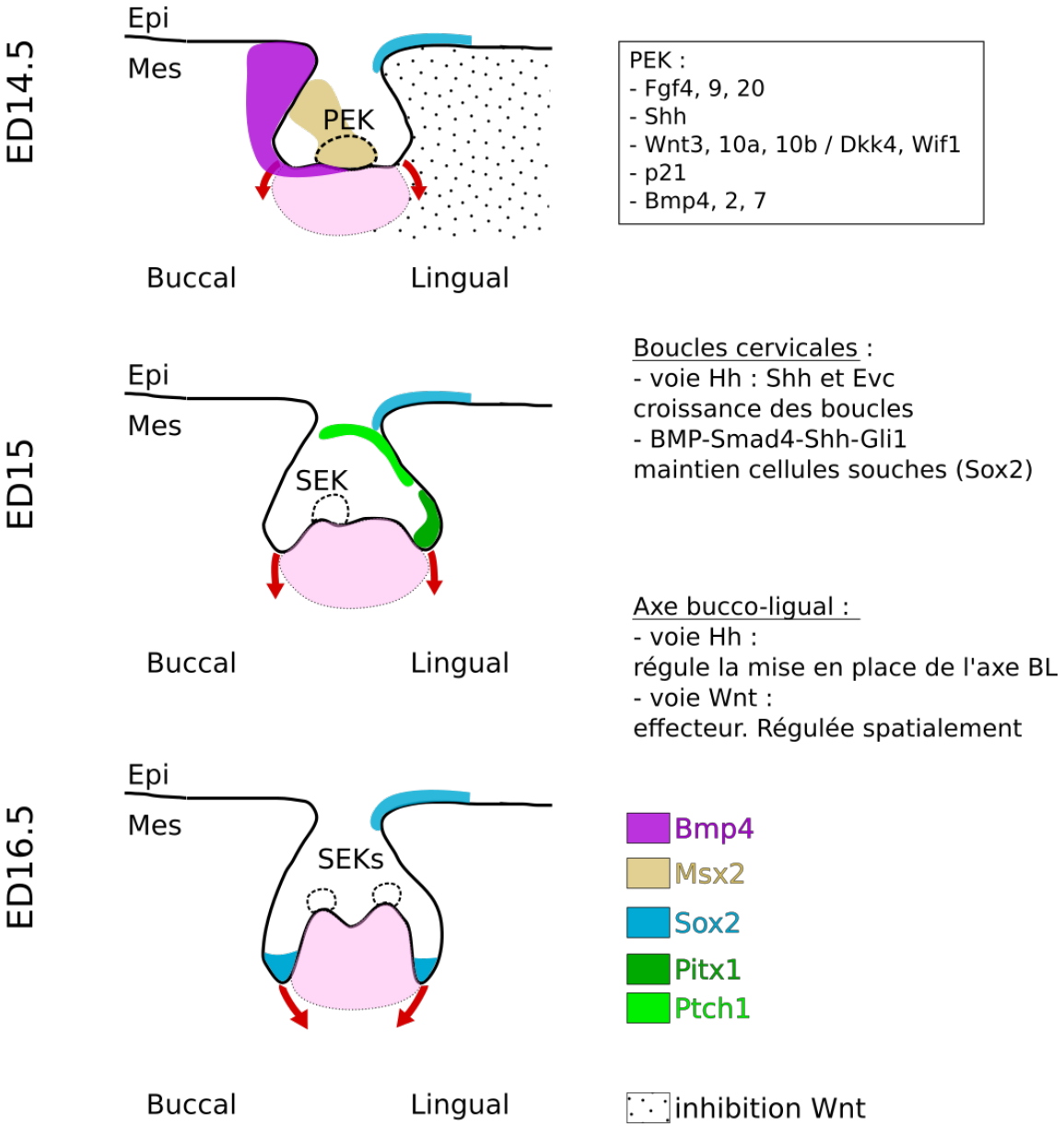



FIGURE 3.7 – **La morphogenèse de la molaire**.
Bilan des différentes voie de signalisation et de leur rôle au cours du développement du germe de molaire pendant la cuspidogenèse. Flèches rouges : croissance des boucles cervicales. Rose : mésenchyme condensé. PEK : nœud d'émail primaire ; SEKs : nœud(s) d'émail secondaire(s)

## 3.4 Comment augmenter le nombre de cuspides ?

### 3.4.1 Spécificités morphologiques des premières molaires de souris

Chez les mammifères, les molaires supérieures et inférieures ont des morphologies différentes. Cependant, dans le groupe des Murinés, groupe dont fait partie la souris, les premières molaires supérieure et inférieure présentent une divergence morphologique encore plus forte, avec un nombre de cuspides différents. La figure 3.4, représente la première molaire inférieure et supérieure de souris. La molaire inférieure de souris possède 6 cuspides majeures réparties sur deux rangées alors que la molaire supérieure possède 8 cuspides majeures, car elle développe une rangée de deux cuspides supplémentaire sur son côté lingual. Ces deux cuspides supplémentaires sont apparues récemment au cours de l'évolution, entre 18 MYA et 14 MYA. D'un point de vue développemental, ces deux cuspides supplémentaires sont les deux dernières cuspides à se former au cours de la morphogenèse de la couronne.

### 3.4.2 Modulation de la formation des cuspides au cours du développement de la première molaire de souris

Les molaires de mammifères montrent une tendance à la complexification de la morphologie de la couronne par une augmentation du nombre des cuspides. Comme le fil conducteur sous jacent des travaux présentés dans cette thèse est de comprendre comment la molaire supérieure de souris développe deux cuspides supplémentaires lingualement, je vais m'intéresser dans cette section aux études qui ont montré comment le nombre de cuspides peut être régulé.

D'une part, un certain nombre de gènes on été décrits comme pouvant moduler la formation des cuspides. Ainsi, Kassai *et al.* [2005] montrent qu'en l'absence d'*Ectodine* (un inhibiteur de *BMP*), le domaine d'expression de *p21*, exprimé exclusivement dans le nœud d'émail [Jernvall *et al.* , 1994, 1998], est élargi. Ce qui aboutit à une distance de formation entre cuspides qui est réduite. Les crêtes des cuspides buccales des molaires inférieures de souris dépourvues d'*Ectodine* sont fusionnées et parfois une dent supplémentaire se développe côté lingual. [Charles *et al.* , 2009a] montrent qu'une diminution de la dose de *Fgf3* réduit fortement la taille de ces cuspides linguales supplémentaires de la molaire supérieure et mime l'état ancestral. Cependant on ne peut pas faire un lien direct entre la diminution de la dose de *Fgf3* et la perte de ces cuspides. En effet, comme le montre le modèle *in silico* de Salazar-Ciudad & Jernvall [2002], la mise en place des cuspides est directement liée à la croissance du bourgeon. Comme les cuspides supplémentaires de la molaire supérieure de souris sont les dernières à se mettre en place, une croissance insuffisante du bourgeon peut expliquer leur échec à se développer. De ce fait, si l'on pousse ce raisonnement à l'extrême, tout gène affectant la croissance de la dent pourrait permettre d'augmenter la taille du bourgeon et permettre le développement de cuspides supplémentaires et ainsi avoir participé à cette innovation morphologique chez les Murinés. La voie de l'ectodysplasine, composée de trois acteurs principaux : le ligand *Eda*, le récepteur *Edar* et l'adaptateur *Edaradd* (revue dans

Sadier *et al.* [2014]) joue un rôle important en régulant le développement des cuspides [Tucker *et al.* , 2000; Charles *et al.* , 2009b]. *Edar* et *Edaradd* sont exprimés dans les nœuds d'émail [Tucker *et al.* , 2000; Headon *et al.* , 2001]. La perde de *Eda, Edar* et/ou *Edaradd* conduit à des défauts dans les nœuds d'émail. Notamment, en l'absence de l'expression de *Eda* dans les nœuds d'émail, une réduction de la taille de ces derniers est observée [Pispa *et al.* , 1999; Laurikkala *et al.* , 2001]. De plus, Tucker *et al.* [2004] montrent que le nombre de cuspides est corrélé à l'activité de *Edar*, qui active la voie NF-$\kappa$B. Ainsi, une sur-expression de la forme active du récepteur *Edar* conduit au développement de cuspides supplémentaires. Enfin, une étude basée sur des bourgeons de molaire inférieure de souris en culture met en évidence que le nombre de cuspides peut être augmenté en ajustant de multiples voies de signalisation en même temps (*eda*, *activin A* et *Shh*) [Harjunmaa *et al.* , 2012].

En résumé, l'augmentation du nombre de cuspides peut être obtenue en modifiant deux paramètres : (1) soit en jouant sur la croissance de l'épithélium, (2) soit en modifiant les paramètres du modèle de réaction-diffusion responsable de la formation des nœuds d'émail. En effet, Salazar-Ciudad & Jernvall [2010] montrent que la variation de ces deux paramètres permettent d'expliquer les différences morphologiques de dents de phoques.

D'autre part, des expériences de recombinaison de mésenchyme et d'épithélium de molaires soit entre molaires inférieure et supérieure soit entre différentes espèces (recombinaisons hétérotypiques) permettent de mettre en évidence le rôle du mésenchyme et de l'épithélium aussi bien en terme de signalisation que de proportion de tissus. Ainsi, des recombinaisons hétérotypiques entre des épithélia et des mésenchymes de molaires de souris et de rat, Cai *et al.* [2007], montrent que la taille des cuspides et la taille de la dent sont contrôlées respectivement par l'épithélium et le mésenchyme. De plus, ils observent que le nombre final de cuspides est dépendant de la taille du germe dentaire et de la taille des cuspides. Ces résultats sont cohérents avec le modèle *in silico* proposé par Salazar-Ciudad & Jernvall [2002] (voir *p. 44*) qui montre d'une part que la mise en place du patron de cuspide est dépendant de la morphologie du bourgeon dentaire. D'autre part, le modèle montre que la mise en place des cuspides est régulée par modèle de réaction diffusion entre un inhibiteur et un activateur, de la sorte plus la cuspide est grosse plus la zone d'inhibition autour d'elle est grande et empêche la mise en en place d'un autre cuspide. Hu *et al.* [2006] associent une même quantité d'épithélium avec différentes quantités de mésenchyme, ils montrent que l'augmentation de la quantité de mésenchyme permet de mettre en place plus de cuspides. Cependant ce n'est pas le seul critère qui permet de développer plus de cuspides puisque Ishida *et al.* [2011] montrent que la surface de contact entre l'épithélium et le mésenchyme détermine la taille de la couronne qui elle même influe sur le nombre de cuspides. Enfin, Schmitt *et al.* [1999] montrent, pour les premières molaires de souris, qu'à partir du stade "coiffe" (ED14.5), les explants recombinés développent un patron de cuspides spécifique du type de molaire dont provient le mésenchyme. Au regard de ces résultats et sachant qu'une plus grande quantité de mésenchyme permet de mettre en place plus de cuspides [Hu *et al.* , 2006], on peut émettre l'hypothèse que la première molaire supérieure de souris aurait

plus de mésenchyme que la première molaire inférieure, ce qui lui permettrait de former ces deux cuspides supplémentaires. Nous reviendrons sur ces conclusions au regard de nos résultats (annexe A, *p. 207*).

## 3.5 L'asymétrie bucco-linguale au cours du développement des dents

### 3.5.1 Le développement des dents de remplacement est lingual

L'identité buccal et lingual est très importante dans le développement des dents de remplacement. En effet, des reptiles [Handrigan *et al.* , 2010; Handrigan & Richman, 2010] aux mammifères [Järvinen *et al.* , 2008, 2009] les dents de remplacement se développent à partir de la lame dentaire de succession lingualement à la dent en développement de la génération précédente (figure 3.8A et B). Dans leur revue, Tucker & Sharpe [2004] recensent les mécanismes et les réseaux génétiques en jeux au cours du développement des dents de remplacement chez diverses espèces de vertébrés. Je vais plus particulièrement détailler deux réseaux d'interaction impliqués dans ce processus. Tout d'abord, il a été démontré que *Sox2*, exprimé dans la lame dentaire avec un biais lingual, marque la compétence épithéliale pour générer des dents de remplacement chez les mammifères (furet et humain) et les reptiles (serpent de maïs, python ballon, gecko, alligators et iguanes) [Juuri *et al.* , 2013] (figure 3.8). Puis, il a été montré qu'une forte activation de la voie Wnt dans l'épithélium oral de souris au stade ED13.5, permet le développement de dents surnuméraires multiples de forme conique simple, à la place des molaires [Järvinen *et al.* , 2006]. Il est intéressant de noter que la voie signalisation canonique de Wnt est constamment active à la pointe de l'épithélium dentaire et promeut la prolifération cellulaire chez les serpents [Handrigan & Richman, 2010].

Ces données suggèrent une conservation des mécanismes permettant de marquer le côté lingual, et ainsi de différencier le côté buccal du côté lingual, pour permettre le développement des dents de remplacement. De plus, Juuri *et al.* [2013] ont aussi identifié l'expression linguale de *Sox2* dans la lame dentaire de la souris au stade ED16, suggérant la réminiscence d'une dent de remplacement (figure 3.8C).

En conclusion, les données de la littérature suggèrent que le côté lingual des germes de dents est spécifique et différent du côté buccal. De plus, ces données sous-entendent que cette différence entre le côté buccal et le côté lingual est une propriété très ancienne de la dent car elle est établi à travers tous les vertébrés à mâchoire.

### 3.5.2 L'asymétrie bucco-linguale dans le développement des dents de souris

Que devient cette asymétrie bucco-linguale au cours du développement des molaires de souris, qui ont la caractéristiques de ne pas posséder de dent de remplacement? La réponse à cette question pourrait nous permettre de comprendre pourquoi les cuspides supplémentaires de la

première molaire supérieure de souris se mettent en place du côté lingual et non du côté buccal.

### 3.5.2.1 Asymétrie bucco-linguale du développement de la première molaire de souris

L'asymétrie bucco-linguale est aussi présente au cours du développement de la première molaire de souris [Cam *et al.* , 2002; Obara & Lesot, 2007; Lesot *et al.* , 2014]. Morphologiquement, cette asymétrie se caractérise par une différence de la taille des boucles cervicales entre le côté buccal et le coté lingual. Cependant, cette asymétrie du développement a reçu peu d'attention et n'est pas vraiment bien comprise.

La morphogenèse de la couronne ainsi que la formation séquentielle des cuspides est intrinsèquement asymétrique dans l'axe bucco-lingual. En effet, au stade ED15, le premier nœud d'émail secondaire (SEK) est mis en place du côté buccal et le second SEK est mis en place du côté lingual une demie journée plus tard (ED15.5-ED16), comme représenté sur la figure 3.9. Cho *et al.* [2007] étudient le devenir du nœud d'émail primaire (PEK), pour savoir si le PEK participe, au moins en

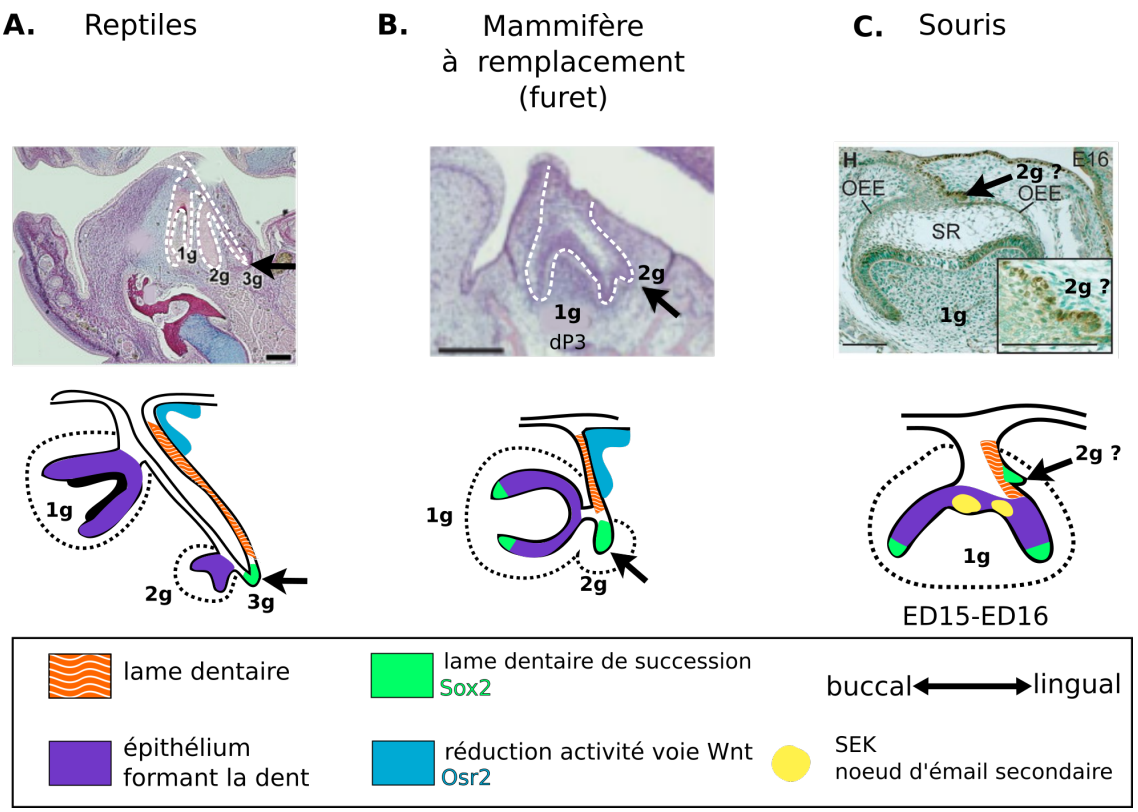



FIGURE 3.8 – **Développement des dents de remplacement.**
Coupes histologiques de germes dentaires au cours du développement : chez les *(A)* les reptiles, **B** les mammifères à dent de remplacement (tel le furet) et **C** la souris, accompagnées d'une représentation schématique. 1g : première génération ; 2g : deuxième génération et 3g : troisième génération. Les traits pointillés délimitent l'épithélium. Flèche : formation de la dent de remplacement. Sections tirées de Buchtová *et al.* [2008]; Järvinen *et al.* [2009]; Juuri *et al.* [2013]

partie, à la formation des SEKs. Par injection de colorant dans des cellules de PEK, les auteurs découvrent qu'une partie de ces cellules contribuent à la formation du premier SEK buccal, la première cuspide à se mettre en place (représentée en vert sur la figure 3.9). Néanmoins, ces résultats sont controversés car le même type de manipulation par injection d'un colorant dans le PEK conclu que les cellules du PEK ne participent à aucun des nœuds d'émail secondaires formés par la suite [Matalova *et al.* , 2005]. Cependant, Cho *et al.* [2007] montrent que sur la base de leurs résultats obtenus par injection couplés à des données histologiques montrant la continuité structurelle entre le PEK et le SEK buccal, les auteurs proposent que le PEK maintienne sa position initiale et devienne le SEK buccal alors que le SEK lingual est nouvellement formé plus tard. Bien que les auteurs ne le mentionnent pas, il semble évident à partir de ces données que le PEK/SEK buccal reste en fait à la même position et que c'est la croissance asymétrique du côté lingual qui permet la mise en place d'un nouveau SEK du côté lingual. Enfin, cette asymétrie bucco-lingual du développement du germe de première molaire est transitoire dans la molaire inférieure (ED15 à ED16) alors que cette asymétrie ne se résout jamais dans la molaire supérieure comme elle va mettre en place deux cuspides supplémentaires linguales.

Les approches gènes-centrées ont permis de mettre en évidence quelques gènes et voies de signalisation essentiels au développement asymétrique bucco-lingual de la première molaire de souris. Ainsi, Keränen *et al.* [1998] observent que les portions buccales et linguales du mésenchyme ont une activité génétique différentes, notamment pour l'expression de *Bmp4*,*Lef-1* et *Msx2* quand le PEK est présent, ce qui suggère une identité moléculaire différente entre les deux côtés. La table 3.1 regroupe les gènes dont l'expression est biaisé dans l'axe bucco-lingual au cours du développement de la molaire. Par exemple, il a été observé que la voie de signalisation Hedgehog jour un rôle majeur dans la croissance asymétrique des boucles cervicales dans l'axe bucco-lingual. Ainsi, il a été montré que l'absence d'expression de *Shh* [Dassule *et al.* , 2000], *Smo* [Gritli-Linde *et al.* , 2002] et *Evc* [Nakatomi *et al.* , 2013] perturbe la bonne formation de cet axe bucco-lingual. Certaine études ont aussi mis en évidence le développement de dent surnuméraires du côté lingual. Ainsi, Zhang *et al.* [2009] montrent que les souris mutantes $Osr2^{-/-}$ sont caractérisées par la formation de dents surnuméraire du côté lingual. Ils répertorient aussi que *Osr2* est exprimé suivant un gradient lingual à buccal au cours du développement de la molaire de souris. De même, il a été observé que le mutant *Runx2* développe un bourgeon lingual supplémentaire au niveau de la première molaire supérieure , au stade ED14, qui fait penser à une dent de remplacement, dans lequel la voie Hh est fortement activée [Åberg *et al.* , 2004a; Wang *et al.* , 2005]. Ainsi, il a été proposé que *Runx2* inhiberait la différentiation et/ou la prolifération de cellules épithéliales *Sox2* positives, qui sont les cellules compétentes à la formation d'un bourgeon [Togo *et al.* , 2016]. Or il a été montré que dans les boucles cervicales, un réseau de signalisation BMP-Shh est responsable du maintien des cellules souches *Sox2* positives [Li *et al.* , 2015], ce qui suggère que la même signalisation pourrait se avoir lieu dans la lame dentaire pour permettre les développement d'un bourgeon supplémentaire. Enfin, Kwon *et al.* [2015] montrent une interaction génétique entre *Osr2* et *Runx2*.

Ces études ont aussi mis en évidence une disparité développemental entre la molaire supérieure et la molaire inférieure. Ainsi, comme mentionné plus haut, le bourgeon lingual se forme uniquement dans la molaire supérieure et il a été observé que *Runx2* est nécessaire seulement dans la molaire inférieure pour l'activation de la voie Hh aux stades ED13-ED14.5 [Wang *et al.* , 2005] permettant la transition au stade coiffe de la molaire inférieure. Ces observations suggèrent que les programmes de développement entre molaire supérieure et inférieure sont différents. Cette différence des programmes de développement entre molaire inférieure et molaire supérieure pourrait expliquer pourquoi la molaire supérieure de souris met en place deux cuspides supplémentaires.

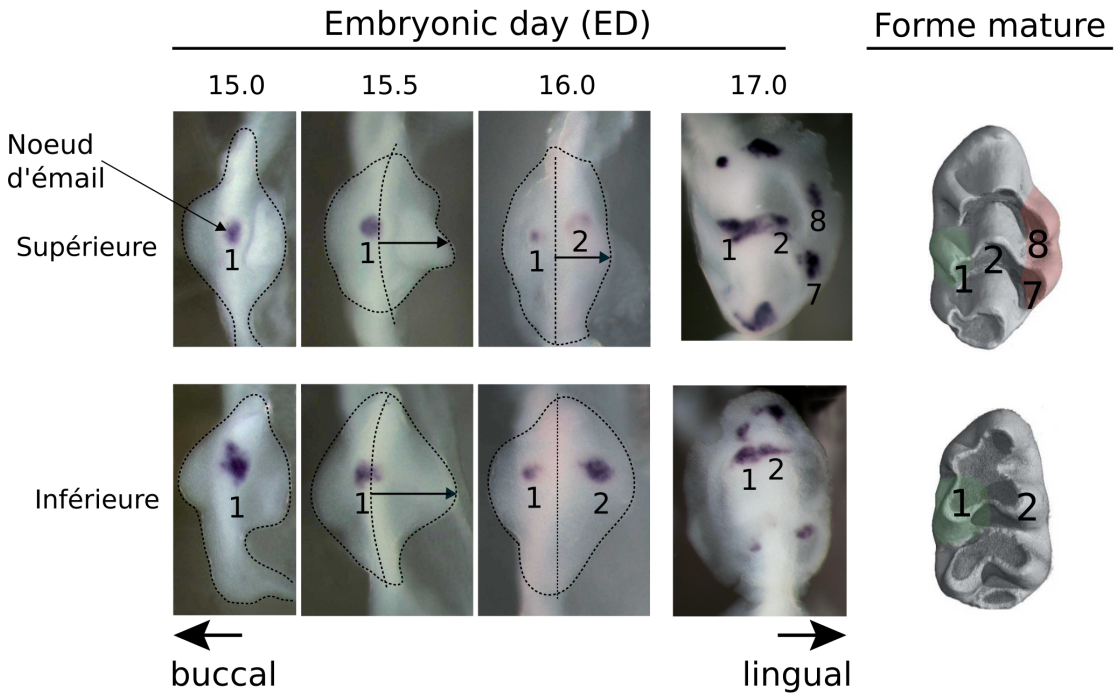



FIGURE 3.9 – **Asymétrie bucco-linguale du développement de la première molaire de souris**. Épithélia issus de germes entier de première molaire de souris, de 15 jours de développement (ED15) à la forme mature chez l'adulte. La formation des cuspides se déroule séquentiellement du côté buccal vers le coté lingual. Les cupides sont numérotées selon leur ordre d'apparition. Elles peuvent être visualisées sur les épithélia grâce à des hybridations *in situ* contre *Fgf4*, qui est un marqueur du nœud d'émail. La première molaire inférieure se compose de 6 cuspides distribuées sur 2 rangées. La première molaire supérieure développe 2 cuspides supplémentaires sur une troisième rangée linguale (colorées en rouge sur la forme mature). Ces cuspides supplémentaires sont les deux dernières cuspides de la molaire supérieure à se former. Notez la croissance asymétrique du côté lingual au stade ED15.5 qui permet la mise en place de la secondes cuspide du côté lingual. L'antérieur est vers le haut, le postérieur vers les bas. (Données provenant du laboratoire).

| Gène | biais | stade | tissu | voie | remarques | ref |
|---|---|---|---|---|---|---|
| *Bmp4* | buccal | 12/13/14<br>13.5<br>14.5 | mes | BMPs | | Åberg *et al.* [1997]<br>Chen *et al.* [1996]<br>Jia *et al.* [2012] |
| *Dlx2* | buccal | 13 | épi | | Expression dans le MES non biaisé BL | Zhao *et al.* [2000] |
| *Dkk2* | lingual | 13.5 | mes | Wnt | expression dans MES non condensé | Jia *et al.* [2016] |
| *Msx2* | buccal | 13.5 | épi | | expression dans corde buccale | Kettunen & Thesleff [1998] |
| | | 15.5 | | | Expression buccale dans le MES de $M^1$ | Cho *et al.* [2007] |
| | | 18.5 | | | Faible expression buccale dans le MES de $M^1$ | |
| *Osr2* | lingual | 13.5 | mes | voie | *$Osr2^{-/-}$ dents linguales surnuméraires* | Zhang *et al.* [2009] |
| *Ptch1* | lingual | 14.5 | épi | Hh | Expression dans le MES non biaisé BL | Dassule *et al.* [2000] |
| *Pitx1* | lingual | 14.5 | épi | Hh | | Dassule *et al.* [2000] |
| *Sfrp2* | lingual | 13.5 | mes | Wnt | expression dans MES non condensé | Jia *et al.* [2016] |
| *Sox2* | lingual | 13/14/16 | épi | | Exprimé dans les boucles cervicales à ED16. // Lien avec dents de remplacement des reptiles | Juuri *et al.* [2013] |

TABLE 3.1 – **Gènes biaisés bucco-lingualement au cours du développement de la première molaire de souris.** épi : épithélium ; mes/MES : mésenchyme ; $M^1$ : première molaire supérieure

## 3.6 Études transcriptomiques du développement des dents de souris

L'étude du transcriptome lors du développement permet une approche large échelle de la compréhension du programme de développement des dents. L'étude du transcriptome est donc un outil précieux pour comprendre le programme de développement des dents et leur dynamique temporelle. Ces études sont encore peu nombreuses, mais elles ont permis de mettre en évidence des réseaux d'interactions génétiques et des processus spécifiques à chaque stade de développement [Jia *et al.* , 2012; O'Connell *et al.* , 2012; Laugel-Haushalter *et al.* , 2013; Landin *et al.* , 2015].

Je vais détailler en particulier l'étude d'O'Connell *et al.* [2012], car je me suis servie de ce jeu données au cours de ma thèse pour effectuer les déconvolutions pour estimer les proportions

d'épithélium et de mésenchymes dans mes échantillons (voir la section déconvolution [*p. 77*] du chapitre 4 pour une explication de cette méthode). De plus cette étude illustre bien l'approche globale qui peut et qui devrait être utilisée pour extraire une information biologique en utilisant l'ensemble des données transcriptomiques. Cette étude ne s'arrête pas à la simple identification de gènes candidats, mais va plus loin dans l'utilisation des transcriptomes.

Le jeu de données de O'Connell *et al.* [2012] se compose de transcriptomes de tissus purs de germes de premières molaires inférieures de souris à différents temps de développement, de ED10 à ED13, échantillonnés toutes les 0.5 jours ; ainsi que du nœud d'émail à ED14.5. Les auteurs s'intéressent au dialogue entre le mésenchyme et l'épithélium lors des stades précoces du développement de la dent, de l'initiation (ED10) à la mise en place du pEK (primary Enamel Knot), le centre de signalisation responsable de l'initiation de la morphogenèse de la couronne (ED14.5). Ils reconstruisent un réseau de régulation génétique dans l'épithélium et le mésenchyme et identifient un circuit de rétroaction entre *Wnt* et *Bmp*. Ces deux voies de signalisation seraient les précurseurs initiant le dialogue entre l'épithélium et le mésenchyme au cours de l'odontogenèse.

Les données transcriptomiques générées au cours de cette étude, plus de 100 puces à ADN, ont donné lieu à l'établissement d'une banque de données : ToothCODE (http://compbio.med.harvard.edu/ToothCODE/index.html) qui donne l'accès aux niveaux d'expression des gènes dans les différentes puces. De plus, cette base regroupe des informations sur 5 des grandes voies de signalisation en jeux lors du développement de la dent : les voies BMPs, FGFs, Wnt, Hh et Activin, basée sur la revue de la littérature effectuée par les auteurs pour cette étude.

# Matériels et Méthodes Part II

# 4 Analyse des données RNA-seq

## 4.1 Présentation des jeux de données

Au total cette thèse présente l'analyse de 92 échantillons RNA-seq listés ci-dessous.
► Large échelle temporelle souris : bourgeons de première molaire supérieure et inférieur de souris (*Mus musculus*) à 8 stades de développement (de ED14.5 à ED18, échantillonnés tous les 0.5 jours et couvrant la période d'acquisition des cuspides) en duplicat (n=32 échantillons).
► Large échelle temporelle hamster : bourgeons de première molaire supérieure et inférieur de hamster doré (*Mesocricetus auratus*) à 8 stades de développement (de ED11.8 à ED14.5, couvrant la période d'acquisition des cuspides) en duplicat (n=32 échantillons).
► Petite échelle temporelle souris : bourgeons de première molaire supérieure de souris au stade ED15.5 (réplicats biologiques) provenant de 3 portées différentes (portées A, B et C ; n=7 échantillons) et bourgeon de molaire inférieure de souris au stade ED15.5 (provenant de la portée A).
► Bucco-lingual (BL) : bourgeons de première molaire inférieure et supérieure de souris au stade ED15 coupés en 2, de façon à isoler la partie buccale de la partie linguale du bourgeon, en duplicat provenant d'un seul et même embryon (n=8 échantillons).
► Tissus purs, épithélium et mésenchyme : bourgeons de première molaire inférieur et supérieure de souris au stade ED15 pour lesquels l'épithélium et le mésenchyme ont été dissociés afin d'isoler les deux tissu et pour lesquels la molaire droite et la molaire gauche du même embryon ont été mis en commun de manière à avoir assez de matériel, réalisé en duplicat.
► bourgeon de molaire de gerbille et de souris épineuse : bourgeons de première molaire inférieure et supérieure de gerbille de Mongolie (*Meriones unguicula*, n=2 échantillons) et de souris épineuse (*Acomys dimidiatus*, n=2 échantillons) à un stade équivalent à ED15.5 chez la souris.

Des analyses préliminaires réalisées dans l'équipe ont montré que les données transcriptomiques sont sensibles au temps passer à disséquer, au sexe et surtout à des fines différences de stades

de développement. Ces fines différences de stades de développement sont présentées dans le chapitre 8. Afin de minimiser ces effets et d'assurer une construction propre des données, toutes les germes ont été disséqué sur des embryons femelles et la molaire supérieure et inférieure à un stade de développement *t*, proviennent d'un même individu, ce qui permet de s'assurer que les différences de développement observées sont bien dues à une différence inf vs.sup et non à une différence entre individus. C'est pour cette raison que nous n'avons pas mis en communs plusieurs germes, cependant cette approche nous oblige à travailler avec des faibles quantités d'ARN.

## 4.2 Extraction d'ARN

J'ai réalisé l'extraction d'ARN pour le jeu de données de tissus purs (table 4.3) (chapitre 10) ainsi que pour le jeu de données de bourgeons de molaires de gerbille et de souris épineuse (données non présentées car trop préliminaires mais discutées). L'extraction d'ARN de bourgeons de molaires nécessite un protocole spécifique car on travaille alors avec de faibles quantité d'ARN. De manière à préserver la qualité de l'ARN, les échantillons ont été placés en RNA*later*®puis stockés à -20°. Il faut veiller à bien débarrasser l'échantillon du maximum de RNA*later*®avant d'effectuer l'expression d'ARN. J'ai ensuite lysé les échantillons à l'aide d'un Precellys®, puis j'ai extrait l'ARN à l'aide d'un kit Quiagen RNeasy micro. Ensuite, par quantification fluorométrique, j'ai quantifié la quantité d'ARN extraits à l'aide d'un Qubit (thermofisher), puis j'ai vérifié la qualité et la quantité des ARN extraits sur TapeStation 2200 (agilent). Une fois les échantillons validés, j'ai coordonné l'envoi des échantillons à la plateforme de séquençage à Strasbourg (IGBMC), avec lesquels j'ai assuré la mise en place et la gestion de ces deux projets.

## 4.3 Estimation du niveau d'expression à partir de données RNA-seq

*Pour les chapitres 8, 9 et 10.*

J'ai réalisé les alignements ainsi que les estimations des niveaux d'expression au Pôle Scientifique de Modélisation Numérique (PSMN, http://www.ens-lyon.fr/PSMN/doku.php?id=accueil:presentation).

J'ai utilisé une approche classique composée de l'aligneur bowtie2 [Langmead & Salzberg, 2012] (version 2.0.2) suivi du compteur de lectures htseq-count [Anders *et al.* , 2014] (version 0.5.3.p9 et python version 2.7), paramétré de façon à ce que ne soient pas pris en compte les lectures chevauchant 2 gènes (-m intersection-nonempty). J'ai aussi testé le pseudo-aligneur kallisto [Bray *et al.* , 2016] (version 0.43.0) Dans les deux cas, j'ai aligné les données sur le transcriptome de la souris (ensembl GRCm38 release 85 de Juillet 2016). J'ai ensuite effectué l'analyse statistique de ces données d'expression à l'aide du logiciel R (version 3.3.1 [R Core Team, 2016])et le paquet ggplot2 [Wickham, 2009] pour réaliser la majorité des graphiques.

J'ai pu observer que les deux méthodes donnaient le même signal de PCA, cependant ces méthodes diffèrent dans l'estimation des enrichissements des gènes différentiellement exprimés (obtenu par une analyse DESeq2 [Love *et al.* , 2014]). Cela tient certainement à l'estimation à la hausse des gènes faiblement exprimé dans l'analyse Kallisto, ce qui n'est pas prévu par les lois statistiques utilisées par le paquet DESeq2 [Love *et al.* , 2014] pour identifier les gènes différentiellement exprimés, causant ainsi un biais dans la normalisation des données et l'estimation des gènes DE. Comme représenté sur la figure 4.1, on voit que la pente de la droite de régression linéaire entre les niveaux d'expressions estimés par les deux méthodes est significativement inférieure à 1 et dont l'ordonnée à l'origine est égale à 0.31, ce qui traduit une estimation à la hausse des faibles comptes (gènes faiblement exprimés) ainsi qu'une estimation à la baisse des forts comptes (gènes fortement exprimés). Les développeurs de Kallisto conseillent d'utiliser le paquet Sleuth [Pimentel *et al.* , 2016] récemment développé. Dans un premier temps, pour des question d'historique et de maitrise du paquet, nous avons préféré continuer d'utiliser l'outil DESeq2.

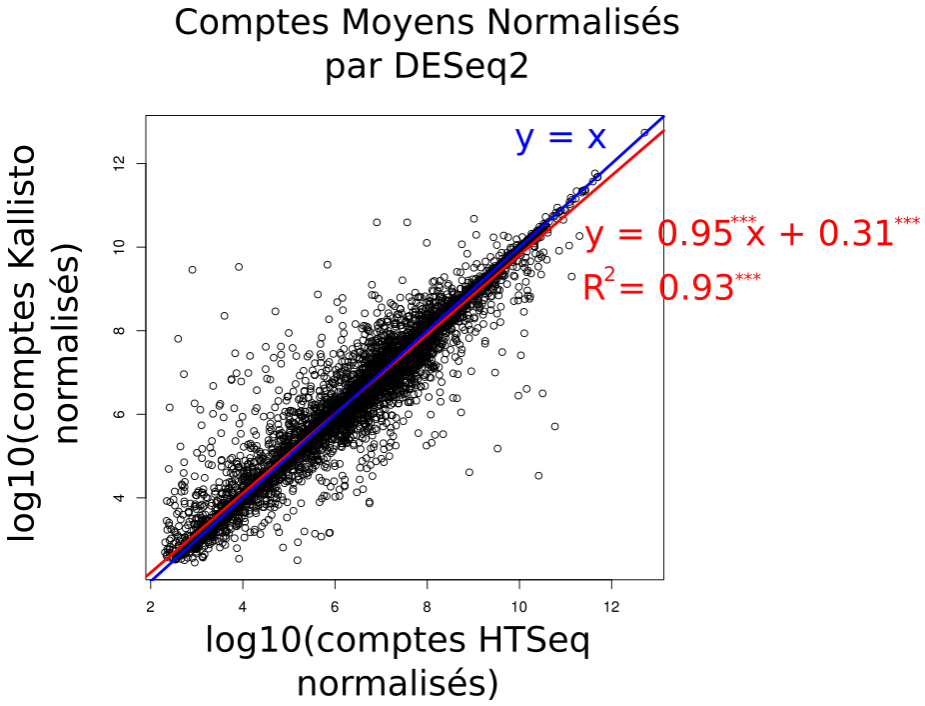



FIGURE 4.1 – **Comptes Kallisto vs. HTSeq** Comptes moyens normalisés avec DESeq2 pour le jeux de données bucco-lingual ayant été obtenus soit avec Kallisto (en ordonnée) soit avec htseq-count (en abscisse). La droite bleu représente la droite y=x, et la droite rouge est la régression linéaire entre les comptes obtenus pas kallisto et ceux obtenus par htseq-count de formule $y = 0.95x + 0.31$. (*** : , *pvalue* < 2.2e-16)

| | Petite échelle temporelle | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| stade | 15.5 | | | | | | | |
| portée | A | | B | | | C | | |
| dent | low | up | up | up | up | up | up | up |
| noms | A | | B1 | B2 | B3 | C1 | C2 | C3 |
| poids(mg) | 410 | | 495 | 495 | 447 | 460 | 461 | 511 |
| Total comptes | 25.1 M | 21.9 M | 17.2 M | 16.9 M | 22.3 M | 15.3 | 20.1 M | 15.4 |

TABLE 4.1 – Données réplicats biologiques RNA-seq de bourgeons dentaires de souris au stade ED15.5. *(M= millions de lectures)*

| | BL | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| stade | 15.0 | | | | | | | |
| poids (mg) | 352-359 | | | | | | | |
| dent | low | | | | up | | | |
| replicats | rep1 (dent gauche) | | rep2 (dent droite) | | rep1 (dent gauche) | | rep2 (dent droite) | |
| côté | buccal | lingual | buccal | lingual | buccal | lingual | buccal | lingual |
| nom | low1 B | low1 L | low2 B | low2 L | up1 B | up1 L | up2 B | up2 L |
| taille lib | 37.4 M | 32.8 M | 33.9 M | 39.8 M | 64.1 M | 44.2 M | 80.2 M | 56.4 M |
| Total comptes | 27.3 M | 24 .2 M | 24.9 M | 29.9 M | 47.2 M | 33.2 M | 60.2 M | 42.9 M |
| % Total comptes | 73% | 74% | 73.5% | 75% | 74% | 75% | 75% | 76% |

TABLE 4.2 – Librairies bucco-linguales de bourgeons de molaire de souris au stade ED15. *(M= millions de lectures)*

| | Tissus purs | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| stade | 15.0 | | | | | | | |
| réplicat | rep1 | | | | rep2 | | | |
| infos | pool dent droite + gauche | | | | pool dent droite + gauche | | | |
| poids (mg) | 370 | | | | 350 | | | |
| dent | low | | up | | low | | up | |
| tissu | epi | mes | epi | mes | epi | mes | epi | mes |
| taille lib | 81.4 M | 108 M | 99.7 M | 99.5 M | 99.7 M | 82.4 M | 130.9 M | 64.7 M |
| Total comptes | 62.3 M | 83.1 M | 76.4 M | 76.3 | 73.7 M | 60.5 M | 93.8 | 47.3 |
| % Total comptes | 76.5% | 77% | 77% | 77% | 74% | 73.5% | 72% | 73% |

TABLE 4.3 – Librairies tissus purs. *(M= millions de lectures)*

### 4.3.1 Comparaison des niveaux d'expression entre espèces

*Cette méthodologie se rapport au chapitre 9.*

Une des difficultés de comparer les niveaux d'expressions entre plusieurs espèces est l'obtention de données d'expression qui soient comparables. Cette difficulté est d'autant plus grande quand on travaille à la fois avec des espèces qui disposent d'un génome bien annoté, comme la souris, et des espèces non-modèles qui n'ont pas de génome, comme le hamster doré, la gerbille et la souris épineuse. Ceci va biaiser l'estimation de l'abondance des transcrits de chaque espèce et ainsi biaiser la comparaison des niveaux d'expression entre ces espèces. Pendant mon master et au tout début de ma thèse, j'ai donc réaliser des études préliminaires afin de tester deux approches : (i) l'alignement direct ( direct mapping) sur génome de référence et (ii) l'assemblage *de novo* d'un transcriptome. Ces tests par approches classiques ont permis de mettre en évidences les problèmes rencontrés par ces deux méthodes et sont à la base des réflexions ayant conduit au développement d'une méthode "maison" développée par Carine Rey, une étudiante en thèse de l'équipe dont le principe est présenté dans l'introduction du chapitre 9. C'est cette méthoque qui a été utilisée pour comparer les niveaux d'expression entre les données de souris et de hamster (chapitre 9.

#### 4.3.1.1 Alignement des lectures sur un génome de référence

J'ai testé l'aligneur TopHat [Trapnell *et al.* , 2009] (version 1.4.0) sur le génome de la souris (NCBIM37) pour deux échantillons de souris (*Mus musculus*). Cependant, le hamster doré (*Mesocricetus auratus*) ne possédait pas de génome de référence (son génome a été publié en 2014 mais la qualité des contigs et de l'annotation n'est pas satisfaisante Tchitchek *et al.* [2014]), j'ai donc utilisé le génome de cellules ovariennes de hamster chinois (CHO, *Cricetulus griseus*) [Hammond *et al.* , 2011] pour effectuer l'alignement de deux échantillons de hamster. J'ai ensuite utilisé htseq-count [Anders *et al.* , 2014] pour estimer les niveaux d'expression des transcrits. Puis j'ai comparé les niveaux d'expression entre le hamster et la souris obtenu par cette analyse.

#### 4.3.1.2 Assemblage *de novo* de transcriptomes avec *Trinity*

En parallèle, j'ai testé un assemblage de transcriptome *de novo* sur des échantillons de souris et de hamster doré. J'ai utilisé Trinity [Grabherr *et al.* , 2011; Haas *et al.* , 2013] (version 4.2.1) pour assembler *de novo* les transcriptomes couplé à RSEM [Li & Dewey, 2011] pour estimer l'abondance des transcrits. J'ai ensuite annoté les contigs reconstruits en effectuant un BLAST entre ces contigs et le génome de la souris (NCBIM37) et en ne gardant que les meilleurs alignements (approche *reciprocal best hit*). Enfin j'ai comparé les niveaux d'expressions entre le hamster et la souris.

#### 4.3.1.3 Comparaison des deux méthodes et stratégie adoptée

L'hypothèse est que de façon générale les niveaux d'expression dans les molaires de souris et de hamster doivent être corrélés, à l'exception d'un petit nombre de gènes qui présentent une expression différente entre les deux espèces et qui seraient responsable des caractéristiques spécifiques de chacune des molaires dans les deux espèces. Ainsi, nous avons émis l'hypothèse que la méthode qui donnerait la meilleure corrélation entre les niveaux d'expression du hamster et de la souris serait la méthode la plus approprié pour estimer des niveaux d'expressions les plus comparables possibles entre les deux espèces. La comparaison entre l'alignement direct et l'assemblage *de novo*) montrent que l'on a une meilleurs corrélation entre les niveaux d'expression de souris et de hamster obtenus par assemblage *de novo*. La comparaison entre les niveaux d'expression obtenus sur le transcriptome *de novo* de la souris et sur l'alignement directe sur le génome de la souris permet d'évaluer les biais introduit par l'assemblage *de novo* d'un transcriptome. Ainsi, nous avons noté que l'approche par assemblage *de novo* n'est pas satisfaisante car on reconstruit un nombre important de contigs, dont certains sont des chimères de deux transcrits et il est fastidieux d'identifier ces transcrit chimères. Enfin, le plus gros point faible de cette approche *de novo* est l'annotation des contigs reconstruits, à cause notamment des duplications et des pertes. C'est pourquoi, une méthode spécifique a été développé au laboratoire afin de pallier à ces problèmes. C'est une méthode qui permet de reconstruire *de novo* de transcrits en se basant sur un transcrit de référence et qui utilise les informations phylogénétiques pour identifier précisément les transcrits dans l'ensemble des espèces traités (voir chapitre 9 et figure 9.1 *p. 111*).

## 4.4 Analyses multivariées

*Analyses réalisées dans les chapitres 8, 9, 10.*

### 4.4.1 Les différentes analyses multivariées

Le but des analyses multivariées est de réduire la complexité des données pour révéler leurs structurations sous-jacentes. Une façon simple d'estimer la similarité entre deux points est le calcul de la distance euclidienne entre ces deux points : si elle est faible cela signifie que les points sont similaires cependant si cette distance est grande on ne peut rien conclure sur ce qui cause cette forte dissimilarité. J'ai testé trois analyses multivariées différentes, une méthode de réduction de dimensionnalité linéaire : (1) l'analyse en composante principale (PCA) et deux méthodes de réduction de dimensionnalité non-linéaire : (2) t-SNE [Maaten & Hinton, 2008] et (3) diffusion map [Coifman *et al.* , 2005, 2008; Angerer *et al.* , 2015].

**La PCA** a pour but de reconstruire la cartographie linéaire entre un objet à haute dimension ($n \times p$, où $n$ est le nombre de gènes exprimés et $p$ est le nombre d'échantillons par exemple) et un sous espace à $d$ dimensions (où $d < n \times p$) tout en capturant le maximum de variabilité dans

les données. Ce sous-espace est défini par $d$ vecteurs orthogonaux : les composant principaux (PCs), qui sont les axes de la PCA [De la Porte *et al.* , 2008] (voir box PCA *p. 26*).

Les méthodes non-linéaires partent du principe que les données d'un objet à forte dimension ($nxp$) sont souvent intégrées dans une structure à faible dimension. Il serait alors possible de caractériser les données ainsi que les relations entre chaque points en utilisant moins de dimensions (2 à 3), en mesurant les distances entre chaque point dans cette structure à faible dimension plutôt que dans l'espace euclidien.

**t-SNE** (t-distributed Stochastic Neighbor Embeding) ou intégration stochastique du voisinage par distribution t de Student, permet d'estimer la distribution du voisinage de deux points. Concrètement, deux matrices sont calculées : (1) la matrice de probabilité $p_{ij}$ constituée de la distance des points $x_i$ et $x_j$ dans l'espace à haute dimension, ceci mesure la proximité de $x_j$ de $x_i$, en considérant une distribution gaussienne autour de $x_i$ avec une variance donnée $\sigma_j^2$, (2) la matrice de similarité $q_{ij}$ qui mesure la distance des points $x_i$ et $x_j$ dans l'espace à faible dimension permettant d'évaluer la proximité des points en considérant une distribution $t$ de Student [Rossant, 2015]. Enfin, le meilleur espace à faible dimension est l'espace dans lequel la distance entre $p_{ij}$ et $q_{ij}$ est la plus faible. Cette méthode à tendance à séparer les données en paquets.

Le principe de la **diffusion map** est le suivant : la connectivité entre deux points $x$ et $y$, est la probabilité de sauter de $x$ à $y$ en une seule étape le long d'une marche aléatoire (définie par une gaussienne). Cette étape permet de définir des noyaux de similarité (kernel) à l'intérieur desquels la connectivité entre les points composant le noyau est forte et diminue rapidement lorsque l'on sort de ce noyau [De la Porte *et al.* , 2008]. À partir de la matrice de diffusion, qui contient la connectivité entre deux points, la distance de diffusion est calculée à partir du processus de diffusion qui correspond à la marche aléatoire le long des noyaux de connectivité pour aller du point $x$ au point $y$. Un nouvel espace de diffusion est alors défini dans lequel la distance euclidienne entre les points correspond à la distance de diffusion. Cette méthode a plutôt tendance à organiser les données en gradient. C'est pourquoi cette méthode est plus appropriée pour les analyse de données différentiation cellulaire et d'engagement du destin cellulaire.

Le problème de la PCA est qu'elle a tendance à donner beaucoup de poids aux points très éloignés et donc à amoindrir les différences faibles (bien que ces différences existent), alors que les deux méthodes non-linéaires présentées ici vont chercher à conserver à la fois ces fortes différences (structure globale) tout en conservant l'information de ces faibles différences (structure locale). Cependant, la PCA a un aspect quantitatif que t-SNE et diffusion map n'ont pas. De plus, la PCA permet d'avoir des informations sur les variables, dans notre cas c'est-à-dire les gènes, ce qui permet de récupérer les coordonnées des gènes sur les axes ainsi que d'identifier les gènes qui contribuent le plus à ces différent axes. C'est pourquoi j'ai principalement utilisé la PCA pour estimer la structuration de mes données.

### 4.4.2 Identifier les axes majeurs de variations et caractériser la participation de chaque tissu à cette variation

J'ai réalisé des analyses multivariées, et majoritairement des analyses en composantes principales, à l'aide du paquet ade4 [Dray & Dufour, 2007] (version 1.7.4) afin d'identifier les axes majeurs de variation dans les différents jeux de données RNA-seq. J'ai aussi interrogé la participation de chaque compartiment tissulaire du bourgeon de molaire (*eg.* le mésenchyme, l'épithélium et le nœud d'émail) en réalisant des PCA uniquement sur des sous-ensembles de gènes marqueurs de ces trois compartiments (*cf.* chapitres 8 et 10). De plus, afin de découpler les effets entre la proportion relative de tissu dans le bourgeon et l'état de maturation du tissu, j'ai réalisé ces PCA soit en conservant les données normalisées sur l'ensemble des gènes, conservant ainsi l'information de proportion de tissu, soit en re-normalisant les niveaux d'expression uniquement sur mon sous-ensemble de gènes marqueurs, éliminant ainsi les informations de proportions relatives des tissus et conservant uniquement l'information de nature du tissu. Une représentation schématique de l'influence de cette re-normalisation sur un sous-groupe de gènes marqueurs d'un compartiment tissulaire est présenté dans l'annexe A (figure supplémentaire 6a). La façon dont j'ai sélectionné ces marqueurs est décrite dans la section déconvolution ci-dessous (*p. 78*)

### 4.4.3 Analyses *between* : identifier la variation associée à un facteur

Un axe de PCA représente un axe de variance dans les données et non pas un facteur biologique. Une façon de connaitre le pourcentage de variation associé à un facteur donné est l'utilisation d'une analyse multivariée supplémentaire/complémentaire appelée *between-group*. La *between* va estimer un individu moyen pour chaque groupe, ces individus moyens sont ensuite utilisés pour effectuer une analyse multivariée puis les échantillons originaux sont projetés dans sur cette analyse (voir figure 4.2). De la sorte, l'analyse en *between-group* va permettre d'identifier le pourcentage de variance associé spécifiquement à un facteur biologique. Ainsi, j'ai pu estimer le pourcentage de variation associé à l'identité buccale ou linguale dans le jeu de données BL (chapitre 10) ainsi que la variation associée au type de molaire (inférieure vs. supérieure) dans le jeu de données BL et le jeu de données souris-hamster (chapitre 9) ou encore la variation associé à identité des tissus (épithélium vs. mésenchyme) dans le jeu de données tissus purs exploité lors des analyses bucco-linguales (chapitre 10).

Cette approche est complémentaire à l'analyse des gènes différentiellement exprimé (DE) car les très faibles différences de niveaux d'expression vont pouvoir être capturées par l'analyse multivariée alors qu'elle ne le seront pas par l'analyse DE qui ce concentre sur les gènes dont le niveaux d'expression varie plus fortement.

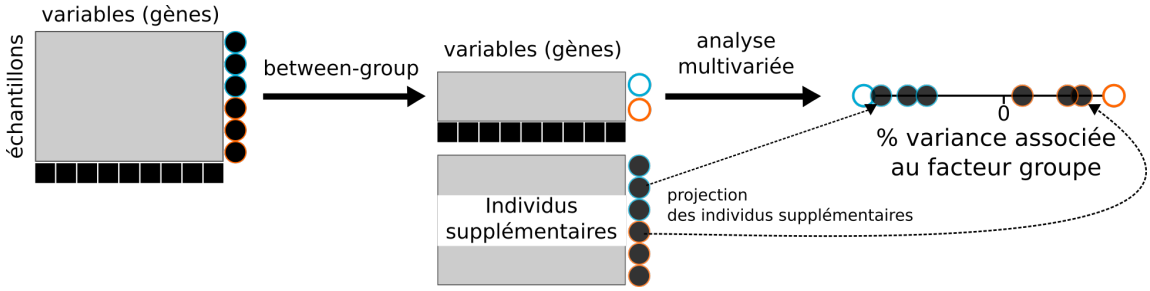



FIGURE 4.2 – **Principe de l'analyse multivariée en *between-group*.** Représentation schématique du principe de l'analyse multivariée *between-group*. Un hypothétique jeu de données composé de 9 variables (les rectangles pleins) et de 6 échantillons (les cercles pleins) répartis dans deux groupes : orange et bleu. Les cercles vides représentent le niveaux d'expression moyen de chaque variables par groupes.

### 4.4.4 Distance BL sur PCA et Bootstrapping

*Cette partie se réfère au chapitre 10 : l'étude de l'asymétrie bucco-linguale dans les germes de molaires de souris.*

#### 4.4.4.1 La distance BL est-elle la même entre la molaire inférieure et la molaire supérieure ?

J'ai estimé la distance entre les librairies buccales et linguales des molaires supérieures et inférieures. Afin de tester si cette distance est statistiquement différente entre la molaire supérieure et la molaire inférieure j'ai réalisé un grand nombre d'analyse PCA (itérations = 1000) sur un sous-ensemble de 500 gènes tirés au hasard à chaque analyse puis j'ai vérifié que l'axe 1 correspondait bien à un axe qui sépare les échantillons buccaux des échantillons linguaux, enfin j'ai mesuré la distance entre les librairies buccales et linguales sur PC1 pour chaque type de molaire (inf. vs sup.). Cependant, la comparaison d'une distance directement entre différentes PCA n'est pas possible, j'ai donc pallié à ce problème en comparant le ratio de la distance BL pour la molaire supérieure et inférieure. De la sorte j'ai obtenu la distribution du ratio BL entre molaire inf. et sup., que j'ai pu modéliser par une loi normale dont j'ai estimé les paramètres.

#### 4.4.4.2 L'excès de différence BL est-il dû à une différence de progression du programme développemental ?

Afin d'estimer si une différence de progression du programme de développement entre la molaire supérieure et la molaire inférieure est responsable de la plus grande distance entre les transcriptomes buccaux et linguaux de la molaire supérieure, j'ai réalisé le même genre d'analyse que celui décrit dans le paragraphe précédent. Je me suis servie de la PCA obtenue sur les données de bourgeons de molaires entières à différents temps de développement (le jeu de données large échelle). J'ai ensuite projeté les échantillons BL sur cette PCA et mesuré la distance BL sur PC1, qui est un axe de temps. Puis j'ai réalisé un grand nombre d'analyse PCA (itérations = 1000) sur

un sous-ensemble de 1000 gènes tirés au hasard à chaque analyse en ne gardant que les PCA pour lesquels PC1 ordonne correctement les échantillons, c'est-à-dire que l'ordonnancement temporel des échantillons de germes entiers est conservé. Enfin j'ai mesuré la distance entre les librairies buccales et linguales sur PC1 pour chaque type de molaire (inf. ou sup.). Puis j'ai extrait les ratios des distances BL entre la molaire supérieure et inférieure, ce qui m'a permis d'ajuster une loi normale et d'en estimer les paramètres.

## 4.5 Analyse des gènes différentiellement exprimés (DE)

*J'ai réalisé cette analyse DE dans les chapitres 8, 9 et 10.*

Une analyse extrêmement classique des données transcriptomiques est d'identifier les gènes dont l'expression est significativement modifiée entre deux conditions. J'ai réalisé l'analyse des gènes différentiellement exprimés (DE) avec le paquet DESeq2 [Love *et al.* , 2014]. Pour le jeu de donnée BL, j'ai pu identifier les gènes DE entre le coté buccal et le côté lingual pour la molaire inférieure et supérieure de manière séparée (formule : $\sim bl$). J'ai aussi pu identifier les gènes DE BL en prenant en compte l'identité inf-sup de la molaire (formule $\sim dent + bl$). Enfin, j'ai identifié les gènes DE BL dit "en interaction" entre le facteur bl et le facteur dent, c'est à dire les gènes pour lesquels le biais bucco-lingual est différent entre la molaire inférieure et la molaire supérieure (*i.e.* intensité du biais, sens du biais) qui sont les gènes que l'on a appelé *gènes exacerbés* , obtenus avec la formule : $\sim dent + bl + dent : bl$.

J'ai également déterminé les gènes DE entre la molaire inférieure et la molaire supérieure dans tous les jeux de données analysés (sauf le jeu de donnée petite échelle temporelle). Pour les jeux de données large échelle hamster et souris, j'ai estimé les gènes DE entre la molaire inférieure et supérieure à chaque temps de développement et pour chaque espèce séparément (formule : $\sim dent$). J'ai ensuite comparé le nombre de gènes DE entre molaire inférieure et molaire supérieure à chaque temps de développement entre chaque espèces et ainsi identifié un plus grand nombre de gènes biaisé dans la molaire supérieure à la fois chez la souris et chez le hamster ainsi qu'un excès de gènes DE à la fin de la mise en place de la deuxième cuspide et au moment de la mise en place de la troisième, commun aux deux espèces.

### 4.5.1 Estimation de "l'âge" des échantillons BL

*Cette analyse se réfère au chapitre 8.*

Pour définir "l'âge" des échantillons BL j'ai d'abord dû définir ce que l'on a appelé les gènes "timers". Ces gènes sont les gènes dont l'expression augmente ou diminue significativement d'un facteur minimum de 2 entre le début et la fin de la cuspidogenèse. J'ai défini ces gènes "timers" à partir du jeu de donnée large échelle souris, en réalisant une analyse de gènes DE, à l'aide de DESeq2. Pour cela, j'ai pris en compte le stade de développement (formule : $\sim temps$,

où *temps* est un facteur numérique contenant les différents stades de développement) pour la molaire inférieure et supérieure séparément. Puis, j'ai récupéré la pente, qui correspond à l'enrichissement calculé par DESeq2, ainsi que l'ordonnée à l'origine aussi calculé par DESEq2 lors de l'analyse DE, ce qui m'a permis d'obtenir des équations de droite pour chaque gène. De la sorte, connaissant le niveaux d'expression d'un gène dans le jeux de données BL, j'ai pu estimer l'âge de l'échantillon pour ce gène en effectuant une simple projection, comme représenté sur la figure 4.3. Ainsi, j'ai obtenu un âge d'échantillon pour chaque gène "timer", j'ai ensuite pu estimer l'âge moyen de l'échantillon basé sur la moyenne de ces âges par gènes. Les résultats de cette analyse sont présentés dans le chapitre 8.

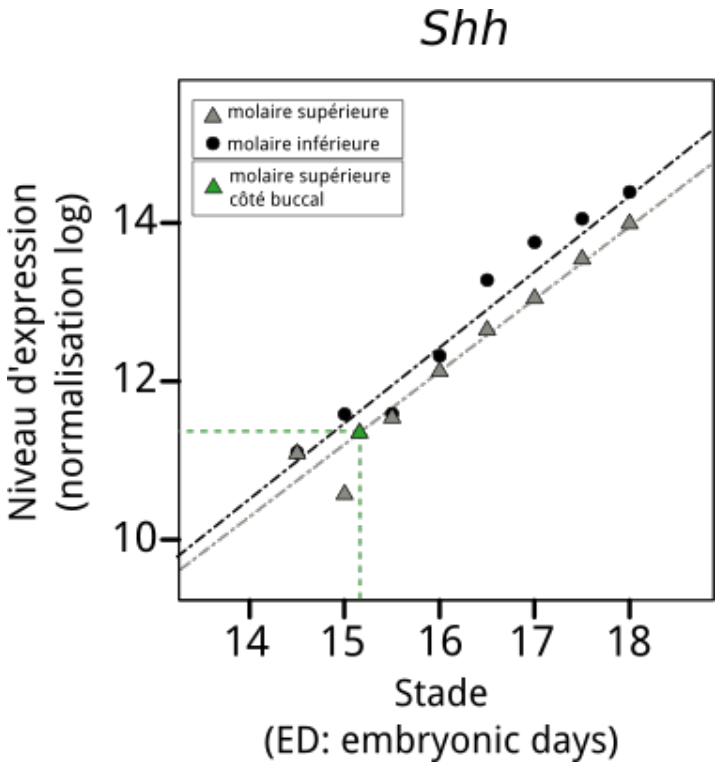



FIGURE 4.3 – **Principe de l'estimation de l'âge des échantillons bucco-lingual.** Dans un premier temps, il faut définir des gènes qui augmentent ou diminuent plus ou moins linéairement au cours de la cuspidogenèse afin d'obtenir l'équation d'une droite. Puis, connaissant le niveau d'expression d'un gène dans un échantillon, on peut définir l' "âge" de l'échantillon en effectuant une simple projection sur cette droite.

### 4.5.2 Estimation du stade de développement des échantillons du jeu de données petite échelle

*cf. chapitre 8.*

Les échantillons du jeu de données petit échelle sont des réplicats biologiques, ils proviennent d'embryons de souris à ED15.5 issus de trois portées différentes, comme explicité dans la table 4.1.

Cependant, lorsque ces échantillons sont projetés sur une PCA obtenue sur le jeux de donné large échelle (voir chapitre 8), ils présentent des coordonnées différentes sur l'axe temporelle PC1 (bien que proches de ED15.5 comme attendu). Ces données suggèrent que ces réplicats biologiques du stade ED15.5 n'ont pas exactement le même état d'avancement de leur programme de développement. Comme on peut le voir sur la figure 4.3, le transcrit *Shh* (exprimé dans le nœud d'émail) augmente de façon linéaire au cours du temps dans les deux molaires, ainsi il pourrait être utilisé comme marqueur de l'avancement du programme développemental. C'est pourquoi, j'ai testé la corrélation entre le niveau d'expression de *Shh* dans le jeu de données petite échelle (n=7 bourgeons de molaire supérieure) et les autres gènes de ce jeux de données. En parallèle j'ai testé la corrélation des niveaux d'expression dans le jeu de données petite échelle avec deux autres facteurs qui pourraient approximer un vecteur temporelle : (i) le poids des embryons (voir chapitre 8) [Peterka *et al.* , 2002] et (ii) les coordonnées des échantillons petite échelle sur l'axe 1 de la PCA large échelle représentant un axe temporelle. J'ai réalisé un test de corrélation de Pearson et de Spearman (sauf pour le poids). Les résultats de cette analyse sont présentés sur la figure 4 du chapitre 8 où l'on voit que poids n'est pas un bon indicateur d'avancée développementale alors que les coordonnées sur PC1 et les niveaux d'expression de *Shh* sont eux de bons indicateurs de l'état d'avancement du programme développemental. Ensuite, par une analyse de gènes DE (formule : ~ *niveau_Shh*), j'ai pu définir les gènes "linéaires" qui sont les gènes dont l'expression augmentent/diminuent significativement dans le jeu de données petite échelle avec le niveau d'expression de *Shh* qui est un proxy de l'état d'avancement du développement et donc du temps à petite échelle.

## 4.6 Corrélations entre jeux de données grande et petite échelle temporelle de souris

Cette section fait référence à la figure 4 du chapitre 8. Le but de cette manipulation était de voir si les variations des niveaux d'expression à petit échelle prédisent les variations dans le jeu de donnée large échelle. Pour cela, j'ai repris mon analyse de corrélation entre les niveaux d'expression petite échelle et les 3 facteurs : poids, coordonnées PC1 et niveau d'expression *Shh* (comme expliqué dans le section précédente *cf. p.4.5.2*). Puis j'ai séparé les gènes en trois catégories : les gènes dont l'expression (1) augmente ou (2) diminue significativement au cours du temps avec un facteur d'au moins 2 dans le jeu de données large échelle (les gènes "timers", *cf. p. 72*) et (3) les gènes dont l'expression ne varie pas significativement avec le temps. J'ai ensuite testé et validé que les distribution de ces trois catégories de gènes appartiennent à des populations différentes pour PC1 et le niveau de *Shh*. Ensuite, j'ai effectué la manipulation réciproque en estimant la corrélation (spearman et pearson) entre les niveaux d'expression des gènes du jeu de données large échelle avec le temps (=les stades de développement) puis j'ai séparé les gènes en trois catégories : les gènes dont l'expression (1) augmentent ou (2) diminuent significativement avec le niveau d'expression *Shh* dans le jeu de données petite échelle (les

gènes “linéaires”, *cf. p. 73*). J'ai testé et validé que les trois distributions représentent bien trois populations différentes. Afin de tester ce résultat, j'ai effectué la même manipulation mais cette fois lors du calcul des corrélations au lieu de donner l'ordre exact des stades de développement j'ai donné un ordre aléatoire de temps pour chaque gène ce qui donne la distribution attendue sous l'hypothèse neutre si les données n'étaient pas structurées.

## 4.7 Exclure les artéfacts statistiques des données bucco-linguales

Cette méthodologie à été appliquée dans le chapitre 10.

### 4.7.1 Ré-échantillonnage des librairies bucco-linguales

Les librairies bucco-linguales ont des tailles très variables (voir table 4.2) avec un facteur allant jusqu'à de 2.5, ce qui est particulièrement gênant car cela va introduire une différence technique entre la molaire supérieure et inférieure et peut donc interférer avec les différences biologiques qui nous intéressent. En effet, les librairies de la molaire inférieure ont en moyenne 1.7 fois moins de lectures que les librairies de la molaire supérieure. Cette variabilité dans le nombre de lectures va biaiser l'estimation du niveau d'expression des gènes car les transcrits faiblement exprimés vont être mécaniquement plus représentés dans un librairie ayant un nombre de lecture plus grand. Indirectement, la taille des librairies va donc biaiser l'estimation des gènes différentiellement exprimés. Pour pallier ce biais, j'ai réalisé 2 ré-échantillonnages aléatoires pour des librairies ayant une taille finale de 20M de lectures chacune. Pour chacun des deux ré-échantillonnage des 8 librairies BL, j'ai utilisé l'aligneur TopHat2 [Kim *et al.* , 2013] (version 2.0.6 et bowtie2 [Langmead & Salzberg, 2012] version 2.0.2) suivi du compteur de lecture htseq-count (version 0.5.3p9 et python 2.7), sur le génome de la souris (ensembl NCBIM37.67 release 67 de Mai 2012). Puis j'ai effectué une analyse de gènes différentiellement exprimés entre le côté lingual et le côté buccal et j'ai mis en évidence qu'il y avait toujours deux fois plus de gènes différentiellement exprimés entre le côté buccal et le côté lingual de la molaire supérieure que de la molaire inférieure (voir chapitre 10, supplementary text).

### 4.7.2 Dispersion des librairies RNA-seq

Les librairies de molaires inférieures sont plus dispersées que les librairies de molaires supérieures, en d'autres termes, les réplicats de la molaire supérieure sont plus similaires entre eux que les réplicats de la molaire inférieure. Ainsi, cette différence de dispersion va biaiser l'estimation des gènes différentiellement exprimés entre le côté lingual et le côté buccal dans les 2 types de molaires analysées séparément. En effet, les librairies présentant une faible dispersion vont bénéficier d'une puissance statistique plus forte et ainsi le nombre de gènes différentiellement exprimés (DE) sera plus important, même si il n'y a pas de différences biologiques entre les deux conditions. Inversement, les librairies présentant une plus forte dispersion vont avoir un nombre

plus faible de gènes DE dû à la diminution de la puissance statistique associée à l'analyse.

On s'est alors demandé si on avait une réplicabilité identique entre molaire inférieure et molaire supérieure, est-ce qu'on aurait le même nombre de gène DE? Pour cela, j'ai extrait les informations de dispersion estimées par DESeq2 au cours de l'analyse des gènes DE BL pour la molaire supérieure et la molaire inférieure séparément. Ainsi j'ai pu utiliser la dispersion estimée par DESeq2 pour artificiellement augmenter (utilisation de la dispersion de la molaire inférieure) ou diminuer (utilisation de la dispersion de la molaire supérieure) la dispersion des librairies. Puis avec ces nouvelles valeurs de dispersion, j'ai effectué l'analyse de gènes différentiellement exprimés et enfin j'ai comparé le nombre de gènes DE obtenu lorsque les dispersion sont similaires.

## 4.8 Analyse GO de listes de gènes

*Ces analyses ont été réalisées pour le chapitre 7 / Annexe A ainsi que le chapitre 10.*

Les analyse GO ont été réalisées avec GOrilla (http://cbl-gorilla.cs.technion.ac.il/) et topGO [Alexa & Rahnenfuhrer, 2016]. J'ai réalisé l'analyse topGO [Alexa & Rahnenfuhrer, 2016] (version 2.24.0) qui permet d'effectuer une analyse d'enrichissement des termes GO en prenant en compte la structure emboitées de GO. Un des problèmes de analyse de GO est la visualisation ainsi que la comparaison des résultats entre analyses séparées, par exemple dans notre cas où nous voulions comparer la liste les termes GO enrichis du côté buccal à la liste des termes GO enrichis du côté lingual. Certains GO étaient communs entre ces deux listes et nous voulions savoir si l'enrichissement était plus important pour l'un ou l'autre côté, j'ai donc utilisé un test de $\chi^2$ pour tester si l'enrichissement est statistiquement plus fort d'un côté que de l'autre. Enfin, pour donner une représentation synthétique et objective des résultats de l'analyse GO, j'ai développé un outils de visualisation sous R, qui est une fonction me permettant de visualiser les résultats de l'analyse pour une sélection de termes GO (définie par l'utilisateur). Nous avons souhaité que cette visualisation comprenne les informations concernant la significativité du terme, la taille du terme (soit le nombre de gènes composant le terme) et la valeur l'odd ratio (ou rapport des chances, qui permet d'estimer/d'évaluer l'enrichissement du terme GO ainsi que significativité). Ces données sont représentées sous forme d'un graphique en bâtons, où la significativité du terme est représentée par la couleur du bâton, la taille du terme est proportionnelle à la largeur du bâton et enfin l'enrichissement est proportionnel à la longueur du bâton. De plus, dans le cas où il s'agit de comparer 2 listes de termes GO, une étoile rouge marque l'enrichissement plus fort d'un terme dans une condition (par exemple entre côté buccal et côté lingual). Nous avons aussi appliqué une approche systématique qui vise a comparer les termes un à un entre le côté buccal et le côté lingual par exemple pour l'ensemble des termes GO contenant les mots *sang* et/ou *vasculogenèse* (figure 4.4). De la sorte, nous avons pu avoir une idée plus précise des processus qui étaient plus représenté d'un côté que de l'autre.

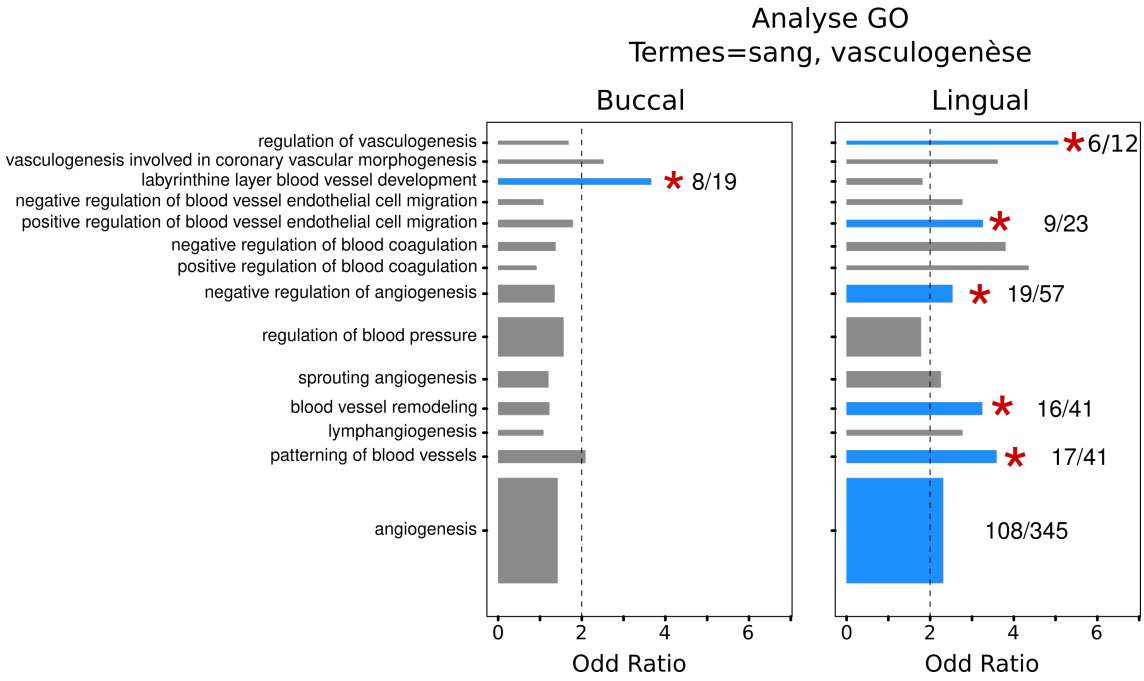



FIGURE 4.4 – **Visualisation des résultats de l'analyse GO.** Termes GO enrichis (*pvalue* < 0.05) en buccal et/ou en lingual contenant les mots *sang* et/ou *vasculogenèse*. La largeur des barres est proportionnelle au nombre de gènes contenus dans le terme GO. La longueur de la barre représente la valeur du rapport des chances (ou odd ratio). La couleur des barres reflète la significativité du terme, beu : termes significativement enrichi (test de Fisher réalisé par topGO, *pvalue*< 0.05) et gris : non-significatif. L'étoile rouge représente l'enrichissement plus fort d'un terme dans une côté que dans l'autre (test $\chi^2$, *pvalue*<0.05).

## 4.9 Estimation des proportions de tissus par déconvolution

*Cette méthode à été utilisée dans le chapitre 10.*

### 4.9.1 Le principe

L'expression des gènes dans un échantillon biologique (par exemple un organe) provient souvent d'un mélange hétérogène de plusieurs cellules ou de plusieurs tissus dont les proportions sont variables, chacun contribuant au niveau d'expression global. Si on prend l'exemple de la dent, le modèle d'étude de ce travail, cette dernière est composées de trois populations cellulaires majeures : l'épithélium, le mésenchyme et l'épithélium qui commence à se différencier en cuspides. De ce fait, cette hétérogénéité est un facteur confondant majeur dans l'analyse des gènes différentiellement exprimés entre deux organes par exemples, où des différences dans les proportions relatives de cellules constituants l'échantillon peuvent biaiser la détection des différences d'expression entre échantillons. La déconvolution est l'estimation des proportions tissulaires d'un échantillon basée sur ses niveaux d'expression. On distingue deux grands types

de déconvolution. La déconvolution partielle et la déconvolution complète. La déconvolution complète permet, à partir de données d'expression des tissus purs, d'estimer la proportion de chaque tissu ainsi que d'estimer le niveau d'expression de chaque tissus de l'échantillon. C'est à dire que l'on peut estimer la participation de chaque tissu au niveau d'expression global [Gaujoux & Seoighe, 2013; Gaujoux, 2013]. Dans cette thèse, j'ai utilisé la déconvolution partielle avec des marqueurs spécifiques de chaque tissu. Cette méthode permet simplement d'estimer la proportion de chaque tissu dans l'échantillon. Voir la figure 4.9 pour une explication graphique de la déconvolution partielle à l'aide de gènes marqueurs. Je n'ai pas encore réalisé les déconvolutions globales, car nous ne disposions pas encore de données RNA-seq de tissus purs (les données brut de séquençage sont arrivées quelques semaines avant la fin de ma thèse). Or pour réaliser des déconvolutions complètes il faut que les jeux de données des tissus purs et des échantillons à déconvoluer soient similaires.

### 4.9.2 Définition des marqueurs

J'ai réalisé les déconvolutions à partir de différents groupes de marqueurs sur les échantillons du jeux de données BL ainsi que sur les échantillons des jeux de données large échelle de souris et de hamster. J'ai réalisé ces déconvolutions à l'aide du paquet CellMix [Gaujoux & Seoighe, 2013; Gaujoux, 2013] et de la méthode PDSA [Zhong *et al.* , 2013]. J'ai utilisé trois sources pour définir ces jeux de marqueurs. (1) Les données de la base *bite-it* (http://bite-it.helsinki.fi/) qui regroupe des données de la littérature de gènes exprimés au cours du développement de la molaire et dont le patron d'expression (au moins à un stade de développement) a été caractérisé par hybridation *in situ*. (2) Les données de puces d'O'Connell *et al.* [2012] qui se composent des niveaux d'expression d'épithélium et de mésenchyme de bourgeons dentaires à des stades précoces (ED12.5, ED13 et ED13.5) ainsi que le niveau d'expression dans le noeud d'émail à ED14.5. (3) Enfin, des données RNA-seq d'épithélium et de mésenchyme de bourgeon de molaire inférieure et supérieure au stade ED15.5.

Pour les données de puces [O'Connell *et al.* , 2012], j'ai calculé un ratio médian d'épithélium : $ratioEpi$, un ratio médian de mésenchyme : $ratioMes$ et un ratio médian de noeud d'émail (EK) : $ratioEK$ défini de façon suivante (voir équation 4.1) où $t$ sont les différents stades de développement (ED12.5/ED13/ED13.5).

$$
\begin{aligned}
ratioEpi &= median(\frac{epi_t}{mes_t}, \frac{epi_{13.5}}{ek_{14.5}}) \\
t \in \{12.5, 13, 13.5\} \qquad ratioMes &= median(\frac{mes_t}{epi_t}, \frac{mes_{13.5}}{ek_{14.5}}) \qquad (4.1) \\
ratioEK &= meian(\frac{EK_{14.5}}{epi_t}, \frac{EK_{14.5}}{mes_t})
\end{aligned}
$$

Pour les données de RNA-seq j'ai défini les marqueurs de l'épithélium et du mésenchyme en utilisant différentes valeurs de centiles d'enrichissement (supérieur à 95% ou 99% des enrichissement pour chaque tissu) pour lesquels j'ai vérifié que epi > 2 mes et mes > 2 epi. Pour l'épithélium j'ai aussi imposé que la $ratioEK$ (tiré des données O'Connell *et al.* [2012]) soit inférieur à 1.1 (voir figure supplémentaire B du chapitre 10 pour la définition du seuil EK).

#### 4.9.2.1 Remarques sur la méthode de déconvolution

Nous avons pu observer que la valeur de la proportion estimée par déconvolution est sensible au groupe de marqueurs utilisé pour réaliser cette déconvolution. C'est pourquoi par la suite nus avons utilisé un approche par bootstrapping des la liste de marqueurs. Cependant, les proportions relatives entre compartiments tissulaire ainsi que la dynamique temporelle (pour les données large échelle de souris) sont communes entre les différents groupes de marqueurs.

#### 4.9.2.2 Ré-échantillonnage des marqueurs par jacknife

J'ai aussi observé que les marqueurs doivent être strictement spécifique au compartiment tissulaire dont on veut estimer la proportion. Notre modèle d'étude est un cas particulier dans la mesure où le mésenchyme est l'épithélium, en dialogue permanent au cours de l'odontogenèse, partagent des voies de signalisation qui sont spécifiques d'un tissu à un stade de développement donné mais plus à un autre. De plus, le nœud d'émail est une structure provenant de l'épithélium et donc des marqueurs spécifiques du nœud d'émail (à partir du stade ED14.5) peuvent être des marqueurs de l'épithélium à des temps précédant l'apparition du nœud d'email. C'est pourquoi j'ai utilisé un ré-échantillonnage par jacknife des marqueurs d'épithélium, c'est a dire en retirant 1 à 2 marqueurs des listes de marqueurs d'épithélium défini à l'aide des donnée de puces [O'Connell *et al.* , 2012].

## 4.10 Définir un seuil d'émaillitude d'un gènes buccal et identifier des nouveaux marqueurs potentiels du nœud d'émail

*Cf. chapitre 10.*

Le nœud d'émail est le centre de signalisation qui permet la mise en place des cuspides. Un certains nombre de gènes sont connus pour être exprimer dans les nœud d'émail au cours du développement, tel que *Bmp4*, *Fgf4*, *p21* (*Cdkn1a*), ... [Jernvall *et al.* , 1994, 1998]. Nous avons voulu tirer parti de notre jeu de données bucco-lingual afin d'identifier de nouveaux marqueurs potentiel du nœud d'émail (EK). En effet, au stade auquel nous avons échantillonné le jeu de données BL, le stade ED15, il y a la présence d'un nœud d'émail uniquement du côté buccal. Ainsi, au milieu des gènes exprimés avec un fort biais buccal (car pas de présence de EK côté lingual), il doit y avoir des marqueurs de nœud d'émail. Afin de faire le tri au milieu des ces gènes fortement biaisé buccalement, j'ai utilise des données d'expression de tissus pures : EK au stade

ED14.5 et épithélium aux stade ED13/ED13.5 [O'Connell *et al.* , 2012]. J'ai utilisé le $ratioEK$ défini plus haut (*p. 78*). Pour chaque valeur de $ratioEK$ j'ai compté, soit de manière cumulative soit de façon ponctuel pour chaque valeur du $ratioEK$, le nombre de gènes qui sont biaisés du côté buccal ou du côté lingual avec un enrichissement d'au moins deux fois ainsi que les gènes non différentiellement exprimés. J'ai ensuite pu représenter graphiquement ces données (voir figure supplémentaire B du chapitre 10). Ceci nous a permis de définir des seuils "d'émaillitude" pour lesquels nous avions un bon compromis entre la spécificité d'un gènes et la sensibilité avec laquelle nous avons pu identifier ce gènes.

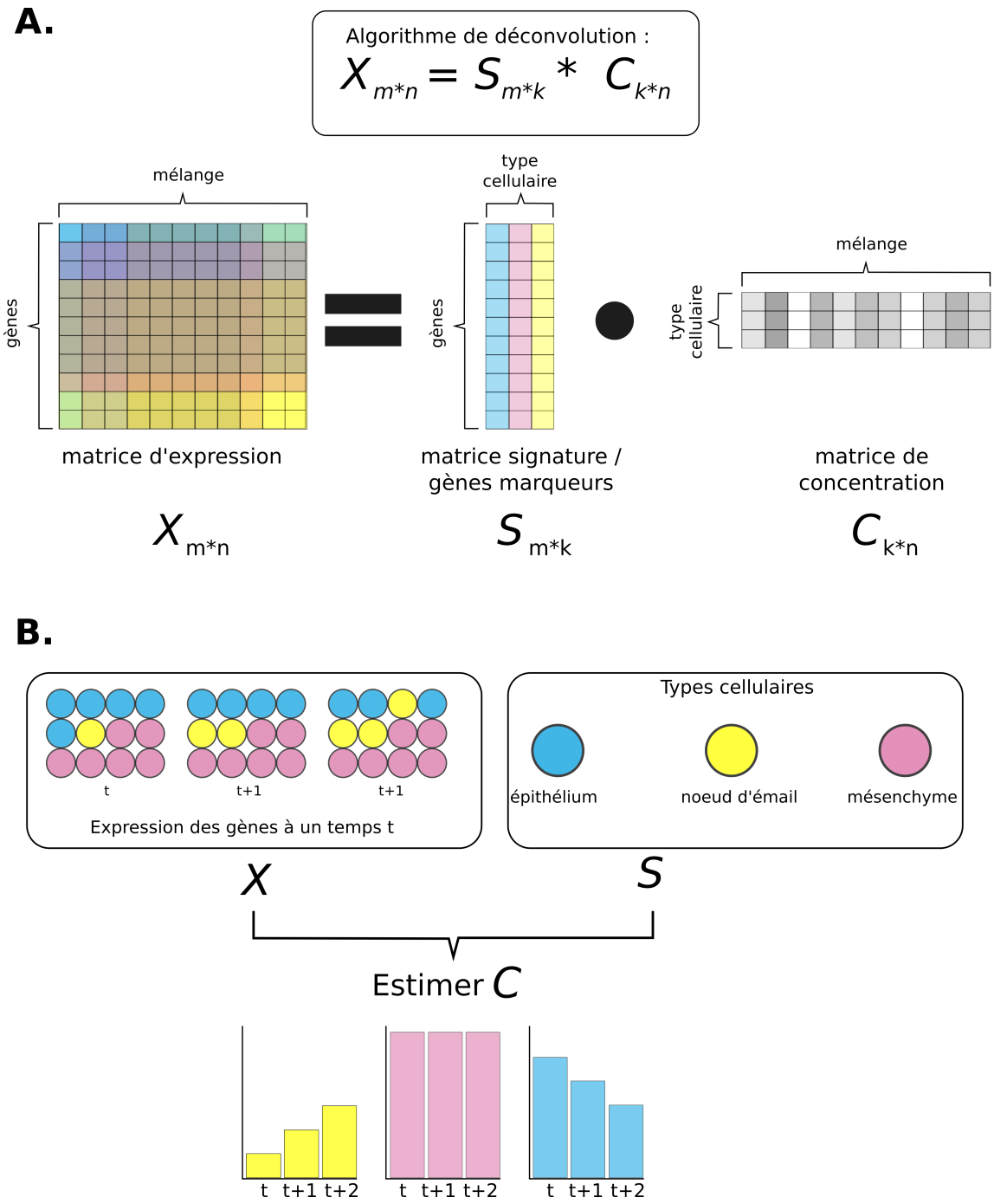



FIGURE 4.5 – **Estimer des proportions tissulaires. A.** Représentation théorique du principe des déconvolutions visant à estimer des proportions tissulaires. La matrice des niveaux d'expression ($X$) est la combinaison linéaire de la matrice de signature ($S$) et de la matrice de concentration ($C$). **B.** Représentation schématique d'un jeu de données de bourgeon de molaire à trois stades de développement différents. Connaissant le niveau d'expression global dans la molaire (matrice $X$) ainsi que les gènes marqueurs des différents tissus qui composent la molaire (matrice $S$), les proportions (matrice $C$) de chaque tissu dans l'échantillon de départ peuvent être inférés par déconvolution.

# 5 Microscopie 3D

*Cette méthodologie se rapport au chapitre 10*

J'ai réalisé la mise au point du protocole permettant de réaliser les reconstructions 3D des germes de molaires, dont le principe général est représenté figure 5.1. D'abord j'ai réalisé l'immunolocalisation *in toto* des bourgeons de molaires avec l'anticorps anti-*laminin α5* afin de marquer la lame basale qui est à l'interface entre l'épithélium et le mésenchyme. Ensuite, j'ai inclus ces bourgeons dans l'agarose low-melting afin d'orienter le germe correctement. Puis j'ai réalisé une étape de clarification au BA/BB (benzyl alcool / benzyl benzoate) afin de rendre transparent les échantillons pour permettre leur acquisition par un microscope à fluorescence. Enfin j'ai utilisé un microscope biphoton (qui reduit le photo-blanchiment) pour réaliser l'acquisition des images en sections transversales. Ensuite j'ai testé deux logiciels d'imagerie pour la reconstruction 3D des bourgeons imagés : le logiciel Imaris (http://www.bitplane.com/) et le logiciel Amira (https://www.fei.com/software/amira-3d-for-life-sciences/). Nous avons retenu le logiciel Amira qui était plus approprié à nos besoins.

Bien que ce protocole permettait de reconstituer correctement la forme de l'épithélium aux stades ED15 et ED15.5, nos données RNA-seq large échelle de souris ont montré que la *laminin α5* est différentiellement exprimée au cours de la cuspidogenèse, ce qui signifie des fluctuations de l'intensité du marquage en fonction du stade de développement. Ceci n'était pas compatible avec notre but de réaliser ces reconstructions 3D pour la totalité de la cuspidogenèse. De plus, le marquage de la lame basale est assez irrégulier ce qui empêchait une reconstruction quasi automatisée, demandant une intervention manuelle de l'opérateur sur la totalité des coupes utilisées pour la reconstruction 3D. Par la suite, c'est l'ingénieur de l'équipe qui a pris le relai sur ces expériences, il a résolu ce problème de reconstruction par l'utilisation d'un mélange d'anticorps anti-cytokératines pour marquer l'épithélium (et non plus la lame basale par anticorps anti-*laminin α*). Les résultats relatifs à ces reconstructions 3D se retrouvent dans le chapitre 10 ainsi que l'annexe A (figure supplémentaire 4).

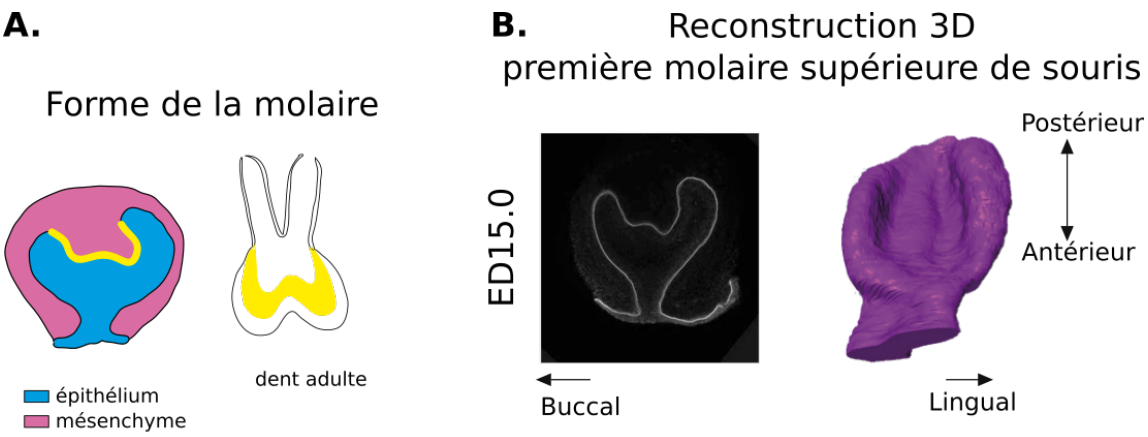



FIGURE 5.1 – **Principe de la reconstruction 3D de la forme de la molaire.**
**A.** La forme à l'interface entre l'épithélium et le mésenchyme va donner la forme de la molaire au stade adulte (représenté en jaune). **B.** Le marquage de la lame basale par la *laminin α5* permet de délimiter l'interface entre épithélium et le mésenchyme. Ce marquage nous a permis de reconstruire la forme 3D de la partie épithéliale du bourgeon de molaire à l'aide du logiciel Amira.

# 6 *Bmper* : gène candidat de l'asymétrie bucco-linguale

*Matériels et méthodes du chapitre 11.*

## 6.1 Hybridation *in situ* : caractérisation du patron d'expression de *Bmper*

J'ai commencé à caractériser l'expression de *bmper* chez la souris pas *in situ in toto*. Pour cela, j'ai cloné une sous-partie du transcrit *Bmper* chez la souris afin de réaliser un sonde pour les hybridations *in situ*. En parallèle, j'ai disséqué des germes de molaire à différents stades (ED15,ED15.5 et ED16) puis j'ai réalisé des hybridation *in situ* pour *Bmper* ainsi que pour *Fgf4* afin d'avoir un contrôle positif de la manipulation et aussi pour pouvoir marquer la position des noeuds d'émail. J'ai observé le patron d'expression de *bmper* au stade ED15.5 sur germe entier. J'ai ensuite inclus ces germes de molaires ayant subi l'hybridation *in situ* en paraffine afin de réaliser des coupes pour effectuer une une meilleure caractérisation du patron spatial d'expression. Pour contrôler l'orientation des germes, j'ai pré-inclus en agarose, puis j'ai déshydraté le tout avant de l'inclure en paraffine et d'effectuer les coupes au microtome. Cependant, le processus de déshydratation (en vu de l'inclusion en paraffine) lavait trop le signal de l'hybridation *in situ in toto*. Pour pallier à ce problème il faudrait mettre au point une *in situ* sur coupe cryostat.

## 6.2 Mutants *bmper* et coupes histologiques

Nous nous sommes demandé si *bmper* jouait un rôle dans le développement de la molaire chez la souris. Comme il existait déjà des mutants de *Bmper* chez la souris, nous avons collaboré avec Pamela Lockyer du laboratoire de Cam Patterson qui nous a fourni le matériel (adultes et embryons) pour caractériser la morphologie dentaire des souris mutantes *Bmper*, dont la

mutation a été obtenu par délétion, créant l'allèle $Bmper^{tm1Cpat}$ [Kelley *et al.* , 2009]. Les mutants *bmper* homozygotes meurent à la naissance car ils ne peuvent pas respirer (malformation de la trachée). Je n'ai donc pas pu travailler directement avec la molaire minéralisée des homozygotes. J'ai cependant pu photographier et comparer la dentition des mutants *bmper* hétérozygotes avec celle de la souche sauvage, mais je n'ai pas observé de différences notoires avec la dentition de la souche sauvage. En parallèle j'ai travaillé sur des embryons au stade ED18.5 pour la souche sauvage, hétérozygote et homozygote mutante. Nous avons choisi le stade ED18.5 car il correspond à la fin de la mise en place des cuspides. De plus, à ce stade le processus de minéralisation n'a pas encore commencé ce qui permet l'analyse en coupes histologiques. J'ai donc reçu les échantillons en Bouin (traitement qui permet de déminéraliser et de fixer les tissus), que j'ai déshydraté et inclus en paraffine. À l'aide d'un microtome, j'ai ensuite réalisé des coupes histologiques (de 6 $\mu$m d'épaisseur) que j'ai coloré au trichrome de Masson afin de mettre en évidence les différents tissus de la dent, puis j'ai analysé ces coupes au microscope. J'ai ensuite photographié des séries complètes (1 section toutes les 3 sections) pour observer et caractériser la morphologie embryonnaire (pour 5 homozygotes mutés / 1 hétérozygote / 3 sauvages).
Dans le but de reconstruire la morphologie 3D du bourgeon de molaire chez des embryons homozygotes *bmper*, j'ai aussi testé, sur des échantillons de souris sauvages, un traitement PTA (Phospho Tungstic Acid) afin de contraster les tissus mous pour permettre d'imager ces tissus mous par micro-tomographie à rayons X (microCT scan) [Metscher, 2009; Hoorebeke, 2013]. Cependant, cette approche n'a pas été concluante due à la difficulté de la reconstruction du bourgeon de molaire qui est encapsidé dans le cartilage de la mâchoire et qui ne permet pas une reconstruction simple de la morphologie du bourgeon et un travail plus important de mise au point serait nécessaire.

# Part III Résultats

# 7 Transcriptomic signature of serial organs : the mouse first upper and lower molar

**Introduction**

How to link differences of the developmental program to changes in morphology? We tackle this question at the whole transcriptomic level in the mouse first upper and lower molars.

The molars are serial organs, meaning that they are the same module patterned in different location of the body. In the case of upper and lower molar they deployed in the upper and the lower jaws. As already discussed in the introduction of this thesis (see chapter 3, *p. 41*), the upper and the lower jaw identities are defined around ED8.25-ED9 by homeobox genes. For example *Dlx5/6* are essential in the first pharyngeal arch to determine the identity of the lower jaw [Depew, 2002; Beverdam *et al.* , 2002]. Afterwards, the odontogenic homeobox code is known to regulate the type of the tooth [Cobourne & Sharpe, 2003; Catón & Tucker, 2009]. For example, regions of the jaw expressing *Dlx1/2* [Qiu *et al.* , 1997; Thomas *et al.* , 1997] and *Barx1* [Tucker, 1998; Miletich *et al.* , 2005] are the presumptive regions where molars will develop.

An interesting feature about mouse first molars is that they present extreme divergent morphologies between the upper and the lower jaw. Thus the mouse first upper molar develop two supplementary cusps. How to increase the number of cusps patterned during development? Experiments of tissue recombination show that cusp patterning is tightly linked to mesenchyme identity and proportion [Schmitt *et al.* , 1999; Hu *et al.* , 2006; Cai *et al.* , 2007; Ishida *et al.* , 2011]. Gene-centered studies have already identified some genes and pathways responsible for the patterning of supplementary cusps [Tucker *et al.* , 2004; Kassai *et al.* , 2005; Charles *et al.* , 2009a; Harjunmaa *et al.* , 2012] (see chapter 3, *p. 52* ). However, most of these gene-centered studies were done in mouse lower molars only and were not interested in identifying the developmental bases responsible for the increase number of cusps in mouse upper molar.

In the lab we are interested in how the mouse upper molar is able to develop these two supplementary cusps. Answering this question implies comparing developmental programs of upper and lower molars. Historically this was done at a gene by gene basis. However, the rise of microarrays and RNA-seq techniques enable to address the developmental program at the scale of the whole transcriptome. The transcriptomics studies already performed in mouse were either focusing on lower molars development [Landin *et al.* , 2015] or in a single developmental stage [Laugel-Haushalter *et al.* , 2013], or were focusing on earlier stages [O'Connell *et al.* , 2012; Jia *et al.* , 2012]. The novelty of our study is to survey both the development of mouse first upper and lower molars at the transcriptomic level during crown morphogenesis, when cusps are patterned sequentially from ED14.5 to ED18 (sampled every 0.5 days). We proposed to study cusps patterning in both upper and lower molars at the whole transcriptomic level first to gain insight into the transcriptomics signatures that characterize crown morphogenesis and second to compare the development programs between upper and lower molars to identify the modifications of the developmental program that enable the patterning of two supplementary cusps in mouse upper molar. Most transcriptomics studies comparing developing embryos or organs looked for the main pattern of variation in transcriptomes, which are typically extracted by multivariate analysis and identified different species or organs through developmental time [Yanai *et al.* , 2004; Liao & Zhang, 2005; Brawand *et al.* , 2011; Lin *et al.* , 2014; Gilad & Mizrahi-Man, 2015]. As discussed by Breschi *et al.* [2016] the transcriptome is made of different signatures. To go beyond the identification of which transcriptomics signature prevails, it is of interest to identify the patterns of expression responsible for each signature and to quantify the relative contribution of those signatures to the final transcriptome. Such a characterization would enable to efficiently compare expression profiles between homologous organs or between species and tackle questions such as : how differences in transcriptomics signatures enlighten us of the developmental basis of morphological disparities between homologous organs?

**Importance of cellular proportion in organ identity**

In this article we aimed to identify the transcriptomic differences that underly the development of mouse first upper and lower molar during crown morphogenesis. We found that part of the upper/lower transcriptomic differences are explained by three transcriptomics signatures that are shaped by relative abundance of cellular type.

First, differences in epithelial/mesenchymal relative proportion participate to differences between upper/lower molars, together with differences of expression within tissue, at least in the mesenchyme. Second, a significant heterochrony explained part of the upper/lower differences. This heterochrony could be modeled by the increase in proportion of the tooth germ occupied by cusp tissue. Third, an early-mid (ED15.5-ED16) transient excess of upper/lower differences have been identified. Analysis of the genes responsible for this transient excess suggested that the upper molar experience exaggerated morphogenetic processes (with less mitotic cells and more migrating cells).

Our analysis suggested that gene expression in a complex tissue is not only the result of trans-

criptional regulation but also the result of tissues proportions (see figure 12.1 in the Discussion section, *p.* 186). The differences in cellular proportions in organ identity is often overlooked, however it provides valuable insight in understanding the cellular processes responsible for differences in organ development, as exemplified also during limb development (*cf.* chapter A figure 7).

**Participation to this work**

The article corresponding to this work has been accepted for publication in Genome Biology. Even though my participation (detailed below) is minor in this analysis, I present the full manuscript (in Appendix A), because it is not yet available online, and I figured that the results may be useful for the understanding of the other sections of the manuscript.

For this work, I developed a visualization tool for GO analysis to overcome one of the problem of GO analysis which is to compare results between two different GO analyses, for example one obtained for genes differentially expressed in lower molar versus one obtained for genes differentially expressed in upper molar (see later chapter 10). I made a visualization tool to look at specific subset of GOTerms (chosen by the user ). Each plot include information on the size on the GOTerm, its odds ratio in the dataset analyzed and its statistical significance. Size, or number of genes in the GOTerm, is proportional to the width of the bar, enrichment in the data set, the odd ratio, is proportional to the length of the bar. An bar are color coded to represent the significance of the GOTerm in the dataset. I coupled it with a pairwise comparison of GOTerms (based on $\chi^2$ test) to find if one GOTerm is more enriched in one condition than in another one, and represent the statistical significance of an enrichment with red stars. I used this tool to perform the Supplementary figure 2 of the paper presented in appendix A. I re-use this visualization tool for my bucco-lingual analysis of upper and lower molars in chapter 10.

# 8 Temporal signal in transcriptome of developing organs

**Temporal signal in transcriptomes throughout development**

Several studies analyzing transcriptomes during development found a temporal ordering of samples when performing multivariate analyses [O'Connell *et al.* , 2012; Levin *et al.* , 2012; Anavy *et al.* , 2014; Tschopp *et al.* , 2014]. For example, Principal Component Analysis (PCA) of Xenopus embryos transcriptomes revealed that samples ordered according to their stages of development, as illustrated in figure 8.1 [Anavy *et al.* , 2014]. Levin *et al.* [2012] note that adjacent stages have more similar expression patterns than temporally distant stages, thus reflecting the progression of development.

The difficulty when working with developmental samples is to be able to stage them correctly. Typically, stages are defined by morphological characters, such as the advancement of limb development in tetrapods. Finer staging in model organisms such as mice can also be inferred from weight. For example, for mice embryos within the same litter, which are supposed to be at the same development stages, exhibit differences in weight that are correlated with their stages of development [Peterka *et al.* , 2002]. There are a lot of cases when the morphological staging is not possible, for example between two closely sampled developmental stages, or at stages when there are no morphological milestones to be exploited. Even the weight staging in mice embryo is not satisfying between litters.

Anavy *et al.* [2014] used the transcriptomic signature of time present in timeseries transcriptomes of whole embryos to order transcriptomic samples according to their stages of development. They develop a method named BLIND, Basic Linear Index Determination of transcriptomes, to order transcriptomic samples of developing embryos. This method takes into account the distance between every two samples on the PCA axis corresponding to time ordering, together with the use of a salesman algorithm to find the shortest path between consecutive samples. However, it is poorly known what kind of expression differences are responsible for this strong temporal signal.

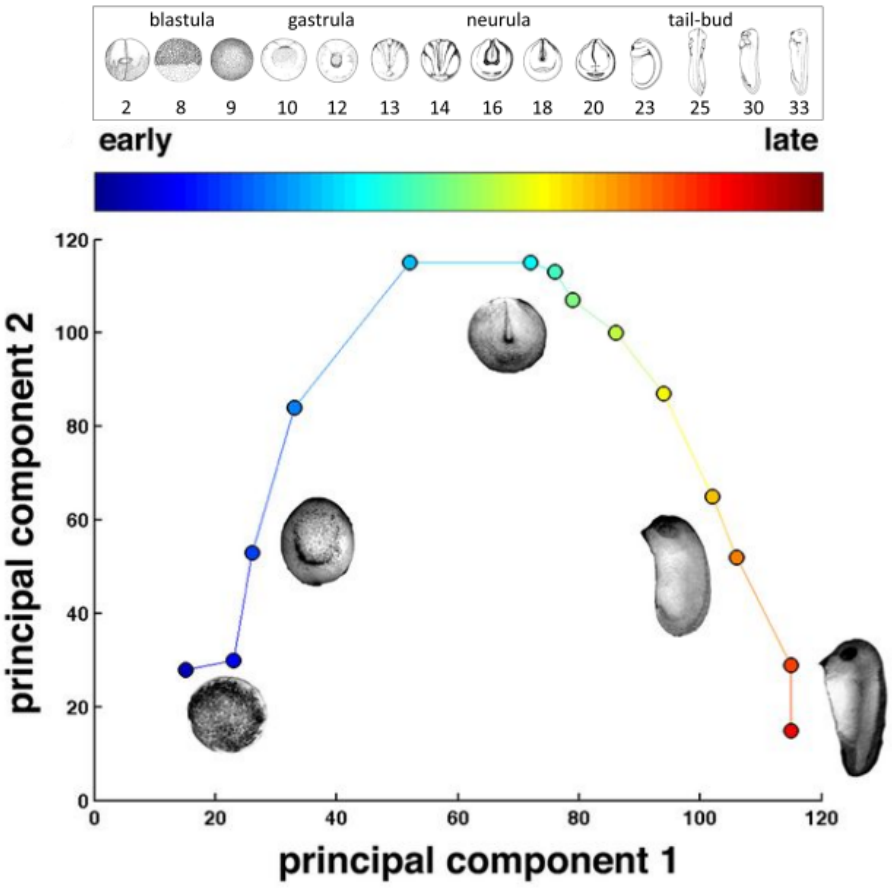



FIGURE 8.1 – **Temporal ordering of timeseries transcriptomes**.
Principal Component Analysis of Xenopus embryos (*Xenopus laevis*) at 14 stages of development. See the ordering of the transcriptomes according to their stages of development on the principal component 1, revealing a strong temporal signal in transcriptomes of developing embryos. (From Anavy *et al.* [2014])

In this work we investigated the temporal signal at three different levels : (1) the whole molar germ level at a relatively large time scale, from ED14.5 to ED18.0 sampled every 0.5days ; (2) the whole molar germ level at a small scale, between samples considered as biological replicates and coming from three different litters and (3) in half molar germs, buccal and lingual sides, at one stage of development (ED15). We were able to identify a temporal signal for these three different levels. Moreover, we showed that the temporal signal is carried by subtle changes between stages, notably we found that this temporal information is encoded in linearly increasing/decreasing genes. Finally, we evidenced that the temporal signal indirectly record changes in the proportion of differentiated cells, in our case epithelial cusp tissue. In chapter 9 we used this temporal signal to homologize developmental stages between mouse and hamster molars in order to compare the dynamics of their developmental programs at the transcriptomic level.

**Participation to this work** The work presented in this section corresponds to one of the main project during my thesis, the other one being presented chapter 10, *p.*145, on bucco-lingual asymmetry during mouse molar development. This analysis stems from the observation that the projection of buccal and lingual samples in the mouse large scale PCA clearly separate buccal and lingual sample on the PC1 axis, which is a temporal axis. These data suggested that buccal and lingual side mature at a different pace. Then we decided to combine different datasets that were available in the lab, to investigate further the accuracy, precision, and biological underpinnings of time signals in transcriptome.

I drafted the first version of the manuscript presented below, and prepared all figures and performed most of the analyses. The text was next amended and modified by my supervisors, and we hope to submit it early 2017.

In this work, I performed the alignment and the estimation of expression levels of the presented data. I performed the multivariate analyses and the partial deconvolution analysis. I used DESeq2 [Love *et al.* , 2014] to extract the "timer genes", which are genes that linearly increase/decrease over time, with at least a two-fold change between DE14.5 and ED18. I also computed the correlations of expression levels between our small scale dataset and large scale dataset. I defined and isolated the "linear" genes in the small scale dataset. I used the "timer" genes to estimate the "age" of the buccal/lingual samples.

-DRAFT -

# Tracking temporal signal in developmental transcriptomes.

December 19, 2016

Coraline Petit[1], Marie Sémon[1], Sophie Pantalacci[1]
[1] *UnivLyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, Laboratoire de Biologie et Modélisation de la Cellule, 15 parvis Descartes, F-69007, Lyon, France*

**Abstract**

Developmental transcriptome studies are challenged by sample staging, whether from whole embryos, or organs. Recently, methods have been proposed to order the samples in a path representing the progress of embryonic development. This was based on the position of the samples in a principal component analysis, or on the identification of stage-associated marker genes. The ordering of the transcriptomes of broadly sampled embryos is easily understandable because many genes are switched on and off between consecutive stages. Here we are interested in the nature and sensitivity of the temporal information at much more fine scale, using mouse molar germ as a model. Molar development involves the sequential addition and maturation of cusps (hills at the surface of the molar crown) from the buccal to the lingual side. Whole tooth germ carry a temporal signal, sufficient to discriminate very small developmental intervals, through many linearly increasing/decreasing genes, but also genes with unimodal temporal profiles. Tooth germs dissected along the bucco-lingual axis also carry a temporal signal, with buccal side looking developmentally "older" than the lingual side. This is more easily explained than appears at first glance: it can be accounted for by differences of expression levels in linearly increasing/decreasing genes, and by tissue proportions, as exemplified by the proportion of cusp epithelium. Thus, ordering of consecutive developmental samples appears because many genes show variations that are subtle, yet consistent between consecutive stages which would be classically considered developmental duplicates. This very fine sample ordering is a way of monitoring differentiation, and reminds us that the continuous nature of development can be seen in transcriptomes.

## Introduction

Connecting progress into embryonic development to underlying gene expression changes is a fundamental question in developmental biology. Determining the relative development age of each sample, through morphological and expression data, is a crucial step in this process. Recently, it was discovered that it is possible, from the whole transcriptome (expression of all genes) of a developmental sample, to extract time-related information. This stage-related information is embedded in developmental transcriptomes, but can be discovered in different manners. Multivariate techniques, such as principal component analysis [1] or canonical correlation analysis (CCA) are very often used to explore the main sources of variation in gene expression. Using such methods, many studies have shown that successive developmental time points are placed in a consistent chronological order on the first principal components [2, 3, 4, 5]. The location of samples in the PCA two-dimensional space has even been considered a metric for transcriptomic state, such that the distance between transcriptomes in this space can be taken as measure of the developmental distance between the samples [6, 3, 4, 7]. Such an ordering was also recovered, perhaps less unexpectedly, using sets of stage-specific genes. This reveals that expression is more similar between adjacent stages than between temporally distant stages, reflecting the progression of development. This previous work has been focused on broadly sampled stages. In this case, transcriptomes register the drastic transformations during such largely spaced developmental stages, characterized by many sets of genes that are switched on and off between consecutive stages. In such case, multivariate analysis will tend to group stages with most expressed genes in common, and leading, from one stage to the next, to a time-ordering of the samples. In line with this idea, interspecific comparisons have discovered conserved "developmental milestones" [3], which are very distinct embryonic stages, characterized by their conserved expression dynamics and activation of key developmental regulators. So at this scale, development at the gene regulatory level reflects the stage-like processes occurring at the macroscopic level. In this view, intermediate stages between the milestones may be seen as

transitions.
What about closely sampled developmental stages? Recent work hints that time information is pervasive even in closely related stages : a meta-analysis showed that 50% of the studies has a developmental component, which is unintentional in 26% of all cases [8]. Self-ordering of individual cells in differentiation gradients is also noted in a growing literature of single cell transcriptome analyses. But the way this information present at the level of the cell translates in a time information at the level of the organ is unknown. Do all cell types / tissue compartments in an organ carry a time information? If yes, is it a congruent time related information? How can this help better understanding the processes acting during development? Although many previous analyses show that there is a strong temporal component in the transcriptomes of developmental samples, there is little insight into what is carrying this temporal signal, so the answer to these questions is unknown.
We choose to ask these questions using the transcriptomes of the first molars in mouse. The molar germ is an appropriate model for the analysis of time signal in a complex developing organ, because it is a relatively simple and independent organ, which is arranged in two compartments, epithelium and ectomesenchyme (herein called mesenchyme for simplicity). A step-by-step description of first molar development is found in Figure 1. Key to the process of crown morphogenesis are epithelial signaling centers called Enamel Knots [9, 10]. The Primary Enamel Knot (PEK) directs the folding of the epithelium around the condensing mesenchyme, defining the molar. Later, the Secondary Enamel Knots (SEK) direct further foldings of the epithelium, defining the cusps. These enamel knots are patterned sequentially (from ED16.0) while the molar germ grows [9, 10]. They also act as epicenters for the progression of proliferation arrest and differentiation. This iterative cusp addition permits to finely follow the advancement of morphogenesis. For the first lower molar, spatio-temporal gene expression data during these different periods of molar development are reported in the "bite-it" database, and transcriptomic data were generated with microarray technology [2, 11] and in our lab, with RNA-seq data. In this previous study, we followed transcriptomic dynamics during a period of tooth morphogenesis that encompasses the iterative formation of several cusps. We obtained closely spaced RNA-seq transcriptomes with 8 developmental stages for upper and lower first molars in mouse, sampled every 12 hours (stars on the Figure 1). With this data we have shown (Pantalacci *et al.*, submitted [12]) that the transcriptome of a whole tooth germ carries a temporal signal, which correlates with the proportion of cusp epithelium in the whole tissue. But this analysis was based on a single replicate, with stages samples every half day.
Here we push the analysis forward, by sampling even more closely related stages, and even replicates. We also ask, whether the temporal signal is spatialized in a developing organ. Here we show that the ordering of consecutive developmental samples appears because many genes show variations that are subtle, yet consistent between consecutive stages which would be classically considered developmental duplicates. Halves of tooth germs, dissected along the bucco-lingual axis, also carry a temporal signal, the buccal side looking developmentally "older" than the lingual side. This can be accounted for by differences of expression levels in linearly increasing/decreasing genes, by changes of tissue proportions over time and between parts of the same organ. Furthermore, different time signals are carried by different tissue compartments in the tooth germ.

## Results and Discussion

We built a dataset of transcriptomes in tooth development, containing samples for eight broadly defined stages covering the whole tooth morphogenesis in duplicates. This is called "large-scale dataset". This dataset is completed with a "small-scale dataset" containing 8 samples at an early stage of development. We also obtained a "bucco/lingual dataset" with tooth halves transcriptomes corresponding to buccal and lingual sides at a stage where there is one cusp patterned on the buccal side, and none on the lingual side (pink/purple stars on the Figure 1). This makes a total of 48 RNA-seq samples to interrogate temporal signal in tooth development. All these samples have been normalized together. In order to match gene expression dynamics with the progression of morphogenesis, we also followed the formation of cusps using *in situ* hybridization in tooth germ (*ISH*).

### Developmental transcriptomes embed a signal of developmental timing.

As a starting point, we used multivariate analysis to extract the main axes of variation the "large-scale" time series for both molars. We combined the lower molar samples, the upper molar samples, or all samples together. In each case, we found that the first axis represents about half of the total variation and orders the samples according to developmental time (Figure 2 and Supp Fig 1). This time ordering is extremely robust and resists subsetting data, either by randomly sampling genes, by taking only the bottom (or top) 20% expressed sets of genes, or by retaining only tissue-specific genes (see later).

Our dataset is sampled at the level of the organ, and not at the level of the whole organism, and is covering a small time range, not the whole embryonic development. So, this is a much smaller time-scale than some previous analyses which recovered a time ordering based on larger time series [2, 3, 4, 5]. In our case, it is not realistic to imagine that this ordering emerges because the expression of many genes is switched on and off. Indeed, on average, only 95 genes are very weakly expressed at least at one stage (DESeq Basemean $< 10$) and expressed (more than twice more) in the consecutive stage. The removal of all these genes (in upper molar germ, 471 genes in total) does

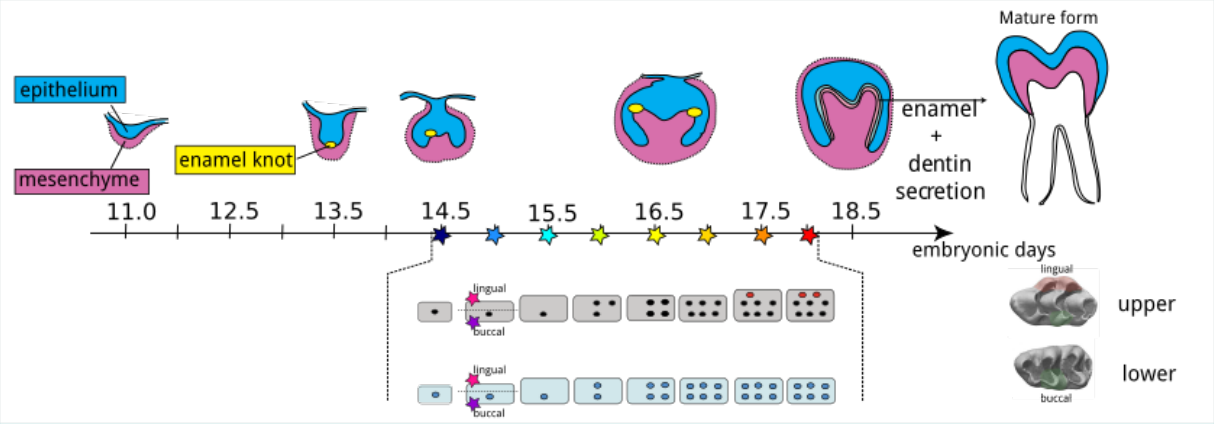



Figure 1 – **Outline of tooth development together with experimental setting for RNA-seq sample collection.** Stages where RNA-seq samples have been collected are indicated by stars.

not impact the strength of PCA1 (45.4% vs. 45.18% for PCA based on Upper molars only), and the ordering of the samples. The self-ordering of developmental samples may rather appear because a sufficient number of genes show variations that are more subtle, yet consistent between consecutive stages. If this is true, virtually any stage difference, even between samples classically considered developmental replicates, could be measurable in the transcriptomes, provided that the other sources of experimental noise are kept minimal. Next we will study the temporal signal among samples considered as duplicates.

## The signal of developmental timing is visible for samples classically considered developmental replicates.

It is obvious on the PCA map that stage duplicates localize correctly, that is, they are more closely related than they are from samples which are not duplicates. Yet, they do not map at the exact same position on the first PCA axis (Figure 2 and Sup Figure 1). Is the offset observed on the first axis of PCA really reflecting a difference of timing, or is it due to unrelated effects? We were as careful as possible in the process of building our dataset: for instance, we retained only female embryos, from a restricted pool of fathers and corresponding to the first litter of the mother, and we minimized dissection time. All this, to minimize all kinds of biological sources of variations. Furthermore, technical sources of variations are visible, separating replicates on the fourth axis of PCA (representing 7.1% of the variation). The libraries corresponding to the two replicates have been prepared with different versions of the kits, and separated by 1 year interval. This batch effect is even the second axis (11.8%) for the PCA made on upper and lower samples only. This is much less than time effect, visible on the first axis (44.8%), or distinction between upper and lower first molars (second axis, 10.3% of variation). One piece of evidence points that at least some of the effect separating replicates on the first PCA axis is developmental timing. Indeed, at a given time point and for each replicate, upper and lower molars have been prepared from the exact same embryo. Hence, it is natural to suppose that a small developmental difference between replicates should result in a coherent offset in upper and lower samples on PC1. Indeed, we observed, at a single exception (stage 17.0), the same relative ordering of the replicates for upper and lower molar (Figure 2). For instance, at stage 16.0, rep 1 appears younger than rep 2 on PC1 for the upper molar, and this is also true for the lower molar. This is true when PCA is drawn on the whole dataset (with both upper and lower samples, as in Figure 2), but also when PCA are drawn separately for upper and lower molars (Sup Figure 1). Another independent observation points in the same direction. Embryos were sampled based primarily on the date of mating, but their weight was also registered, which is useful here, because embryos that are more developmentally advanced tend also to be heavier [13]. This is very accurate within the same litter, but is simply a tendency when comparing different litters. Even though, the correlation between PC1 and weight is slightly better than between PC1 and stage based on the date of mating (Sup Figure 2), suggesting that PC1 differences reflects *bona fide* differences in developmental advancement between replicates.

This observation motivated us to study further the time-related information embedded in developmental replicates. To check whether even finer-scale development stage variations are visible in the transcriptomes, we studied a smaller-scale dataset, which contains 7 upper molar samples, from three independent litters sampled at ED 15.5. Within each litter, we build two "exact replicates" with embryos having the exact same weight, and "normal replicates", with embryos whose weight is slightly more different. Hence, base on weight we can rank the advancement of development within each litter. However, as said above, weight is very accurate within the same litter, but is simply a tendency when comparing different litters.

We thought to use also Sonic hedgehog (*Shh*), as a marker to estimate the age of the embryo, because it is a

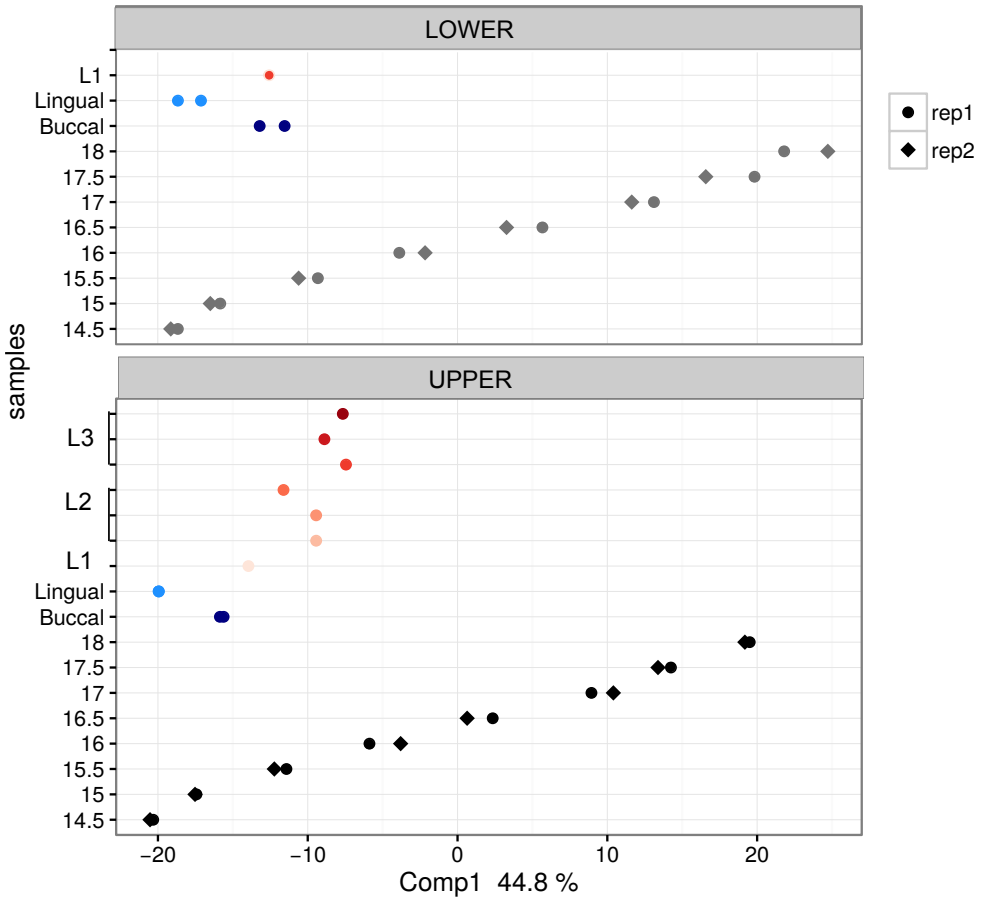



Figure 2 – Ordination of the RNA-seq samples on the first component of a PCA analysis (Comp1) based on the dataset of 32 whole tooth samples covering 8 broad stages during the whole period of cusp morphogenesis. 16 supplementary samples (colors) have been mapped to this axis. "UPPER": 16 samples of lower tooth ranging from ED 14.5 to ED 18 order on this PCA, whose first axis represents 44.8% of the total variation. An independent sample (red, L1= litter 1) of a tooth aged about ED 15.5 maps directly onto this axis. Buccal and lingual tooth halves (blue) also map around their corresponding stage, ED 15.0. "LOWER": 16 samples of upper tooth ranging from ED 14.5 to ED 18 order on this PCA. Seven independent samples (shades of red: small-scale) of a upper tooth aged about ED 15.5 map correctly onto this axis, and show a subtle gradient of development stages. This small-scale dataset was built using 3 independent litters, labeled here L1, L2 and L3. Buccal and lingual tooth halves also map around around their corresponding stage, ED 15.0.

known marker of cusp formation whose expression territory increases progressively in the tooth germ. We confirmed by *in situ* hybridization that this is seen during the whole period of morphogenesis, and we checked that it is also true during a short period of time. As expected given these in situ data, Shh expression level in the transcriptomes shows a linear increase with time (this is true in upper and lower molar), making it a good indicator of developmental time during the whole period of cusp patterning (Figure 3).

This suggested *Shh* could also be a good indicator in the small time window of our "small scale" dataset. In consequence, we used *Shh* expression level in the samples of the small-scale dataset to estimate their age based on a linear regression (pink-red dots Figure 3). The estimates all fall within a ED 15.0-ED 16.0 window, as expected

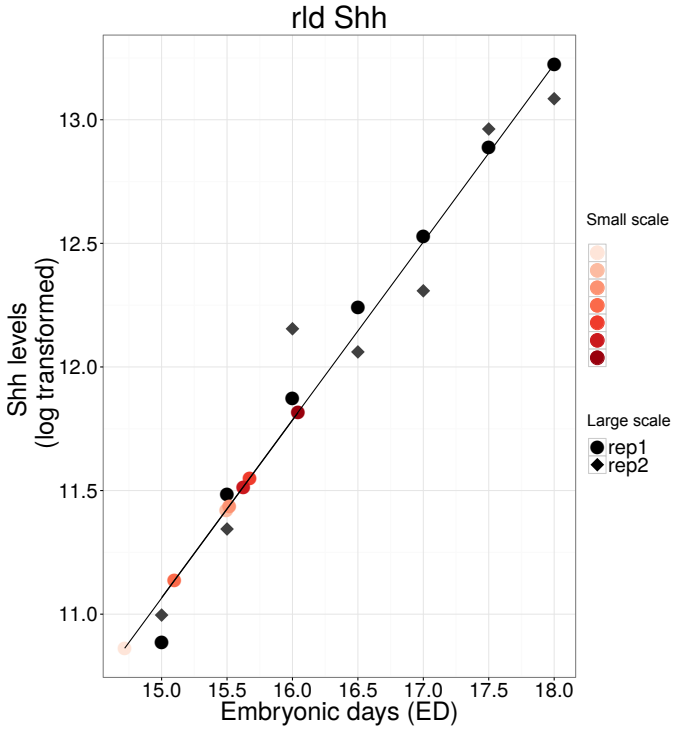



Figure 3 – Shh expression recapitulates the correct temporal ordering of samples. Shh expression level increases regularly during development, which is visible in RNAseq data (rld normalization) from samples ranging from ED 15.0 to ED 18.0. Shades of red: The small-scale samples were dated based on the linear relationship between the days of development and shh normalized expression levels (excluding stage 14.5 for which expression level is out of the linear part of Shh temporal profile)

given the sampled development stages. Of note, within litters, embryos were ranked by Shh as expected given their weight.

The small developmental differences are visible at the whole transcriptome level, since *Shh* ranking gives the higher number of DE genes with a correlated linear profile (highest number out of 500 random permutations of *Shh* ranking, P-value < 10-3). A multivariate analysis of this small scale dataset does not produce a time ordering. This is consistent with the fact that, in this dataset, the dissection time and sex were not as well controlled for as in the large scale dataset. However, when mapped on the PCA of the large timescale dataset, the small timescale samples ordered as expected (Figure 2), and moreover their coordinate on the first component of the PCA dataset correlated almost perfectly with the Shh expression level. This is true whether or not *Shh* is removed from the dataset used to build the PCA (R2 = 92%, P-value = 5e-4). These results suggest that consistent developmental time information is present in this small timescale set dataset, although embedded in noise coming from other signals (like dissection time, sex, or genetic differences between litters, as we are working with outbred mice).

## Developmental time signal is congruent between small and large developmental scales.

We then looked for genes that vary significantly with *Shh* expression level (increasing or decreasing, next called "linear genes"; DESeq2 adjusted P-value < 0.1) in the small timescale data set and studied their profile in the large-scale expression data set. We found 457 of these "linear genes" (269 decreasing genes, 188 increasing genes). These linear genes (in the small scale dataset) very often significantly vary in the large time scale dataset (81%, P-value for Chi-squared test <10-16) . Furthermore, they are largely varying in the same way (twice more often than expected, see Figure 4 bottom and associated P-value for Wilcoxon test = 10-16). Similar results were found in the reciprocal analysis, where linear genes were defined in the large-scale dataset, and the correlation of their expression in the small-scale dataset with PCA1, weight and Shh was computed (Figure 4 top). Shh and PCA1 coordinates are very coherent and equally good indicators of time in our dataset. Of note, and as expected, embryo weight is a poorer indicator of time especially when comparing different litters. We conclude that samples taken in a very small time window contain time information which may be hidden by other sources of variation, yet is biologically

relevant, and is carried at least in part by "linear genes" showing consistent increase or decrease in the small and large timescale. Finally, this analysis shows that RNA-seq "small timescale" data capture very fine increase or decrease in gene expression levels, which biological meaning might be questionable on a gene-by-gene basis but which is biologically relevant when sets of genes are considered.

### Cellular proportions carry the temporal signal and sub-organs also carry a development time signal.

What is carrying the temporal signal in a complex organ? The period we are studying here encompasses major morphogenesis events in the molar crown, and, among them, the shaping of the crown with the sequential addition of cusps. We recorded the number of patterned cusps in each samples selected for RNA sequencing (using *in situ* hybridization with *Fgf4*). As expected, the number of cusps increases with developmental stage (Figure 5 bottom). Of course, cusp number is correlated with PCA1 in the transcriptomes. However, cusp number does not increase quick enough to explain the differences between replicates. Hence, there are only a few cases where the number of cusps differs between replicates (eg 15.5 and 17 in the lower samples Figure 5 bottom). Even though, in these cases the coordinates on the PCA does not correlated with the number of cusps (e.g. ED 17 in the lower). Hence, the cusp number is well correlated with PCA1 on the large scale, but it does not explain the difference between duplicates (Figure 5 bottom). This confirms our previous results (Pantalacci *et al.*, submitted [12]), with more data. Why is transcriptome more subtle than simply cusp number to register the advancement of development? When patterned, each cusp expands, that is, induces the differentiation of neighboring cells, which can be simplified as an "naive epithelial tissue" being converted into a "cusp tissue". Hence, mechanically, the proportion of this "cusp tissue" over the whole tooth germ is increasing from the patterning of each cusp and until the end of the process. Because replicates were not sampled at the exact same developmental stage, it is possible that they show the same number of cusps in an *in situ* hybridization, but yet, the cusps were patterned earlier in one replicate than the other. Hence, the relative quantity of "cusp tissue" is expected to be slightly different from one replicate to the next. Hence cusp proportion is well correlated with transcriptomic age at a large scale, but not at smaller scale. This may be due to several technical reasons.

To prove that cusp proportion was a strong component in dating complex organs in development, we cut a tooth germ in two halves, at a stage where it is strongly regionalized, E15.0. Indeed, at this stage, a cusp is already patterned on the "buccal" side, but not on the "lingual" side. If the temporal signal is carried by tissue proportion, we should be able to estimate the relative "age" of these samples. In our case, the buccal part of the tooth germ with more cusp tissue would look older than the other part. We used again a deconvolution method to estimate the proportion of cusp tissue epithelium. As expected, we found that the transcriptomes of the buccal sides are clearly enriched in cusp epithelium (Sup Fig 3). In accordance with the fact that they contain more cusp tissue, the buccal sides are clearly ahead of the lingual side along the temporal axis of the PCA (Figure 2).

This results holds true under the examination with another method. We selected 4412 “timer genes” which are genes for which the expression increases or decreases linearly during development more than twice (DESeq2, FDR <0.1, large-scale dataset of 8 time points ). For each of these timer genes we fitted a linear regression to the expression levels during the 8 stages of development. Separate analyses were performed for upper and lower tooth germs. Using this linear models we predicted the "stage" of halve tooth germs, reasoning that the buccal side should appear older than the lingual side on average on this set of linear genes. Although they were computed independently, the timing were well correlated between upper and lower molar (Adjusted R-squared: 0.46, p-value <2.21-e-16), meaning that, although noisy, these estimated ages are meaningful biologically. Indeed, as expected, the median age of buccal transcriptomes is 15.1 days of development and the median age of lingual transcriptomes is 14.6 days, a difference of about 0.5 days of development that is significant in both upper and lower molars (Wilcoxon test, p-values < 2.2e-16 ). So, there is a temporal signal which permits to date each tooth germ halves (both project around E15.0) but with subtle differences. We hypothesized that these differences could be related to the presence of a more advanced cusp structure at the buccal side. This was further confirmed by the study of tissue-specific timer genes (extracted from the set of 4412 timer genes). The set of "cups tissue" specific genes clearly dates the buccal samples older than the lingual samples, but a set of "mesenchyme specific" genes does not (Figure 6).

## Conclusion

Many previous transcriptome analyses had noted in passing the presence of a time-related signal in developmental transcriptomes. But this was often done at a large developmental scale. Here we show a temporal signal which is observable even between closely related samples, that would normally be considered replicates. We show that this is a temporal signal, because the variation between these closely related samples mirrors the variation observed at a larger developmental stage, although with less amplitude. In other words, there is no true developmental replicates in transcriptomics. Already some studies have been using multivariate analyses to obtain a measurement of the developmental advancement of their samples [4, 7]. Here we show a very fine ordering of samples, but also of subsamples, on a given timeseries. This is demonstrating that staging is possible at a very precise scale.

The reasons beyond the self ordering of developmental transcriptomes have been largely overlooked. In a com-

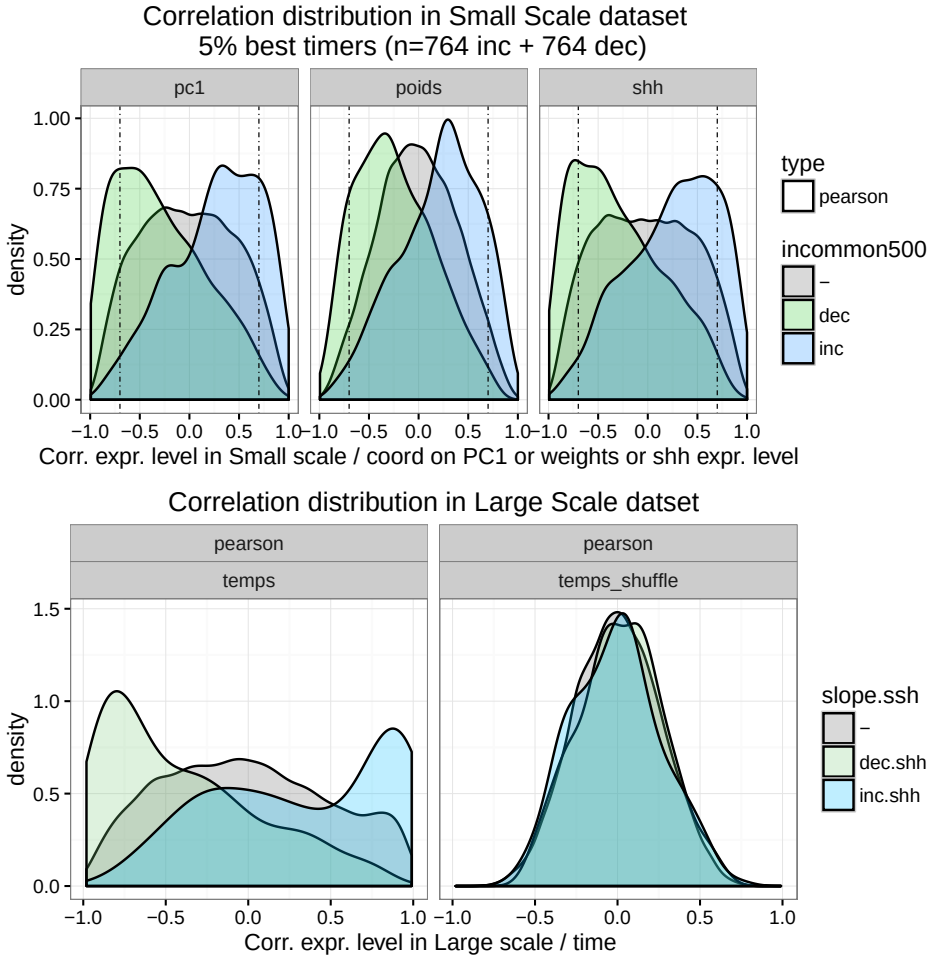



Figure 4 – Congruent temporal signal between large and small timescale. A For each gene, a Pearson correlation between expression level in small scale dataset and (i) the coordinates of the sample on PC1, (ii) the weight of the embryo from which the sample come from , or (iii) the expression level of Shh in small scale dataset was computed. Genes were grouped in three categories according to their expression profiles in the large scale dataset: significantly decreasing with time (green), significantly increasing with time (blue) or no significant increase or decrease (grey). Genes that decrease in the large scale dataset are negatively correlated with PC1 (or weight or shh expression), meaning that there are also decreasing in the small scale dataset. The increasing/decreasing behavior of a gene in the small scale dataset is congruent with the increasing/decreasing behavior in the large scale dataset. Note that we see a better congruence of small and large scale datasets with coordinates on PC1 and Shh expression than with the weight of the embryo. B Reciprocal analysis. For each gene, a Pearson correlation between expression level in large scale dataset and the stage of the sample (time in days) was computed. Genes were, in parallel, grouped in three categories based on their temporal profile in the small scale dataset: (blue) genes significantly increasing with Shh expression, (green) genes significantly decreasing with Shh expression, and genes with no significant correlation with Shh (grey). The 3 sets of genes are clearly separated, which is not the case when correlations with randomly shuffled stages are computed (right).

parative study in flies, it was noted that distinct developmental stages map closely on the developmental map. The presence of early and late mappings was consistent with the fact that flies undergo two rounds of large-scale proliferation of undifferentiated tissue followed by differentiation: once during embryonic stages and again during larval and pupal stages . Here we show that the time-related signal in transcriptomes is linked to the proportion of cusp tissue, which is a differentiated epithelium tissue. Here we link cell proportions and development advancement at a large scale, but to demonstrate this at a smaller scale, more subtle proportion differences need to be recovered. It would be necessary to analyze other cell populations than cusp tissues. Dissection effects also

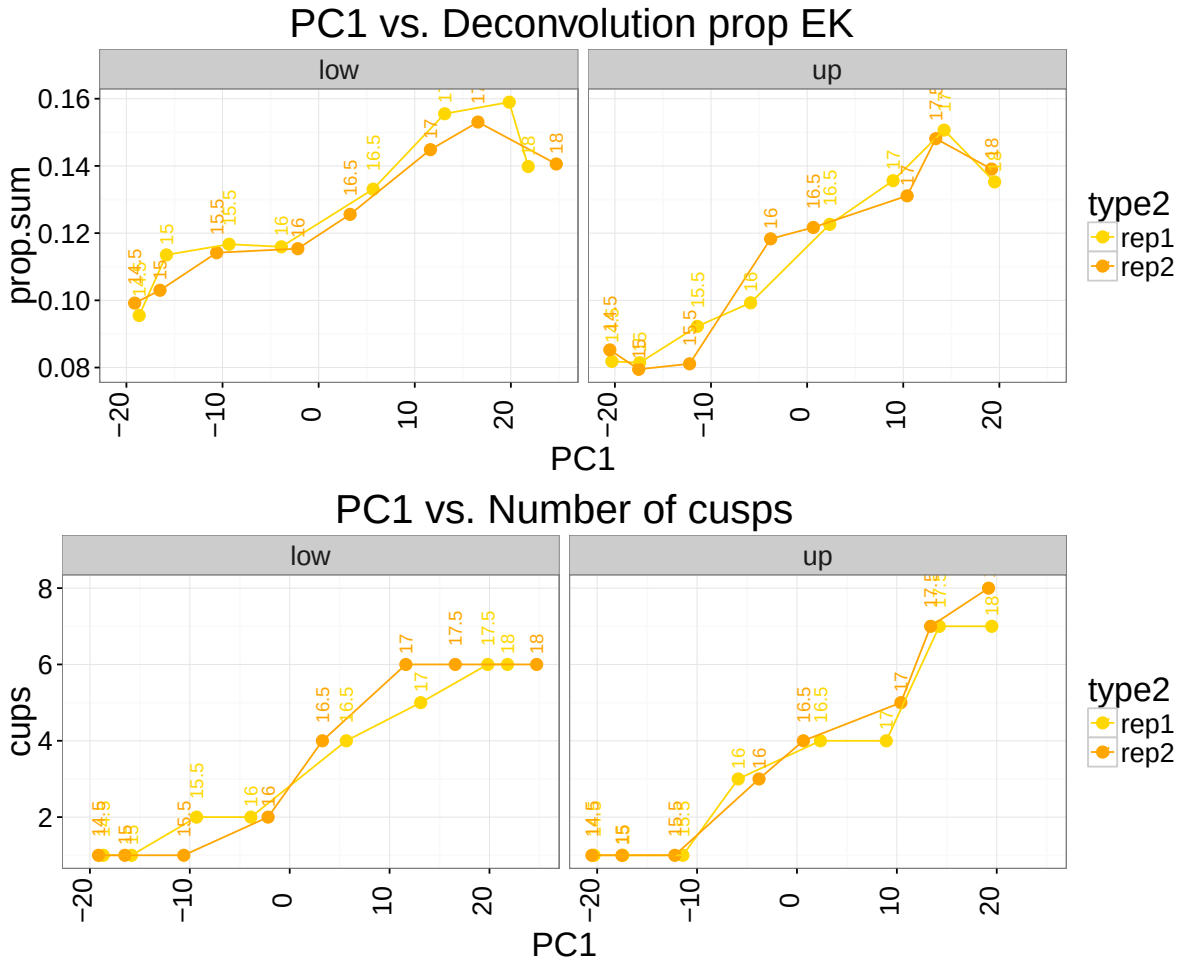



Figure 5 – **Top**: Proportion of cusp tissue estimated by deconvolutions in the large-scale dataset. **Bottom**: Number of cusps patterned assessed by in situ hybridization with Shh.

tend to interfere with proper cell type quantifications at this scale of development. Another consequence of our observations in that transcriptomes ordering is a way to monitor cell differentiation.

A surprising result of our analysis is that tooth halves can be dated, and that they harbor subtle stage differences. Again, this may be explained by the different relative proportion of cusp tissues in buccal and lingual sides. It should also be noted that the temporal signal is true in all tissue compartments but with subtle differences. Epithelium, Mesenchyme and Cusp tissue all date the organ correctly around ED 15.0, but the offset between buccal and lingual sides are different in the different compartment.

Finally, transcriptome developmental timing is a way to date whole organs or subparts of organs, and indirectly records changes in the proportion of differentiated cells.

# Methods (preliminary draft)

## Mice breeding and embryo sampling

CD1 (ICR) adult mice were purchased from Charles River (France). Females were mated overnight (night at 20:00 pm) and the noon after morning detection of a vaginal plug was indicated as embryonic day (ED) 0.5. Other breeding pairs were kept in a light-dark reversed cycle (night at 12:00 pm), so that the next day at 16:00 pm was considered as ED1.0. Pregnant females were killed by cervical dislocation and embryos were harvested on cooled Hank's or DMEM advanced medium, weighted as described in [13] and immediately decapitated. This study was performed in strict accordance with the European guidelines 2010/63/UE and was approved by the Animal Experimentation Ethics Committee CECCAPP.

Epithelium dissociations and in situ hybridizations Complete or hemi mandibles and maxillae were dissected in Hank's medium and treated with Dispase (Roche) 100% at 37°C for 1h30 to 2h20 depending on embry-

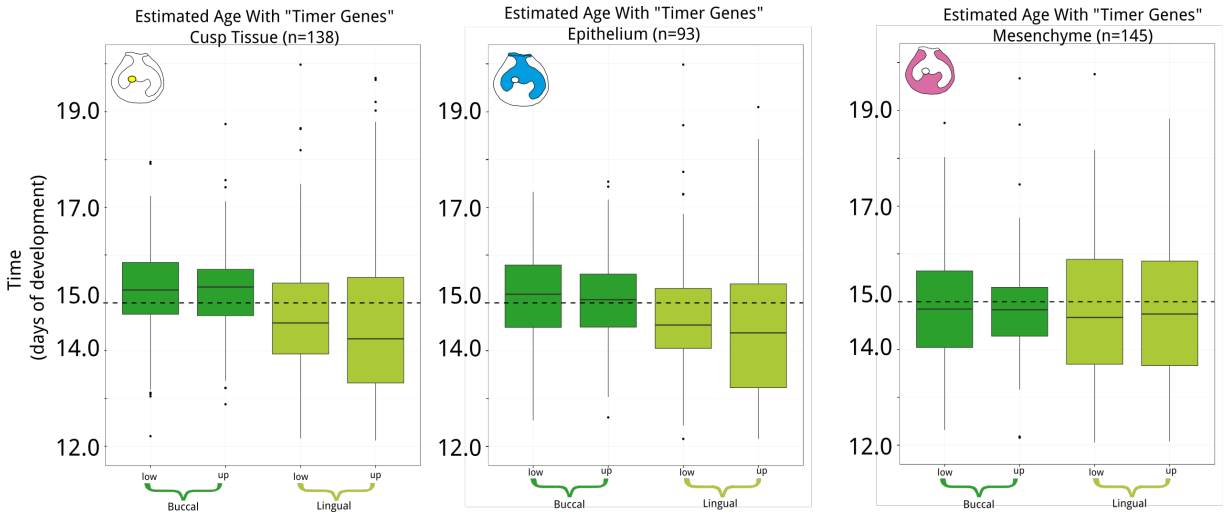



Figure 6 – Estimation of the developmental age of buccal and lingual samples based on linear markers. Among genes with an increasing or decreasing expression profile during development (detected in the large-scale dataset of 8 time points) we isolated 3 sets of tissue-specific genes (cusp tissue, epithelium and mesenchyme). This was based on published pure-tissue microarray data [2]. The number of genes in each category is indicated. Boxplots represent the estimated age for these sets of genes, in buccal and lingual samples.

onic stage. Epithelium was carefully removed and fixed overnight in PFA 4% at 4°C. DIG RNA antisense Fgf4 probe were prepared from plasmids described elsewhere [14]. In situ hybridization was done according to a standard protocol. Photographs were taken on a Leica M205FA stereomicroscope with a Leica DFC450 digital camera (Wetzlar, Germany) or on a Zeiss LUMAR stereomicroscope with a CCD CoolSNAP camera (PLATIM, Lyon).

## RNA-seq sample preparation

A total of 32 samples were prepared for the RNA-seq analysis, that correspond to 3 datasets, which we called "small-scale" (8 samples around ED 15.5), "large-scale" (16 samples ranging from ED14.5 to ED 18.0) and "Bucco-lingal" (8 samples containing tooth halves around ED 15.0) The "small-scale" dataset contains 7 samples of upper molar, and 1 sample of lower molar. Each sample comes from a pool of 3 individuals of the same litter which have essentially the same weight. We used 3 litters, named L1, L2 and L3 in the text. L1 produced the lower sample and one upper sample. L2 and L3 each produced 3 samples of upper molars. The dataset was built such that 2 samples were really close stages as assessed by the weight of the embryo, while another sample was a bit farther apart. The "large-scale" dataset contains 16 samples representing 8 stages (ED14.5, 15.0, 15.5, 16.0, 16.5, 17.0, 17.5, 18.0), coming from 8 individuals. This builds a replicate for a previously published large-scale dataset in the lab (Pantalacci *et al.*, submitted [12]). Each sample contained 2 tooth germs, the left and right first molars (M1) of the same male individual, and for a given stage, the upper and lower samples came from the same individual. The "bucco-lingual" dataset contains 8 samples representing the 4 first molars of the same individual, sampled around ED 15.0. Each molar was cut in half, and each half was prepared independently for sequencing. The heads of harvested embryo were kept for a minimal amount of time in cooled PBS (small scale) or advanced DMEM medium (large scale). The M1 lower and upper germs were dissected under a stereo-microscope and stored in 200ul of RNA later (SIGMA). Another embryo of the same litter and same weight was processed as indicated above for *Fgf4 in situ* hybridization. Total RNA was prepared using the RNeasy micro kit from QIAGEN following lysis with a Precellys homogenizer. RNA integrity was controlled on a Bioanalyzer (Agilent Technologies, a RIN of 10 was reached for all samples used in this study). PolyA+ libraries of the large-scale dataset were prepared with the Truseq V2 kit (Illumina), starting with 150ng total mRNA and reducing the amplification step to only 12 cycles and sequenced on an Illumina Hi-seq2000sequencer at the GENOSCOPE (Evry, France). Samples for the Small-scale dataset were sequenced using GAIIX at the IGBMC platform (Illkirch, France).

## Expression levels estimation and differential expression analysis.

Sequences were obtained for each samples through Illumina sequencing (Large-scale samples: 100 bp paired-end sequences; Small-scale and bucco-lingual samples: 50 bp single end sequences). Expression levels were obtained by pseudo alignments using Kallisto [15] (version 0.43.0) and the mouse reference genome (GRCm38, release ensembl 85).

Statistical analysis was performed using R (version

3.1.2, 2014-10-31) [16], with ggplot2 for graphics (ggplot2-1.0.0; [17]). Two main analysis used in the text include differential gene expression analysis (DE) and multivariate analysis. Differential expression analysis was performed using DEseq package (DESeq2, version 1.6.3 [18]), with settings indicated in the main text. Multivariate analyses (mainly principal component analyses) were performed using ade4 package (ade4, version 1.6.2 [19]).

### Estimating tissue proportions from RNAseq data: deconvolutions

Performing gene expression deconvolution, is estimating cell type proportions (and/or cell-specific gene expression signatures) from global expression data in heterogeneous samples. Here, we used the R package CellMix (CellMix, version 1.6.2, [20]), to estimate the proportions of the three tissue compartments in our tooth germ transcriptomes. CellMix implements in particular the method Digital Sorting Algorithm (DSA) proposed by Zhong *et al.* [21] which performs complete gene expression deconvolution using a set of marker genes only. We needed therefore to define 3 sets of markers, as specific as possible for each tissue compartment. We used markers obtained from microarray data [2] 18 EK (defined such as expression level in ek $>$ 1.2x expression level in mesenchyme and level in ek $>$ 1.2x level in epithelium), 9 for epithelium (such as epithelium $>$ 1.5x mesenchyme and epithelium $>$ 1.2x ek) and 27 mesenchyme (such as mesenchyme $>$ 1.5x epithelium and mesenchyme $>$ 1.2x ek).

## References


[1] M. Ringnér, "What is principal component analysis?," *Nature Biotechnology*, vol. 26, pp. 303–304, mar 2008.

[2] D. J. O'Connell, J. W. Ho, T. Mammoto, A. Turbe-Doan, J. T. O'Connell, P. S. Haseley, S. Koo, N. Kamiya, D. E. Ingber, P. J. Park, *et al.*, "A wnt-bmp feedback circuit controls intertissue signaling dynamics in tooth organogenesis," *Science Signaling*, vol. 5, pp. ra4–ra4, jan 2012.

[3] M. Levin, T. Hashimshony, F. Wagner, and I. Yanai, "Developmental milestones punctuate gene expression in the caenorhabditis embryo," *Developmental Cell*, vol. 22, pp. 1101–1108, may 2012.

[4] L. Anavy, M. Levin, S. Khair, N. Nakanishi, S. L. Fernandez-Valverde, B. M. Degnan, and I. Yanai, "BLIND ordering of large-scale transcriptomic developmental timecourses," *Development*, vol. 141, pp. 1161–1166, feb 2014.

[5] P. Tschopp, E. Sherratt, T. J. Sanger, A. C. Groner, A. C. Aspiras, J. K. Hu, O. Pourquié, J. Gros, and C. J. Tabin, "A relative shift in cloacal location repositions external genitalia in amniote evolution," *Nature*, vol. 516, pp. 391–394, nov 2014.

[6] I. Yanai, L. Peshkin, P. Jorgensen, and M. W. Kirschner, "Mapping gene expression in two xenopus species: Evolutionary constraints and developmental flexibility," *Developmental Cell*, vol. 20, pp. 483–496, apr 2011.

[7] M. Levin, L. Anavy, A. G. Cole, E. Winter, N. Mostov, S. Khair, N. Senderovich, E. Kovalev, D. H. Silver, M. Feder, S. L. Fernandez-Valverde, N. Nakanishi, D. Simmons, O. Simakov, T. Larsson, S.-Y. Liu, A. Jerafi-Vider, K. Yaniv, J. F. Ryan, M. Q. Martindale, J. C. Rink, D. Arendt, S. M. Degnan, B. M. Degnan, T. Hashimshony, and I. Yanai, "The mid-developmental transition and the evolution of animal body plans," *Nature*, vol. 531, pp. 637–641, feb 2016.

[8] L. B. Snoek, M. G. Sterken, R. J. M. Volkers, M. Klatter, K. J. Bosman, R. P. J. Bevers, J. A. G. Riksen, G. Smant, A. R. Cossins, and J. E. Kammenga, "A rapid and massive gene expression shift marking adolescent transition in c. elegans," *Sci. Rep.*, vol. 4, jan 2014.

[9] J. Jernvall and I. Thesleff, "Reiterative signaling and patterning during mammalian tooth morphogenesis," *Mechanisms of Development*, vol. 92, pp. 19–29, mar 2000.

[10] I. Thesleff, S. Keranen, and J. Jernvall, "Enamel knots as signaling centers linking tooth morphogenesis and odontoblast differentiation," *Advances in Dental Research*, vol. 15, pp. 14–18, aug 2001.

[11] M. A. Landin, S. Nygård, M. G. Shabestari, E. Babaie, J. E. Reseland, and H. Osmundsen, "Mapping the global mRNA transcriptome during development of the murine first molar," *Frontiers in Genetics*, vol. 6, feb 2015.

[12] S. Pantalacci, L. Guéguen, C. Petit, A. Lambert, R. Peterkova, and M. Sémon, "Transcriptomic signatures shaped by cell proportions shed light on comparative developmental biology," *Genome Biology*, Dec. 2016.

[13] M. Peterka, H. Lesot, and R. Peterková, "Body weight in mouse embryos specifies staging of tooth development," *Connective Tissue Research*, vol. 43, pp. 186–190, jan 2002.

[14] J. Jernvall, T. Aberg, P. Kettunen, S. Keranen, and I. Thesleff, "The life history of an embryonic signaling center: Bmp4 induces p21 and is associated with apoptosis in the mouse tooth enamel knot," *Development*, vol. 125, no. 2, pp. 161–169, 1998.

[15] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," *Nature Biotechnology*, vol. 34, pp. 525–527, apr 2016.

[16] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[17] H. Wickham, *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media, 2009.

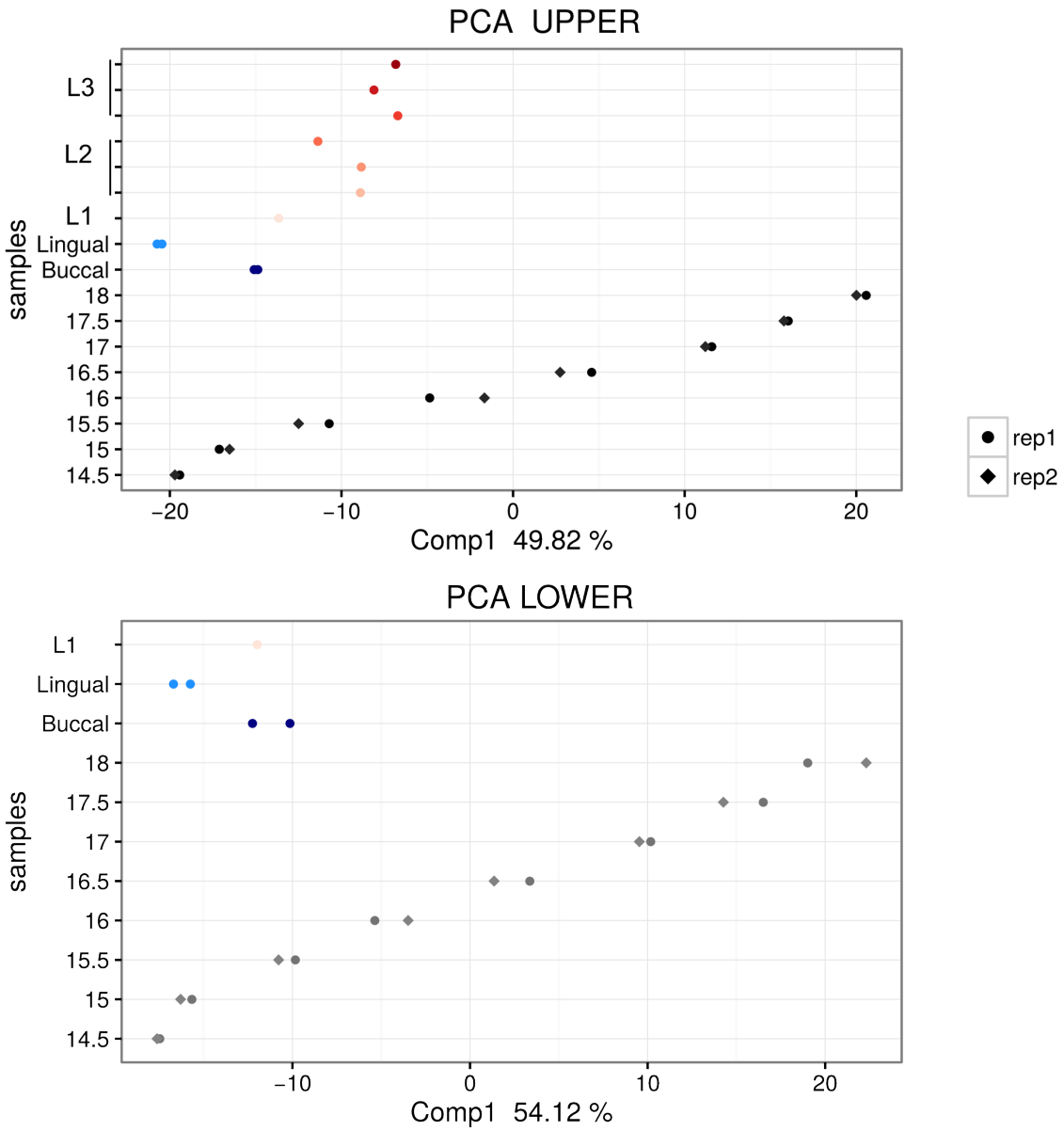
[18] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol*, vol. 15, dec 2014.

[19] S. Dray and A.-B. Dufour, "The ade4 package: implementing the duality diagram for ecologists," *Journal of statistical software*, vol. 22, no. 1, pp. 1–20, 2007.

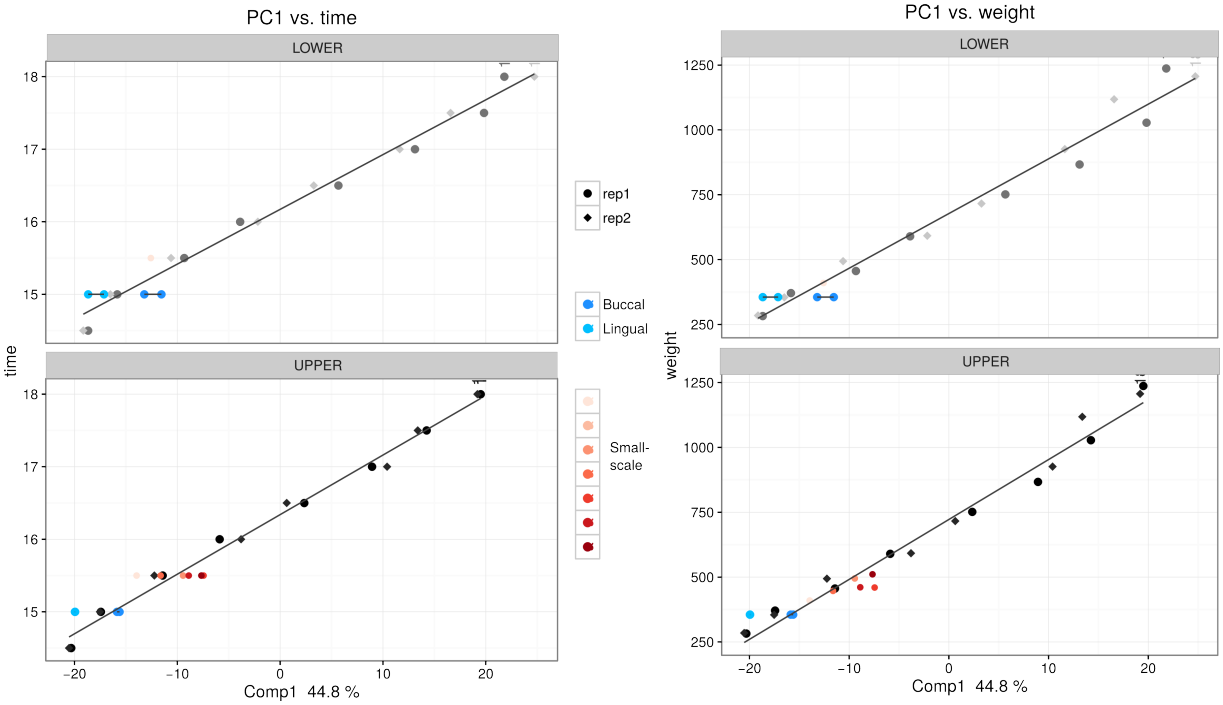
[20] R. Gaujoux and C. Seoighe, "CellMix: a comprehensive toolbox for gene expression deconvolution," *Bioinformatics*, vol. 29, pp. 2211–2212, jul 2013.

[21] Y. Zhong, Y.-W. Wan, K. Pang, L. M. Chow, and Z. Liu, "Digital sorting of complex tissues for cell type-specific gene expression profiles," *BMC Bioinformatics*, vol. 14, no. 1, p. 89, 2013.

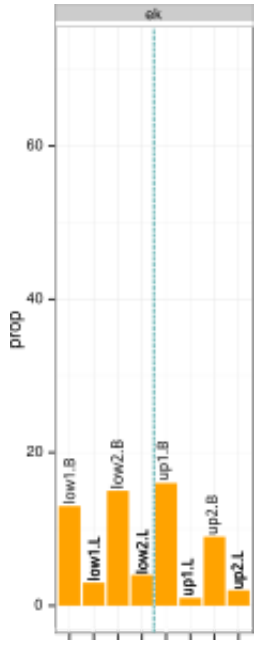
## Supplementary figures




**Supplementary figure 1**




**Supplementary figure 2**




**Supplementary figure 3**

# 9 Comparative transcriptomics of rodent molars

## Introduction

How to link differences of the developmental program to changes in morphology? From our comparative transcriptomic study presented in the previous chapters, we identified differences in developmental program between the upper and the lower first molar of mouse. Upper and lower molars are intrinsically different. First, they are deployed in different "environment" : the upper and the lower jaw (see homeotic code *p. 41* ). For example, *Dlx5/6* are necessary to specify the identity of lower jaw, loss of expression of these genes lead to its development into an upper jaw [Depew, 2002; Beverdam *et al.* , 2002]. Moreover, teeth are serial organs, meaning that they are the same type of organs deployed in different places, for the teeth it is in different "jaw environment". Tooth identity is defined by a combination of homeotic genes expressed along the jaw. For example, *Barx1* is essential to direct molar identity, hence expression of *Barx1* in the presumptive incisor territory, transform the tooth germ into a molar [Tucker, 1998] (see homeotic code *p. 41* ). Second, upper and lower molars have a distinct evolutionary history. The most striking differences in morphologies arise 200 MYA, with the evolution of the tribosphenic molar [Butler, 1990] in early mammals. This clear separation in term of evolution and position on the jaw suggests that the upper and lower molars evolved different developmental program. Obviously these programs should share a fair amount of characteristics because they form organs that are serial homologues [Wagner *et al.* , 2007]. Back to our question of how the mouse first upper molar developed two supplementary cusps. The upper-lower differences we identified at the transcriptomic levels in mouse are thus two fold : they are the intrinsic differences between upper and lower molar developmental programs and they are the differences that allow the mouse upper molar to form these two supplementary cusps. In order to disentangle these two aspects of the upper-lower variations, we used comparative transcriptomic of upper and lower molar in hamster *Mesocricetus auratus*. Contrary to the mouse, the hamster (figure 9.2) has the same number of cusp on its first upper and lower molar. We thus used the hamster as a control

species to identify the upper-lower differences in a case where the same number of cusps are formed. The strategy of this study is to compare molar germs transcriptomes during crown morphogenesis of both upper and lower molar in mouse and hamster.

**How to compare species expression levels throughout development?**
There are still technical challenges when comparing the evolution of expression patterns in different species [Roux *et al.* , 2015]. The first one is to obtain comparable expression levels between species.The second, more related to developmental studies, is to determine homologous developmental stages between compared species.

**Comparable expression level**

Comparative studies often deal with non-model organisms that don't have a sequenced genome. However, to obtain expression data from RNA-seq, the essential step is to align the RNA-seq reads to transcripts of reference and to count how many transcripts aligned to the reference in order to infer the level of expression of this transcript. Previous studies [Lin *et al.* , 2014; Gilad & Mizrahi-Man, 2015] have shown that batch effect could bias the outcome of the analysis. The quality of the annotation of the genome, to which the reads are align, is also going to bias the level of expression estimated and it would ultimately bias the comparison of these levels between species.

Carine Rey, a PhD student in the lab, developed *amalgam* (https://github.com/CarineRey/amalgam), a method to introduce RNA-seq data in existing multi-species alignments and reconstruct reliable phylogenies. I would briefly introduce how we used this pipeline to obtain comparable expression levels between mouse and hamster. Lets define an hypothetic $gene\ i$. The first step of the method is performed by *apytram* ( https://github.com/CarineRey/apytram) which is an assisted *de novo* assembler for transcripts. It will reconstruct a reference transcript for gene $i$ in each compared species from its RNA-seq library, with the help of a scaffold of $gene\ i$ from a model organism. The principle of *apytram* is schematically represented in figure 9.1. Once $gene\ i$ transcript is reconstructed in every species, only the orthologous parts of the reconstructed transcripts are kept, thus every transcript of $gene\ i$ will have the same length in every species. Then the position of the transcript in a phylogeny is checked. If all transcripts for $gene\ i$ (one for each species) placed correctly in the phylogeny, the transcripts are retained and used to estimate expression level of $gene\ i$ in every species.

**Homologous developmental stages**

A second challenge, that is more related to working with timeseries transcriptomes, is to order transcriptomes according to their developmental stages and also to be able to homologize developmental timepoints between species. Previous work on how to scale developmental staged between different species have already been done. For example, Goltsev & Papatsenko [2009] developed a method based on Pearson distance matrices, in combination with other

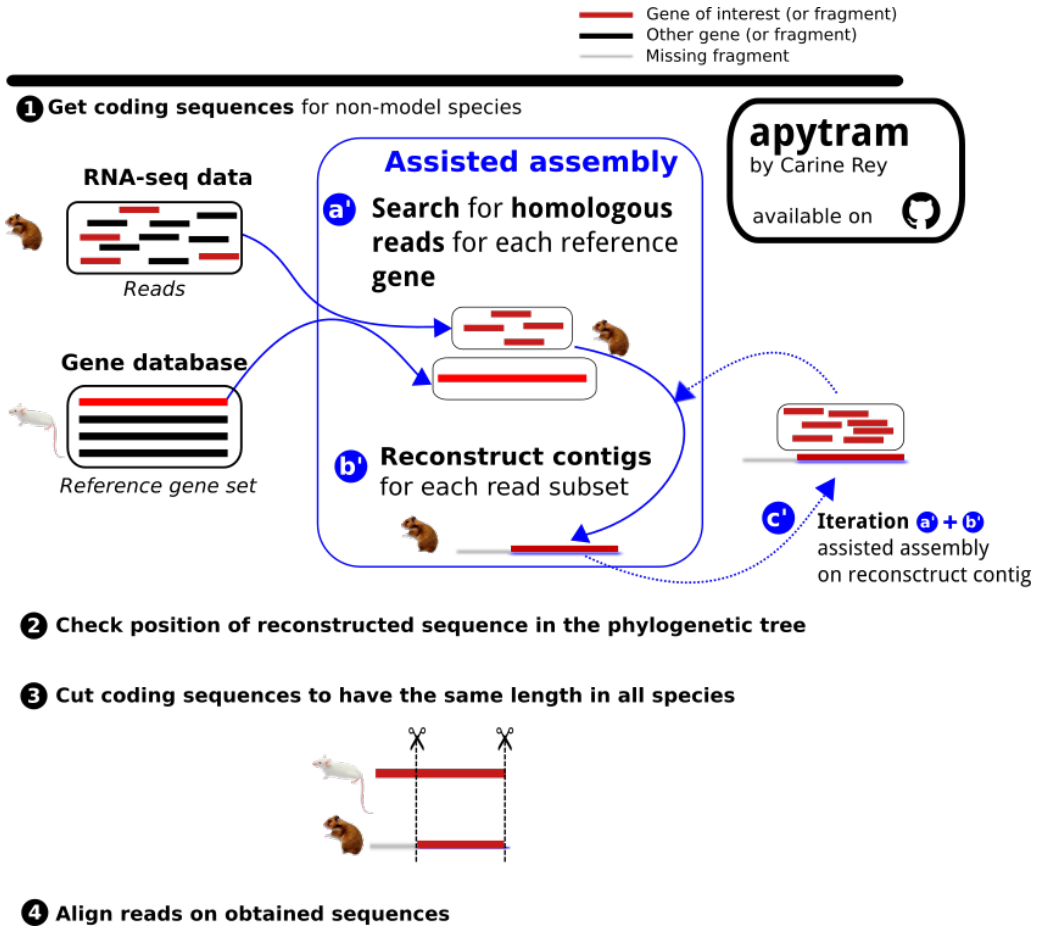



FIGURE 9.1 – **How to estimate comparable expression levels between species?**
Method developed by Carine Rey to estimate comparable expression levels between species.

noise-suppression techniques, to "synchronize" expression profile between different species, this method allow to align expression profiles between species. Papatsenko *et al.* [2011] used this method to compare expression profiles between *Drosophila* and *Anopheles* at different developmental stages, to identify incongruities in their developmental program that could be accounted for morphological differences. Previous works on embryos timeseries [Anavy *et al.* , 2014; Levin *et al.* , 2016] used the temporal signature present in developing embryos to order timeseries transcriptomic sample. From chapter 8, we have shown that temporal signal is a major signature in transcriptomes of developing organs. Multivariate analysis of upper and lower molars transcriptomes of mouse and hamster during crown morphogenesis revealed that the second axis of variation (PC2) is a shared developmental dynamics. Thus in order to obtain comparable timepoints between mouse and hamster transcriptomes, we took advantage of the common temporal signature in both species. Using coordinates on PC2 allowed us to define new developmental timepoints homologous in both species. This allows for the comparison of expression profiles and it also enable to align expression profiles between species in order to

identify which genes have a conserved expression pattern and on the contrary which genes have different expression pattern resulting either from neutral evolution or adaptive selection.

### Evolution of molar shapes and developmental program in rodents

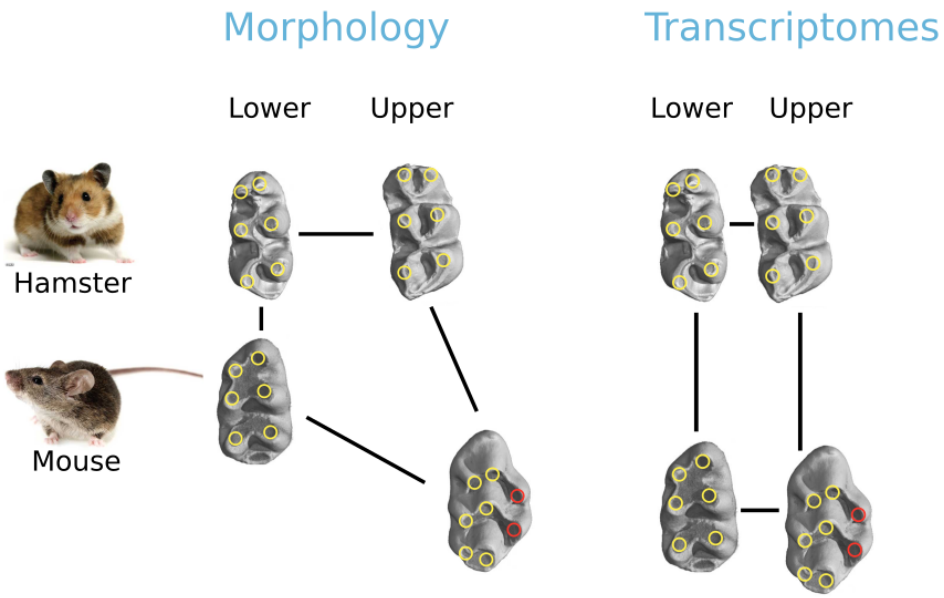



FIGURE 9.2 – **Schematic representation of variations between mouse and hamster molars at the morphological and transcriptomic level.**
**Morphology** Mouse and hamster lower molars are morphologically more alike. Mouse upper molar is morphologically different from the other 3 teeth as it develop two supplementary cusps on its lingual part (cusps highlighted in red). The black bars are proportional to the morphological distance between each teeth. We asked if the morphological differences correlate with the transcriptomic differences? **Transcriptomes** Schematic representation of the transcriptomic differences between upper and lower molars of mouse and hamster. Our comparative analysis revealed that the main pattern of variation between transcriptomes come from the species identity, that account for more than 50% of total variation according to multivariate analysis. Differences in development is almost as high for upper molars as for lower molars despite less morphological changes in lower molars crown. These discrepancies between morphological evolution and the evolution of developmental programs reveal extensive Developmental System Drift (DSD) in molar development.

#### Evolution of gene expression between species

From this comparative analysis, we found that the species identity represent 53.7% of total variation. We are quite convinced that this is not a batch effect [Gilad & Mizrahi-Man, 2015; Roux *et al.* , 2015] because we were very careful in the estimation of expression level in each species (as explained above). Moreover the transcriptomic differences between mouse and hamster are as large in upper molars as in lower molars despite the morphological similarities between mouse and hamster lower molars. We confirmed these results for genes known to be functionally

relevant for tooth development. Thus, the amount of differences between mouse and hamster transcriptomes does not correlate with morphological similarities/dissimilarities, suggesting a rapid evolution of expression levels in each species indicating a Developmental System Drift (DSD) in molar development.

**Participation to this work**

I worked on this project, in parallel with the bucco/lingual analysis, during the last year of my Ph.D. This is a collaborative work, which is still underway, and for which I participated for some analyses (listed below). The manuscript presented here is a draft, on which I commented but which I did not write.

In this analysis, I used the reference transcript generated by Carine Rey to align (using bowtie2 [Langmead & Salzberg, 2012]) and estimate expression levels (using htseq-count [Anders *et al.* , 2014]). I then performed multivariate analysis of the subsequent data to identify the main axis of variation between these 64 transcriptomes. More precisely I performed principal component analysis, together with *between-group* analysis, -a particular case of a principal component analysis, to extract variation from one specific factor. I used it to quantify the percentage of variation associated to the species effect and to the upper-lower effect. I used DESeq2 to identify genes that are differentially expressed in upper and lower molars throughout development in mouse and hamster separately and then compare the number of genes differentially expressed at every stages of development in both species. I also compared the upper-lower difference in mouse and hamster.

This chapter presents two main parts. First, a part on the evolution of developmental program between mouse and hamster molars in the form of an article that aims to be submitted soon. Second, a part on common upper-lower transcriptomic difference in mouse and hamster (see from *p. 142*). This part aims at identifying in more details the part of the developmental programs that are either conserved or that differs between mouse and hamster molar.

# DRAFT. Provisional title: Comparison of developmental genome expression in rodent molars reveals extensive developmental system drift.

Marie Sémon[1]*, Laurent Guéguen[2], Coraline Petit[1], Carine Rey[1,2], Anne Lambert, Manon Peltier, Sophie Pantalacci[1]*

[1] *UnivLyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, Laboratoire de Biologie et Modélisation de la Cellule, 15 parvis Descartes, F-69007, Lyon, France* [2] *Laboratoire de Biométrie et Biologie Évolutive (LBBE), Université de Lyon, Université Lyon 1, CNRS, Villeurbanne, France* *Equal contribution, and corresponding authors

## Abstract

In evolution, it is widely believed that phenotypic changes root in developmental changes and phenotype conservation, in developmental conservation. Seeming phenotype conservation may however hide evolutionary changes in the underlying developmental mechanisms by which a trait is produced. This cryptic evolution is also called Developmental System Drift, and the extent of this phenomenon unclear. We used a well-characterized evo-devo model system, rodent molars, to test the correlation between phenotypic and developmental evolution. Between mouse and hamster, the morphology of the lower molars has much less diverged than the morphology of the upper molars. Is development accordingly more conserved? We compared molar crown formation with a standard approach, and with a tight transcriptome time-series to get a quantitative molecular profiling of developmental states. Our data identify common trends in the development of all molars. Upper and lower molars have their specificities since the early steps of morphogenesis, at the levels of the pattern of cusp formation, cellular composition and biased gene expression. The extent of difference in lower vs. upper molar development within one species does correlate with the extent of difference in final morphology. However, the specificity of lower vs. upper molar development is drowned among the rapid evolution of development, which is highly species-specific in term of expression levels and temporal profiles. Divergence in developmental systems is almost as high for lower as it is for upper molar, despite much lesser morphological changes in lower molar crown. Hence, our results point an extensive drift in this developmental system. Because serial organs are largely sharing gene networks, it supports previous theoretical work that suggest a causal link between pleiotropy and DSD.

## Introduction

Comparing embryonic development to understand the evolution of forms is at the roots of evolutionary biology. A generally accepted idea is that the variation in development between species should be proportional to their variation in adult shape. When a phenotype differs between species, it is tacitly assumed that the development differs. When

the phenotype is similar in different species, then it is tacitly assumed that they have identical underpinnings, and little attention is given to the comparison of their development. But even when a phenotype is similar in different species, this seeming conservation may hide evolutionary changes in the underlying developmental mechanisms by which that trait is produced. This cryptic evolution is also called Developmental System Drift (DSD, [11, 12, 32, 17]). An emblematic example is the development of nematodes vulva, where quantitative changes of the gene signalling network in different nematode species have no impact on the patterning of epidermal cells [17]. Another example is the segmentation in larval development of different dipteran insects [36]. At a first glance, DSD is a less intuitive concept than developmental conservation. It may be useful to remember that natural selection mainly acts on the final product of development. Then, if divergent developmental paths are used to reach the same final phenotype, this should not make a big difference. Further taking into account that genomes are constantly mutating, DSD appear as a likely alternative to developmental conservation.

Because empirical evidence is still scarce, the extent of developmental conservation, and reciprocally, the prevalence of DSD is still unknown in most systems. Furthermore, how developmental gene expression is globally modified in the case of DSD remains unexplored. The factors that favour or prevent DSD are also largely unknown: for instance, theoretical work suggests that pleiotropy, which characterizes developmental genes, may be a factor facilitating DSD [23].

Rodent first molar is an appropriate model for asking whether variation in development mirrors variation in crown morphologies or, on the contrary, whether DSD is prevalent. Basic steps of tooth development have been described a long time ago, and lower molar morphogenesis is well known in mouse ([4, 16, 28] Figure 1-A): Initially, the jaw homeotic-code is transferred to the molars presumptive field (upper: Dlx1,2; lower: Dlx1,2,5,6; molar: Barx1). Later in development, dental epithelium grows and folds to first embed the dental mesenchyme and then form the cusps. Cusps initiate in a sequential manner in the growing tooth germ, and are marked by a signaling center called secondary enamel knot (SEK; Figure 1-A). This involves the iterative use of a regulatory network with key signalling pathways which is thought to be evolutionary conserved [16]. Importantly, crown shape is determined by morphodynamic mechanisms, which is a mode of development where patterning, morphogenesis and growth occur concomitantly. Morphodynamic development confer particular variational properties [29]. For example, the genotype-phenotype relationship is far from straightforward: close genotypes may underlie distant phenotypes, and reciprocally.

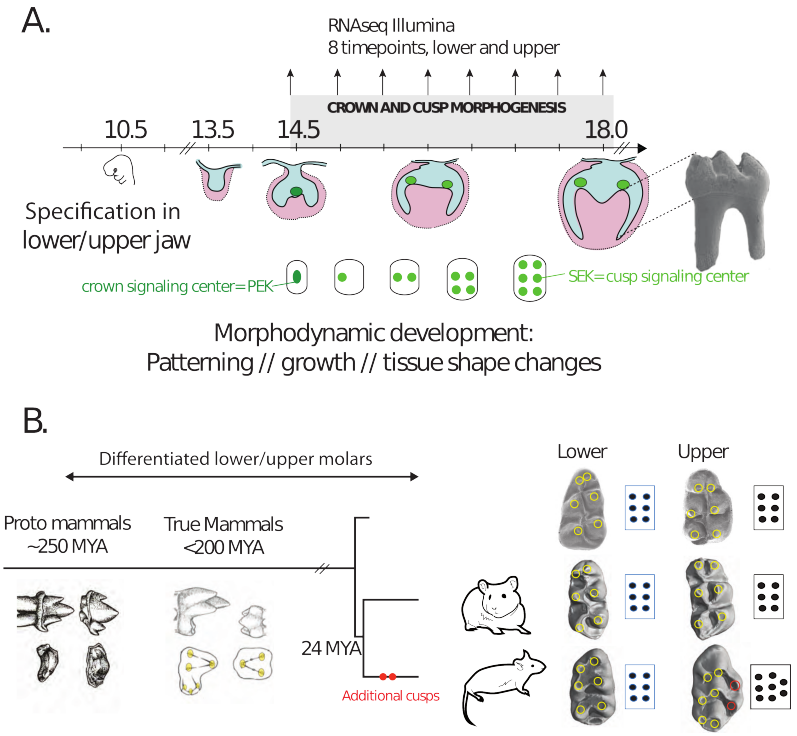



**Figure 1.** Main characteristics of rodent molar crown evolution and development. A- Molar crown morphogenesis involves morphodynamic mechanisms, where patterning and growth are concomitant. The dental epithelium (blue) grows and folds to first embed the dental mesenchyme and later form the cusps. Cusps initiate in a sequential manner and are marked by a signaling center called secondary enamel knot (SEK). B- Morphological evolution of lower/upper molars. Upper and lower molars displayed different morphologies since their origins in early mammals. In the mouse lineage, the upper molar was drastically transformed with the acquisition of a third cusp row (red). In contrast, the lower molar displayed much more limited morphological changes. In hamster, both the lower and upper molar kept a dental plan similar to the ancestral plan.

The evolution of molar morphology is extremely well documented. Thanks to their abundant fossil record, teeth have had a central role in documenting the evolution of vertebrates since the 19th century. In mammals particularly, molar crown shape provides characters for both fine and large-scale phylogenetic reconstructions. It can also directly reflect adaptations such as changes in diet, with typical features being recurrently found in carnivorous or herbivorous species [16]. More recently, the molar has become a forefront model in evo-devo, with the development of an *in silico* model of tooth development linking variation in development with variation in morphology [31]. Within rodents, many changes have been reported, with alternating periods of stasis, gradual changes and more rapid morphological evolution [25, 24]. One particularly drastic change occurred in the mouse lineage: the crown of the upper molar was drastically transformed with the acquisition of a third cusp row, making a pattern of cusp number and arrangement which is called the “murine dental plan” (Figure 1-B [18]). In contrast, the lower molar displayed much more limited morphological changes. In ham-

ster, both the lower and upper molar kept a dental plan which is similar to the ancestral plan. Hence, in mouse and hamsters, the crown shape in lower molars has been much less modified in evolution than in upper molars. Is development accordingly more conserved in lower molars?

We quantify similarities in development during the period of crown morphogenesis, and interpret them in relation with variation in adult morphology. We combine a descriptive approach with a cusp marker gene followed in large embryonic series, and a comparative transcriptomics approach on a tightly sampled RNAseq dataset. The extent of difference in lower versus upper molar development within one species does correlate with the extent of difference in final morphology. However, the specificity of lower versus upper molar development is drowned among the rapid evolution of development, which is highly species-specific in term of expression levels and temporal profiles. Divergence in developmental systems is almost as high for lower as it is for upper molar, despite much lesser morphological changes in lower molar crown. Hence our results are fully consistent with an extensive developmental system drift in this system.

## Results

### Patterns and timing of cusp formation is overall conserved in all types of tooth, with specificities attributable to history and morphological differences

To measure the variation in hamster and mouse crown molar development, we started by investigating the timing and pattern of cusp patterning. For this purpose, we followed the spatial pattern and the temporal dynamics of cusp formation with series of fixed embryos hybridized against the cusp marker *Fgf4* and dated according to embryo body weight. We collected 153 lower (and 141 upper) molars in mouse, and 86 lower (87 lower) molars in hamster, which makes a dataset amenable to quantitative analysis. We then built an in house method, based on Markov models, to infer from these data the amount of time spent in each stereotyped stage (see methods). The rationale underlying this method is very simple: under the hypothesis that embryos are sampled uniformly during development, long-lasting developmental stages should be sampled more often than ephemeral stages. An interesting feature of this model is to handle statistical comparisons of the relative duration of the different stages, within and between species and tooth types.

Cusp formation exhibits clear spatial and temporal similarities in all types of tooth (Figure 2). The first cusp patterned is bucco-central in all cases (paracone for lower mo-

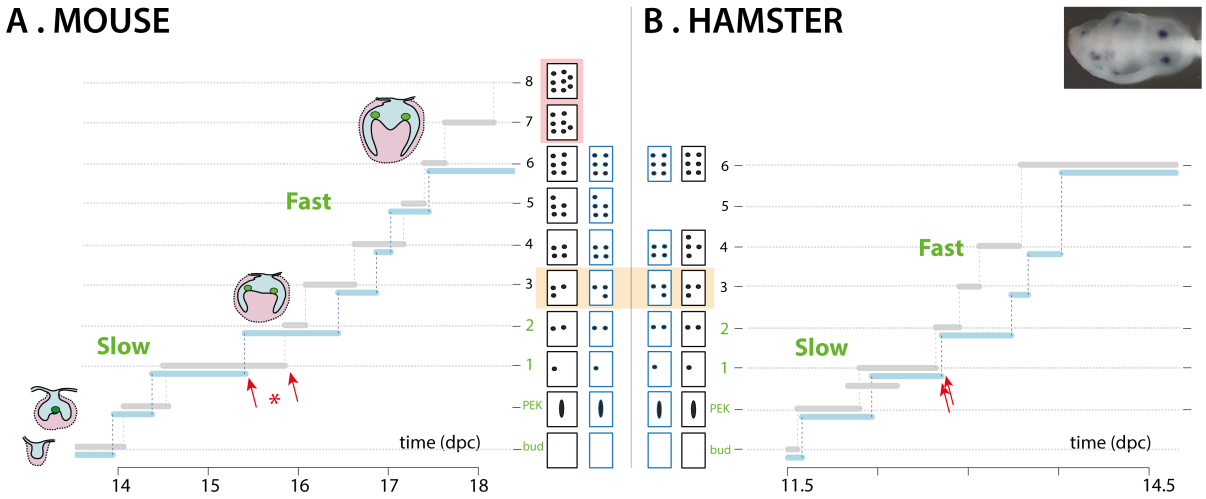



**Figure 2.** Conservation and divergence in dynamics and pattern of cusp formation in mouse (left) and hamster (right). Pattern and dynamics of cusp formation was followed in series of fixed embryos hybridized against the cusp marker Fgf4 and dated according to embryo body weight. We built an in house method, based on Markov models, to infer from these data the amount of time spent in each stereotyped stage. In mouse (left) the best Markov model fits cusp patterning with a succession of stages with 3 different durations represented by the length of blue and grey bars for lower and upper molars respectively. In hamster, the best models have 2 different durations. Globally the sequences are resembling in both species, with for the lower molar, two long first stages ( $1 >>> 2 >>> 3 > 4 > 5 > 6$) and for the upper molar, an extra long duration of the 1 cusp stage, but a short 2-cusp stage ( $1 >>>> 2 > 3 > 4 > 5 > 6 (> 7 > 8)$.

lar, protoconid for upper molar), followed by a second cusp at its lingual side. The third cusp is posterior is all cases, with later addition to the anterior cusps. The similarities extend also to the rates of cusp addition. The best Markov models fit cusp patterning with 3 different rates in mouse and hamster (LRT to compare model with 3 to model with 2 rates, upper mouse: p-value = $1.710^{-9}$; lower mouse : p-value = $3.510^{-5}$, upper hamster p-value = 0.008, lower hamster p-value = 0.05). In all cases, the rate of cusp acquisition is biphasic, early cusps being added slowly, and then more rapidly in late crown morphogenesis.

Interestingly, lower and upper molar present distinguishing features, the most notable being the third cusp patterned buccally in upper molars, and lingually in lower molars. Our preliminary results suggest that this is also true in gerbils and acomys (data not shown). Hence, the patterning of the 3 first cusps show a deep conservation (lingual patterning of the second cusp) and a feature distinguishing lower and upper molar (third cusp). Altogether, the conservation of this specific patterning of is possibly reminiscent of early mammalian molar evolution. A further specific characteristic of upper molars is the duration of the 2-cusps stage which lasts 3 to 7 times more in lower by comparison to upper molars (Markov model estimates).

Because two supplementary cusps add to the upper molar dental plan in the mouse lineage, we searched for individual characteristics in mouse upper tooth development. These additional cusps form last, which is conform to expectations of the model of terminal addition where most recent evolutionary features should be added in the later steps of development (see discussion). The spatial pattern of the upper crowns at 4-cusp stage differs between mouse and hamster. Unexpectedly, earlier stages of upper molars development also display specificities, with a significant delay in the patterning of the second cusp (stage 1 is twice longer in mouse compared to hamster, as estimated from the Markov models). Conversely, lower molar development is very similar in mouse and hamster, which mirrors the fact that their dental plan is similar.

Overall, cusp formation exhibit clear similarities in all teeth, with specificities for lower versus upper molar, possibly reminiscent of early mammal molar evolution. As expected from morphology, lower molar development is very similar in mouse and hamster. Mouse additional cusps form last, according to their phylogenetic history (recapitulation [1]). Clear differences are also seen earlier in upper molars (stage 1 and 4). Finally, this morphological analysis of cusp patterning shows clear specificities of upper and lower tooth and comparatively much less characters distinguishing the species. Is this pattern attributable to evolutionarily conserved underlying gene regulatory networks? Temporal genome expression is now fully recognized to register the course of development [2] and transcriptomes are widely used to quantify differences between species and organs e.g. [27, 38, 19, 14, 6, 5]. Next we constructed a transcriptome dataset to compare upper and lower molar development in mouse and hamster.

## Lower and upper molars in mouse and hamster largely share a common dynamics of genome expression but also exhibit distinguishing features

We prepared RNA-seq libraries from upper and lower whole tooth germs in mouse and hamster, corresponding to 8 development stages roughly equally spaced during the period of crown development. We quantified expression levels of 10,516 protein-coding genes in mouse and hamster, for which the sequence of the transcript was made rigorously similar, despite the fact that no genome sequence is available in hamster (see methods). In a first step, we conducted the analyses separately for each species. Visually inspecting the projection of the transcriptome samples into a space of reduced dimension has become a classical first step in comparative transcriptomics [21, 2, 5]. PCA plots in each species show that the main source of variation reflects a temporal dynamics which is common to both lower and upper molar (31.7% in hamster, 29.2%

in mouse; Figure 3).

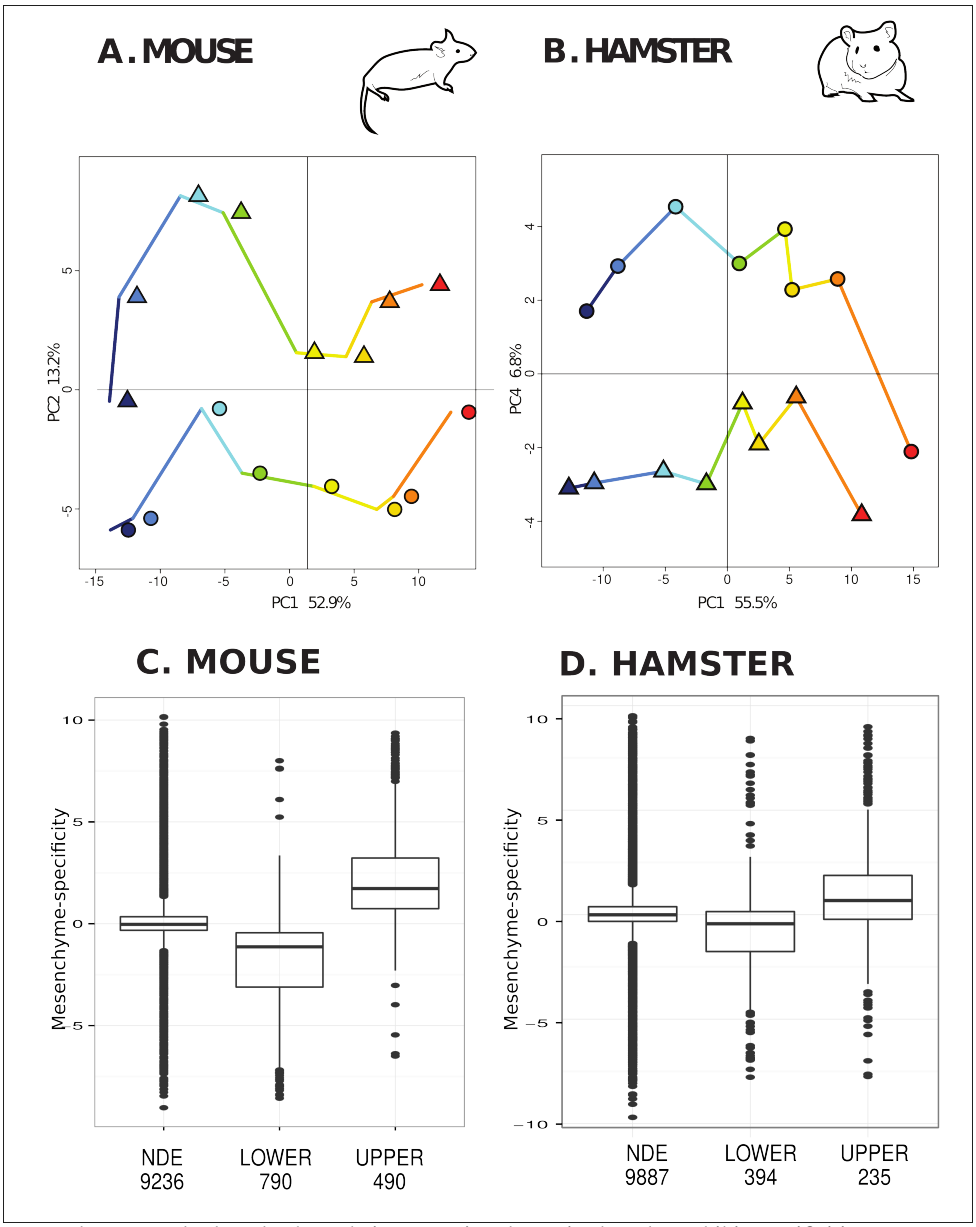



**Figure 3.** Lower and upper molar largely share their expression dynamics but also exhibit specificities, some are conserved in mouse (A) and hamster (B). (Top) Map of 16 transcriptomes from upper (triangles) and lower (circles) molar germs, at 8 stages of development, on the two first principal components of a PCA analysis (orthologous sets of 10,516 genes in each species). PCA1 orders samples with time, hence represents shared temporal variation. PCA2 separates upper and lower molar germs. The percentage of variation explained by each axis is indicated. (Bottom) Upper genes are more mesenchyme specific and lower genes are more epithelium specific in mouse tooth germs, but not in hamster. Upper and lower genes were defined based on classical test for differential expression between upper and lower molar, taking developmental stage into account (adjusted p-value $< 0.05$). The number of genes in each category is indicated. Mesenchyme specificity was estimated as the ratio of mesenchyme and epithelium expression levels from pure tissue transcriptomes (log2 fold change).

A clear organ identity is also visible from the second axis of the PCA plot, which

is separating upper and lower samples and represents 11.6% of the total variation for hamster, and 16.9% for mouse. The ordinations of the genes on the main PCA axes, which were obtained independently from separate analysis in mouse and hamster, are nonetheless significantly correlated (measurement of the global correlation by coinertia analysis, p-value $< 10^{-3}$). The correlation between mouse and hamster are R = 0.39 for PCA1 (p-value $< 10^{-16}$) and R =0.14 for PCA2 (p-value $< 10^{-16}$). This is congruent with an overall correlation of the ratios of upper/lower expression between species (Pearson R = 0.31, p-value $< 10^{-16}$). An examination of the lists of differentially expressed genes between lower and upper tooth germs reveals a moderate but significant overlap between mouse and hamster: we observed about 3.4-5.2 times enrichment between the lists of DE genes depending on the threshold used (respectively adjusted p-value = 0.1 and 0.05; Chi-squared test p-value $< 10^{-16}$).

A temporal analysis of the main clusters of coexpression also reveals a visible tooth type "identity", with clusters in the lower molars being smoother than clusters in the upper molar (Figure 4; see later analysis of the conservation of these clusters). In a previous analysis in mouse, we had shown that this corresponds to processes of morphogenesis being transiently more active in upper than in lower molars. Here we confirm, using GO analyses, that early tooth germs likely contain less mitotic cells and more migrating cells in upper molar than in lower molar. This pattern is common to mouse and hamster. The timing of this process corresponds to a developmental stage where the transcriptome is most different between lower and upper molars (test of the polynomial term of the linear model, p-value = 0.005 see supplementary Figure 7). Most interestingly, this timing also matches the "3-cusp" developmental stage, where the difference in the arrangement of the third cusp patterned is the main specificity distinguishing lower from upper molars (Figure 2).

The second PCA axis, which separates lower and upper samples, is 45% stronger in mouse than in hamster (16.9% and 11.6%), as would be expected from the difference in adult morphologies. This is congruent with a between-group multivariate analysis (37% stronger, p-value = 0.0045 see Sup figure 7) and with differential analysis data, which detects at least 3 times more genes differentially expressed between upper and lower samples in mouse than in hamster (4.1 more for p-adjust=0.05, 3.70 for p-adjust=0.1). Our previous analysis of mouse data (Pantalacci et al, submitted) has shown that this second axis is partly explained by a difference in the relative mesenchyme proportion between lower and upper molars. This was based on marker genes and pure-tissue microarray data. To explore this further in mouse and compare with the patterns observed in hamster, we prepared RNAseq libraries from pure epithelium and mesenchymal cells

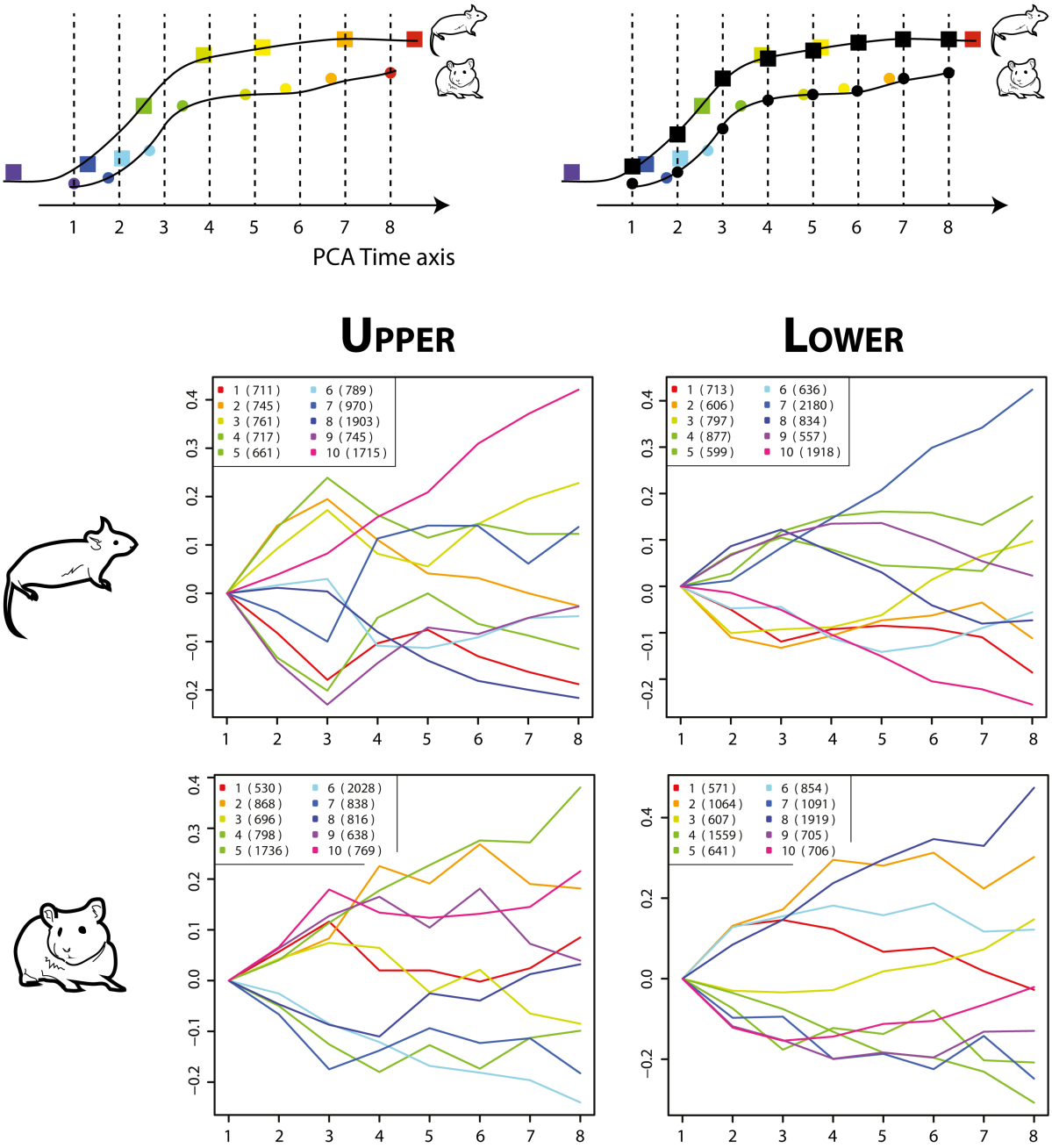



**Figure 4.** Ten main temporal profiles of expression determined independently for the lower (right) and upper (left) molar, in mouse (top) and in hamster (bottom). The second axis of PCA with all samples (Fig 3) was taken as a global index for developmental time. This PCA2 was split in equally sized 'time' windows. The timepoints were made comparable between species by estimating expression levels in each window. 10 clusters were obtained by K-mean clustering (see methods), and associated number of genes are indicated. Cluster numbering and colours are arbitrary and do not indicate a correspondence between the four conditions. Each profile represents the median of relative expression level (normalized with expression level at the first stage) for all genes associated to a particular cluster.

in both species (see methods). In mouse, this new pure-tissue data confirm our previous observations (Figure 3] and Sup figure 8): Using deconvolutions, we confirm that the relative proportion of mesenchymal cells is greater in upper than in lower tooth germs, the average of the 8 time points being respectively 53.1% and 44.0% for upper and lower molar in mouse (Wilcoxon test p-value = 0.000622). This quantification is based on 6 markers for enamel knot known from previous literature, 333 epithelial markers and 655 mesenchyme markers based on our pure-tissue transcriptomes (log2 ratio $> 3$ and adjusted p-value $< 0.1$), but the result is very robust to changes in marker lists. We computed an index of mesenchyme-specificity which is simply the log-ratio of mesenchyme and epithelium expression levels. This index is significantly correlated with the second PCA axis in mouse (Spearman Rho = 0.42, p-value $< 10^{-16}$) and to the average log ratio of upper and lower expression levels in mouse (Spearman Rho = 0.58, p-value $< 10^{-16}$). Of note, the nature of differentially expressed genes is impacted by this difference in tissue proportions. This is illustrated Figure 3, which shows that genes with a significant bias towards the upper molar are more mesenchyme-specific than genes biased towards the lower molar (Wilcoxon test, p-value $< 10^{-16}$). The median "upper" genes is 3.5 times more expressed in mesenchyme (median log2 fold change of 490 UPPER genes = 1.72) than in epithelium, while the median "lower" DE genes is twice more expressed in epithelium (median of 790 LOWER genes = -1.13). In mouse, the relative proportion of epithelial and mesenchymal cells has a pervasive effect on the difference between upper and lower transcriptomes.

The results of symmetrical analyses conducted in parallel in hamster stand in contrast with these observations, although the index of mesenchyme specificity are extremely well correlated between the two species (R = 0.799, p-value $< 10^{-16}$). The difference in mesenchyme tissue proportion between upper and lower molar is very moderate in hamster, which can be seen in mesenchyme proportions estimated from deconvolutions (respectively 49.9% and 49.2%, Wilcoxon test p-value=0.96, predicted using 6 enamel knot markers, 393 epithelial markers and 568 mesenchyme markers from pure tissue transcriptomes such as log2 ratio $> 3$ and adjusted p-value ¡ 0.1). This is also visible from the weaker correlation between PCA2 and the index of mesenchymal-specificity in hamster (Spearman Rho = 0.14, p-value $< 10^{-16}$). Mesenchyme-specificity is also much less correlated to the ratio of upper and lower expression levels (Rho = 0.19, p-value $< 10^{-16}$). In consequence, differentially expressed genes biased towards the upper molar are much less mesenchyme-specific in hamster than in mouse (median = 1.6 for 235 UPPER genes, Wilcoxon test, p-value $< 10^{-16}$, Figure 3).

In conclusion, these results on temporal genome expression mirror our findings on the dynamics and pattern of cusp formation, with a shared developmental program, lower/upper specificities found in both species and a supplementary difference between lower and upper tooth germs in mouse which is congruent with their final morphologies. The patterns and timing of cusp formation does not display any obvious species specificity. What is the situation for genome expression?

## Expression levels are highly species-specific, including for functionally relevant genes.

The PCA plot of a pooled dataset including all samples from mouse and hamster displays a stricking difference between species (Figure 5-A). Species effect explains 53% of variance, whereas developmental timing drops to 11% and the type of teeth only 4%. RNA-seq analyses are highly sensitive to technical artifacts, such as sequence composition and batch effects. To take care of the first issue, we carefully annotated orthologs based on phylogenies and cropped the sequences of the reference transcriptomes before measuring expression levels, so that they become fully comparable between species (see methods). A controversy related to the second issue erupted recently, when the Mouse ENCODE Consortium reported that comparative gene expression data from human and mouse tend to cluster more by species rather than by tissue on a PCA plot [20]. This analysis was challenged when a confounding sequencing batch effect was removed [10]. To prevent similar batch effects, we analysed an homogeneous part of the dataset (8 stages in mouse and in hamster), for which library preparation and sequencing were performed synchronously in mouse and hamster. This does not change the results (Sup figure 9).

This major species effect is also confirmed by other methods, such as “between” multivariate analyses that quantify the relative effects of the species (54.3% on the homogenized dataset) and of the tooth (2.33%). Using differential expression analysis, we find 8,387 genes that vary significantly according to the species, which is much more than the 130 genes that vary according to the upper/lower difference (DESeq2 analysis taking development stage into account, adjusted p-value $<0.1$).

Hence, when the dataset is taken globally, the variation in expression between species exceeds largely the variation in expression between organs. In pioneering comparative transcriptomics studies [37] similar patterns were interpreted as the product of rapid evolution of gene expression, and considered as an evidence for its neutral evolution.

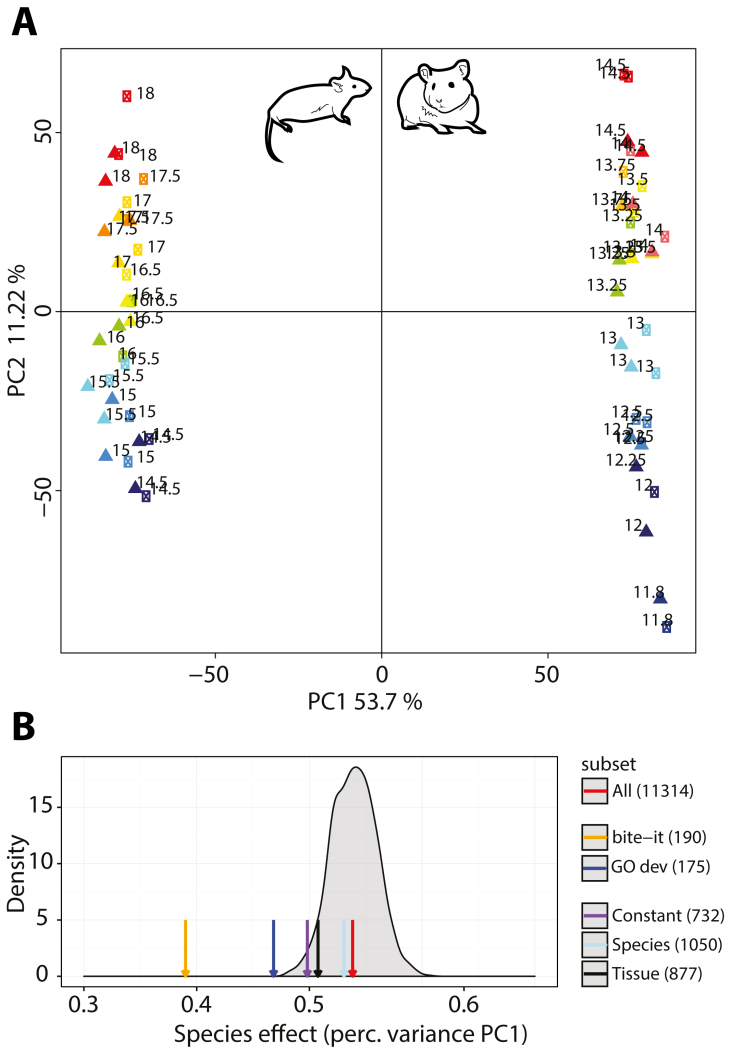



**Figure 5.** Species effect is the major source of variation when comparing all four transcriptomic series. A- Principal component analysis with 64 samples from both species and tooth types. Expression levels are so divergent between species that tooth type differences (circles and triangles) are negligible in comparison. PCA1 separates mouse samples (circled in black) and hamster samples (circled in white). The color gradient indicates the time (in days) which runs from 11.8 days to 14.5 days in hamster, and from 14.5 days to 18 days in mouse. PCA2 sorts the samples with developmental age. B- Distribution of the species effect for the whole genome and for specific subsets of genes. In grey: distribution of %variation associated to PCA1 for random samplings (200 genes) in the whole genome. Red arrow: value for the whole genome. Other arrows show the %variation associated to PCA1 for specific subsets, genes with acknowledged function in tooth/organ development (bite-it database, and "development" GO category), and genes with different evolutionary rates in adult organs (computed using the dataset from [5]).

According to this interpretation, the changes in expression level that are observed in our dataset may be relatively neutral, without much impact on molar development. However, most comparative transcriptomic studies [7, 5, 3, 22] now demonstrate an overall conservation of expression levels between species, at least for protein-coding genes, which is compatible with strong constraints on the evolution of expression.

Hence, we reasoned that the strong species effect that is observed in our data may be carried by a proportion of the genome only, while the rest is evolving under strong selective constraints. We randomly sampled subsets of 200 genes in the genome, and for each set of genes independently, we conducted a PCA analysis and measured the percentage of variance associated to PCA1 as an index of the species effect. The resulting distribution shows that species effect represents at least half of the total variation in the transcriptomes in all cases, and never less than 40%.

Then, we considered that the species effect should be even more pronounced than the genome average for genes known to be species-specific, and much smaller for genes known to play an important role in tooth development. From a large published dataset providing expression levels in adult organs across a wide range of species, we defined three sets of genes : genes for which expression level varies significantly according to the species (N=1,050), according to the tissue (N=877) and genes that do not vary (N=732 "Constant" genes; see methods). We then measured the species effect in our dataset for these three sets of genes. As expected, the species effect is greater for the set of "species" genes than for "constant" and "tissue" genes (Figure 5-B). But importantly, even genes whose expression level evolve slowly in adult organs, for which the evolution of expression level is unlikely to be neutral, also show a strong species effect (>50%). We then reasoned that changes in the expression levels of many genes may not have much consequence on tooth development. We then measured the species effect on the group of genes annotated to "development" function in Gene Ontology (N = 175 genes), and finally, on a set of genes functionally relevant for tooth development (190 genes from the bite-it database). Even on these sets of genes which are much likely to be important for tooth development, there is a very strong effect (40-46%). In conclusion for this part, expression levels are so divergent between species that tooth type differences are negligible in comparison.

## Not only expression levels, but also temporal dynamics is highly species-specific, including for functionally relevant genes.

Decades of studies in evolution and development have highlighted that temporal profiles of gene expression, and not just their expression levels, is instrumental for correct organ development. Hence we compared the profiles of expression in hamster and mouse lower and upper molars. Embryo were sampled so that developmental stages are as comparable as possible between species, and we further refined staging based on transcriptome developmental clock (see methods). Then, we build the ten main expression

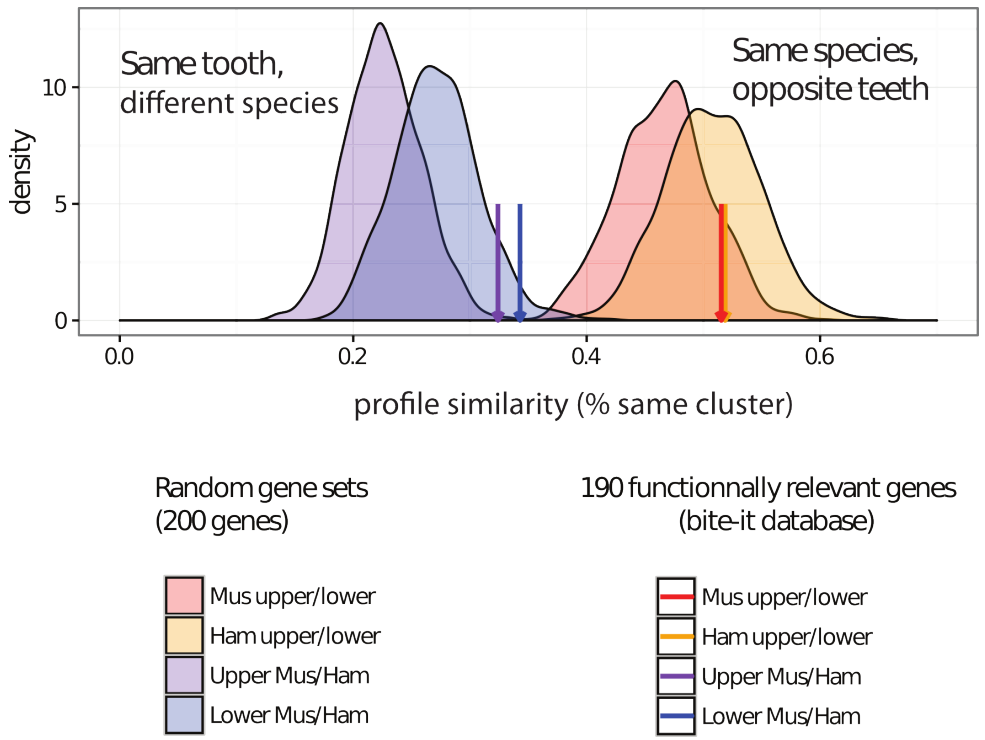



**Figure 6.** Temporal dynamics is highly species-specific, including for functionally relevant genes. X-axis indicates the similarity of expression profiles between pairs of conditions, as measured by the proportion of genes that adopt the same broad temporal profile of expression. Histograms represent the distribution obtained by randomly sampling sets of 200 genes, and the color indicates the type of comparison. The arrows represent the profile similarities obtained for functionaly relevant genes (190 genes from bite-it database).

profiles independently for each tooth (by clustering analysis, Figure 4) and computed the resemblance between the dynamics of transcriptome in the four conditions. Let's consider we are comparing upper molars in mouse and hamster: For each gene, we took its gene expression profile in the upper molar (VUp), and calculated the correlation of this vector VUp with the 10 main expression clusters drawn from lower molar expression. This way, we obtain the cluster with which the vector VUp is best associated. By construction the expression profile of the same gene in the lower molar (VLo) is associated with one of the ten main expression clusters. Are VUp and VLo associated with the same cluster of expression? Through this simple reasoning, we build an index of similarity between temporal profiles. We sampled randomly sets of 200 genes, to build a distribution of the resemblance in temporal profiles. Lower and upper molar of the same species have twice more resembling temporal profiles than lower molars or upper molars of different species (Figure 6). This is again a strong indicator that profiles of expression are more specific to the species than to the type of tooth. The same effect is seen with a set of genes functionally relevant for tooth development (190 genes from bite-it database). Gene expression temporal dynamics is poorly conserved for a given

tooth type, whereas there are extensive similarities between the two teeth of a given species. This is best explained by Developmental Systems Drift (see discussion).

Furthermore, among the 30% of genes whose profile is similar in the same tooth of different species, 75% have in fact a similar profile in all four teeth. This leaves only 7% of genes with a conserved lower or upper specific profile. Then, a minor proportion of the genes display a profile of expression which distinguishes lower and upper molars. Lower molars (6 cusps in mouse and hamster) have slightly more resembling temporal profiles than upper molars (6 cusps in hamster, 8 cusps in mouse: the average profile similarity as measured Figure 6 is 20% higher). Hence, expression dynamics is slightly more conserved in lower molars than in upper molars. This is coherent with adult morphologies, but unexpectedly small : From the adult morphologies, much sharper difference could have been expected.

## Discussion

Like in many other developmental systems, global properties of molar development are conserved for very long periods of time. For instance, playing with levels of signaling molecules, mouse teeth have been engineered in vitro to generate tooth shapes which are widespread in the mammalian fossil record [13]. Moreover we know that molar development in different mammal species use similar set of key signaling pathways and transcription factors [15]. Then tooth development was expected to be strongly conserved, especially in closely related species like mouse and hamster (estimated divergence time = 24 MYA [18]). Here, as a read-through of the development process we quantify the similarities in genome expression during molar crown morphogenesis in mouse and hamster. We find that expression profiles are twice more similar within species than within tooth type. This is also true for developmental genes, even for genes known to function in tooth development. This is even more pronounced for expression levels. How to interpret this observation? We did expect a difference in upper molar developmental processes, because the crown morphology has been drastically transformed in mouse. Seeing that the development was nearly as different between lower molars, is much more astonishing. Such rapid evolutionary changes in the developmental programs of both molars, without relation to the final morphology, is best explained by Developmental Systems Drift. Below we hypothesise that the molars, and serial organs in general, may be particularly prone to DSD.

Lower and upper molars morphological evolution was described up to their origins in early mammals. They already displayed different morphologies at the time [34]. Hence, it would be far-stretched to pretend that they have ever shared the same developmental

mechanism. But it is both expected [35, 34]) and observed (Pantalacci et al, submitted) that similar gene regulatory network are directing the development of upper and lower first molars. Indeed, different types of teeth in mammals, but also fore- and hindlimbs, plant leaves and insect appendages, are serial organs, that is, multiple realisations of a very similar developmental program, which share the same genes except 1-3 "selector" genes (homeotic genes). This causes extreme pleiotropic effects. A mutation at one gene can affect the development of all iterations of the same serial organ [23]. In accordance with this idea, upper and lower molar shape covary in wild rodent populations/species [26]. Intuitively, one may think that such gene pleiotropy will induce a strong conservation of the development (Fisher's models) by the effect of a strong purifying selection. But at the same time, the idea that pleiotropy would strongly constrain the evolution of these organs is contradicted by the fact that the morphology of different types of serial organs is extremely variable in the same organism and in evolution. Counter-intuitively, pleiotropy may induce extensive developmental system drift, according to a model proposed by Pavlicev and Wagner [23]. This model, termed "selection, pleiotropy and compensation model" (SPC model), is in two steps. One genetic change which has an strong adaptive effect on one organ is fixed (Step 1). Because of pleiotropy, it is associated with antagonist (deleterious) effects on other organs. Hence, this initial adaptive change is followed by subsequent selection in favor of supplementary changes to compensate these pleiotropic effects (Step2). Importantly, although such compensatory changes are selected for by stabilizing selection, they amplify the divergence in development between species, producing extensive developmental systems drift. Serial organ pleiotropy is very high, thus this hypothetic mechanism could work strongly and serial organ be prone to DSD. In our specific case, accommodation of lower molar for drastic upper morphological change in mouse may have further amplified DSD. DSD may facilitate differential morphological evolution of serial appendages, because it is enabling one appendage to cope with extensive changes in the other.

Furthermore, DSD may be facilitated by because molars have a peculiar mode of development. During tooth development, signalling and tissue growth are intermingled, in what is called a morphodynamic mechanism. It has been proposed, based on modelling and experimental data, that Genotype-Phenotype relationship is complex in these cases, with several genotypes producing the same phenotype [30, 31]. In other words, different developmental trajectories can lead to the same phenotype. DSD may be facilitated in morphodynamic systems like rodent molar [33]. We note that many serial organs develop through morphodynamic mechanisms (arthropod appendages, limb, teeth, leaves…). Hence the model proposed here may apply to other serial appendages

in general. To push the model of Pavlicev and Wagner further, we suggest that DSD and morphodynamic mechanisms may work hand in hand to alleviate the pleiotropic constraint. This may have render possible the emergence of generalized pleiotropy in developmental systems, while preserving their evolvability.

## Methods (preliminary draft)

### Mice breeding and embryo sampling

CD1 (ICR) adult mice and AURA adult hamsters were purchased from Charles River and Janvier Laboratories, respectively (France). Females were mated overnight (night at 20:00 pm) and the noon after morning detection of a vaginal plug was indicated as embryonic day (ED) 0.5. Other breeding pairs were kept in a light-dark reversed cycle (night at 12:00 pm), so that the next day at 16:00 pm was considered as ED1.0. Pregnant females were killed by cervical dislocation (mice) or medicamentous overdose (hamster) and embryos were harvested on cooled Hank's or DMEM advanced medium, weighted as described in [73] and immediately decapitated. This study was performed in strict accordance with the European guidelines 2010/63/UE and was approved by the Animal Experimentation Ethics Committee CECCAPP (Lyon, France; reference ENS_2012_045).

### Epithelium dissociations and in situ hybridizations

Complete or hemi mandibles and maxillae were dissected in Hank's medium and treated with Dispase (Roche) 100% at 37°for 1h30 to 2h20 depending on embryonic stage. Epithelium was carefully removed and fixed overnight in PFA 4% at 4°. DIG RNA antisense Fgf4 probe were prepared from plasmids described elsewhere [74]. In situ hybridization was done according to a standard protocol. Photographs were taken on a Leica M205FA stereomicroscope with a Leica DFC450 digital camera (Wetzlar, Germany) or on a Zeiss LUMAR stereomicroscope with a CCD CoolSNAP camera (PLATIM, Lyon).

### RNA-seq sample preparation

In mouse, 16 samples were prepared for upper and lower whole tooth germ RNA-seq analysis. This adds to the 16 samples described previously (Pantalacci et al., submitted). We obtained a total of 32 samples representing 8 stages (ED14.5, 15.0, 15.5, 16.0, 16.5, 17.0, 17.5, 18.0), coming from 8 individuals. In hamster, 32 samples were prepared, representing roughly 8 stages (from ED12.5 to 14.5), coming from uppers and lower first molars of 16 individuals.

Each sample contained 2 tooth germs, the left and right first molars (M1) of the same male individual, and for a given stage, the upper and lower samples came from the same individual. The heads of harvested embryo were kept for a minimal amount of time in cooled PBS (small scale) or advanced DMEM medium (large scale). The M1 lower and upper germs were dissected under a stereomicroscope and stored in 200ul of RNA later (SIGMA). Similarly dissected tooth germs from the same litter and same weight were fixed overnight in PFA 4% for immunolocalization and 3D reconstruction (see later). Another embryo of the same litter and same weight was processed as indicated above for Fgf4 in situ hybridization. For epithelium and mesenchyme specific samples, tooth germs were dissected as above, then the two components were separated following a dispase treatment (15 minutes, 37°, ROCHE) and immediately stored in RNA later (SIGMA).

Total RNA was prepared using the RNeasy micro kit from QIAGEN following lysis with a Precellys homogenizer. RNA integrity was controlled on a Bioanalyzer (Agilent Technologies, a RIN of 10 was reached for all samples used in this study). PolyA+ libraries of the large-scale dataset were prepared with the Truseq V2 kit (Illumina), starting with 150ng total mRNA and reducing the amplification step to only 12 cycles and sequenced (100 bp paired-end sequences) on an Illumina Hi-seq2000sequencer at the GENOSCOPE (Evry, France). Raw data is available in the gene expression omnibus repository (XX).

## Expression levels estimation using RNA-seq and differential expression analysis.

Comparable sequences for 10,516 pairs of orthologous genes were extracted by an automated and accurate pipeline (Amalgam, GitHub `https://github.com/CarineRey/apytram`) which combines data from RNA-Seq and existing multi-species gene family alignments. In Amalgam, reconstructed sequences are first attributed at the good gene family and then aligned in the corresponding existing multi-species gene family alignment. Then an accurate gene tree is inferred to use phylogenetic proximity between sequences as gene annotations. This allows taking into account complete genomic knowing of this clade (duplication, and losses) and not of an unique reference species. Reads were mapped to theses sequences using Bowtie2 (version 2.0.2) with standard settings. We obtained a median of XX million reads that mapped uniquely.

Raw counts for the published datasets by Brawand et al [5] were obtained, and anal-

ysed by DEseq2 to obtain the lists of genes differentially expressed with species or tissue factor. A subset of the whole dataset (primates and rodents only) was used in this analysis.

Two main analysis used in the text include differential gene expression analysis (DE) and multivariate analysis. Differential expression analysis was performed using DEseq package (DESeq2, version 1.6.3 [21]), with settings indicated in the main text. Multivariate analyses (principal component analyses, between-class analysis and coinertia analysis) were performed using ade4 package (ade4_1.6-2 [8]).

## Gene Ontology analysis

Gene ontology (GO) analysis were performed and visualized with GORILLA and GOStats (version 2.34.0, with `Mmusculus.UCSC.mm10.ensGene_3.1.2` and GO.db_3.1.2), using the full list of genes expressed in the corresponding dataset as a background.

## Clustering of upper and lower time-series and estimation of similarity between temporal profiles in pairs of conditions.

First, to allow the comparison of temporal profiles between different species and different tooth types, we constructed a dataset with comparable timepoints. For this, we considered the second principal component (PCA2) as an index of developmental timing. We divided the range of coordinates on PCA2 in 8 windows of equal size, and for each gene and each window we extrapolated the expression level that is expected at each window boundary (taking the median expression levels of the samples that belong to this window do not change qualitatively the results). This is schematized on the top of the figure 4.

Then, for each gene and at each stage of development, we took the level of expression divided by the initial level of expression (i.e. level at stage ED 14.5). To classify the genes we first generated theoretical profiles representing all possible combinations of transitions between stages (expression level being either increasing, decreasing or flat between the 8 consecutive stages), making up a total of 2,187 profiles. We then correlated each real expression profile to each theoretical profile using Spearman correlations. We clustered the obtained correlation matrix using K-means (choosing 10 clusters, and checking that the obtained profiles are repeatable over several iterations of the method). The median of the expression profiles of these 10 clusters are drawn on figure 4. Similarities between the different time-series was computed as follows, taking

the example of lower/upper comparison in mouse, but same reasoning was applied to all comparisons: We computed the number of genes which lower timeprofile is robustly assigned to one of the 10 main "lower" clusters (Pearson R>0.7). We counted, the number of genes whose timeprofile in upper molar associates best (and with R>0.7) to the same lower cluster: such genes have similar timeprofile during lower and upper molar development. The ratio of this number to the total of genes is an index of profile similarity between lower and upper molars un mouse.

## Estimating tissue proportions from RNAseq data : deconvolutions

Performing gene expression deconvolution, is estimating cell type proportions (and/or cell-specific gene expression signatures) from global expression data in heterogeneous samples. Here, we used the R package CellMix ($CellMix_1$.6.2, [9]), to estimate the proportions of the three tissue compartments in our tooth germ transcriptomes. CellMix implements in particular the method Digital Sorting Algorithm (DSA) which performs complete gene expression deconvolution using a set of marker genes only.

We needed therefore to define 3 sets of markers, as specific as possible for each tissue compartment (epithelium, mesenchyme, enamel knot). We used 6 enamel knot markers with a large pattern of expression obtained from bite-it database (Fgf4, Slit1, Wnt3, Wnt10a, Dkk4 and Fgf20). Epithelium and mesenchyme markers were defined separately for mouse and hamster, based on our pure-tissue RNAseq data. We selected genes for with significant difference in expression levels between mesenchyme and epithelium (adjusted p-value $< 0.1$) with a log fold change greater than 5. The result does not change depending on the threshold used (3x up to 7x). The threshold used was set up to guarantee enough markers relatively equilibrated among the 3 tissues, and yet that are specific enough for the tissue under study. To get an idea of the sensitivity of the results to the choice of the set of markers, for each analysis, we computed 500 bootstrapped values by randomly resampling half of the epithelial and mesenchymal markers and recomputing deconvolutions on these 500 random subsets. The average of these resamplings are indicated in the barplots (Sup figure 8).

## Modeling the timing of cusp patterning

We modeled the timing of cusp patterning as a continuous time Markovian process, with successive states the numbers of cusps, from 0 to the final number. Each state $N$ has a specific rate of exit $q_N$, to the next state $N+1$. It means that sojourn time in $N$ follows an exponential distribution of mean $\frac{1}{q_N}$. The final state is the absorbing state. From this,

we build the generator $Q$ of the Markovian process of succession between states. The probability to observe a number $n$ of cusps at time $t$ is the value of transition probability matrix $e^{Qt}$ at coordinates (0,n). If all states have the same mean sojourn time, all parameters $q_N$ are equal, and the model has only one parameter. On the contrary, when all $q_N$ are independent, there are as many parameters as the final number of cusps.

Concretely, the input data corresponds to N=141/153 teeth sampled for upper/lower molar in mouse, and 87/86 teeth sampled for upper/lower molar in hamster. On each tooth, patterned cusps were counted on Fgf4 in situ hybridizations of the dissociated epithelium. The development time was estimated in each sample through the weight of the embryo (in milligrams). In hamster, the relationship is the following:

$$Dev.Time.Hamster = \frac{weight^{0.23} + 5.3791}{0.7426}$$

The best transformation was obtained by a boxcox procedure, but is essentially similar to squared root of the weight. In mouse, the relation is the following :

$$Dev.Time.Mouse = \frac{weight^{0.23} + 2.4141}{0.4192}$$

## Statistical analysis

Statistical analysis was performed using R (version 3.2.2, 2015-08-14), with ggplot2 for graphics ($ggplot2_1.0.0$). The data and R scripts, which permit to reproduce the figures and tests presented here, are provided in supplementary material XX.

## References

[1] Arhat Abzhanov. Von Baer's law for the ages: Lost and found principles of developmental evolution. *Trends in Genetics*, 29(12):712–722, dec 2013.

[2] Leon Anavy, Michal Levin, Sally Khair, Nagayasu Nakanishi, Selene L Fernandez-Valverde, Bernard M Degnan, and Itai Yanai. BLIND ordering of large-scale transcriptomic developmental timecourses. *Development (Cambridge, England)*, 141(5):1161–6, mar 2014.

[3] Nuno L. Barbosa-Morais, Manuel Irimia, Qun Pan, Hui Y. Xiong, Serge Gueroussov, Leo J. Lee, Valentina Slobodeniuc, Claudia Kutter, Stephen Watt, Recep Çolak, TaeHyung Kim, Christine M. Misquitta-Ali, Michael D. Wilson, Philip M. Kim, Duncan T. Odom, Brendan J. Frey, and Benjamin J. Blencowe. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science*, 338(6114), 2012.

[4] Marianna Bei. Molecular Genetics of Tooth Development. *Curr Opin Genet Dev*, 19(5):504–510, 2009.

[5] David Brawand, Magali Soumillon, Anamaria Necsulea, Philippe Julien, Gábor Csárdi, Patrick Harrigan, Manuela Weier, Angélica Liechti, Ayinuer Aximu-Petri, Martin Kircher, Frank W Albert, Ulrich Zeller, Philipp Khaitovich, Frank Grützner, Sven Bergmann, Rasmus Nielsen, Svante Pääbo, and Henrik Kaessmann. The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369):343–8, oct 2011.

[6] Alessandra Breschi, Sarah Djebali, Jesse Gillis, Dmitri D. Pervouchine, Alex Dobin, Carrie A. Davis, Thomas R. Gingeras, and Roderic Guigó. Gene-specific patterns of expression variation across organs and species. *Genome Biology*, 17(1):151, dec 2016.

[7] Esther T Chan, Gerald T Quon, Gordon Chua, Tomas Babak, Miles Trochesset, Ralph A Zirngibl, Jane Aubin, Michael J H Ratcliffe, Andrew Wilde, Michael Brudno, Quaid D Morris, and Timothy R Hughes. Conservation of core gene expression in vertebrate tissues. *Journal of biology*, 8(3):33, 2009.

[8] Stéphane Dray and A. B. Dufour. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 22(4):1 – 20, 2007.

[9] Renaud Gaujoux and Cathal Seoighe. CellMix: A comprehensive toolbox for gene expression deconvolution. *Bioinformatics*, 29(17):2211–2212, 2013.

[10] Yoav Gilad, Orna Mizrahi-Man, Yoav Gilad, and Orna Mizrahi-Man. A reanalysis of mouse ENCODE comparative gene expression data. *F1000Research*, 4, may 2015.

[11] Eric S. Haag. The Same but Different: Worms Reveal the Pervasiveness of Developmental System Drift. *PLoS Genetics*, 10(2):1–3, 2014.

[12] Eric S. Haag, Helen Chamberlin, Avril Coghlan, David H A Fitch, Andrew D. Peters, and Hinrich Schulenburg. Caenorhabditis evolution: if they all look alike, you aren't looking hard enough, mar 2007.

[13] Enni Harjunmaa, Kerstin Seidel, Teemu Häkkinen, Elodie Renvoisé, Ian J Corfe, Aki Kallonen, Zhao-Qun Zhang, Alistair R Evans, Marja L Mikkola, Isaac Salazar-Ciudad, Ophir D Klein, and Jukka Jernvall. Replaying evolutionary transitions from the dental fossil record. *Nature*, 512(7512):44–8, aug 2014.

[14] Jennifer W Israel, Megan L Martik, Maria Byrne, Elizabeth C Raff, Rudolf A Raff, David R McClay, and Gregory A Wray. Comparative Developmental Transcriptomics Reveals Rewiring of a Highly Conserved Gene Regulatory Network during

a Major Life History Switch in the Sea Urchin Genus Heliocidaris. *PLoS biology*, 14(3):e1002391, 2016.

[15] Jukka Jernvall and Isaac Salazar-Ciudad. The Economy of Tinkering Mammalian Teeth. In *Tinkering: The Microevolution of Development*, pages 207–224. wiley, jun 2007.

[16] Jukka Jernvall, Irma Thesleff, Y. Ahn, B. W. Sanderson, O. D. Klein, and Et Al. Tooth shape formation and tooth renewal: evolving with the same signals. *Development (Cambridge, England)*, 139(19):3487–97, 2012.

[17] Karin Kiontke, Antoine Barrière, Irina Kolotuev, Benjamin Podbilewicz, Ralf Sommer, David H A Fitch, and Marie Anne Félix. Trends, Stasis, and Drift in the Evolution of Nematode Vulva Development. *Current Biology*, 17(22):1925–1937, nov 2007.

[18] Vincent Lazzari, Cyril Charles, Paul Tafforeau, Monique Vianey-Liaud, Jean Pierre Aguilar, Jean Jacques Jaeger, Jacques Michaux, and Laurent Viriot. Mosaic convergence of rodent dentitions. *PLoS ONE*, 3(10), oct 2008.

[19] Michal Levin, Leon Anavy, Alison G Cole, Eitan Winter, Natalia Mostov, Sally Khair, Naftalie Senderovich, Ekaterina Kovalev, David H Silver, Martin Feder, Selene L Fernandez-Valverde, Nagayasu Nakanishi, David Simmons, Oleg Simakov, Tomas Larsson, Shang-Yun Liu, Ayelet Jerafi-Vider, Karina Yaniv, Joseph F Ryan, Mark Q Martindale, Jochen C Rink, Detlev Arendt, Sandie M Degnan, Bernard M Degnan, Tamar Hashimshony, and Itai Yanai. The mid-developmental transition and the evolution of animal body plans. *Nature*, 531(7596):637–41, mar 2016.

[20] Gong-Hua Lin, Kun Wang, Xiao-Gong Deng, Eviatar Nevo, Fang Zhao, Jian-Ping Su, Song-Chang Guo, Tong-Zuo Zhang, and Huabin Zhao. Transcriptome sequencing and phylogenomic resolution within Spalacidae (Rodentia). *BMC genomics*, 15:32, jan 2014.

[21] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12):550, 2014.

[22] Jason Merkin, Caitlin Russell, Ping Chen, and Christopher B Burge. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science (New York, N.Y.)*, 338(6114):1593–9, dec 2012.

[23] Mihaela Pavlicev, Günter P Wagner, T.F. Mackay, et Al., F.W. Stearns, D.J. Kliebenstein, G.P. Wagner, J. Zhang, A.E. Tsong, et Al., R.R. Ackermann, J.M. Cheverud, N.M. Young, et Al., N.M. Young, B. Hallgrimsson, R. Berg, A.R. Palmer, T.A.

Williams, L.M. Nagy, J.M. Cheverud, et Al., M. Pavlicev, et Al., L.J. Leamy, et Al., M. Pavlicev, et Al., M. Pavlicev, et Al., L.F. Delph, et Al., A. Monteiro, et Al., H.S. Kim, J.C. Fay, H.S. Kim, et Al., S. Wright, R.A. Fisher, M. Baatz, G.P. Wagner, R.E. Lenski, S.J. Schrag, et Al., C. Chevillon, et Al., S. Maisnier-Patin, D.I. Andersson, A. Bossi, B. Lehner, Z. Wang, et Al., G.P. Wagner, et Al., A.Y. Albert, et Al., J.P. Kenney-Hunt, J.M. Cheverud, D.L. Stern, V. Orgogozo, J.R. True, E.S. Haag, E.S. Haag, N.A. Johnson, A.H. Porter, G.P. Wagner, T.F. Hansen, et Al., T.F. Hansen, D. Houle, K.R. Takahasi, et Al., M.Z. Ludwig, et Al., M.Z. Ludwig, et Al., M.T. Weirauch, T.R. Hughes, B.Z. He, et Al., C.R. Landry, et Al., A.S. Kondrashov, S.A. Shabalina, R.J. Kulathinal, R.S. Singh, D. Kuo, et Al., D.S. Falconer, T.M. Williams, S.B. Carroll, O. Pellmyr, H.W. Krenn, M.A. Bell, et Al., L.V. Roth, M.D. Shapiro, et Al., M.C. Brandley, et Al., T. Kohlsdorf, et Al., C.D. Siler, R.M. Brown, T. Kohlsdorf, G.P. Wagner, Q. Zhong, et Al., M. Dreze, et Al., B. Charloteaux, et Al., R.D. Barrett, et Al., G.P. Wagner, et Al., R.E. Lenski, T. Flatt, T.J. Kawecki, K.M. Pepin, et Al., S. Li, et Al., and N. Blow. A model of developmental evolution: selection, pleiotropy and compensation. *Trends in ecology & evolution*, 27(6):316–22, jun 2012.

[24] Sabrina Renaud, Mouloud Benammi, and Jean-Jacques Jaeger. Morphological evolution of the murine rodent Paraethomys in response to climatic variations (Mio-Pleistocene of North Africa). *Paleobiology*, 25(3):369–382, 1999.

[25] Sabrina Renaud, Jacques Michaux, Jean-Jacques Jaeger, and Jean-Christophe Auffray. Fourier analysis applied to Stephanomys (Rodentia, Muridae) molars: nonprogressive evolutionary pattern in a gradual lineage. *Paleobiology*, 22(2):255–265, 1996.

[26] Sabrina Renaud, Sophie Pantalacci, Jean Pierre Quéré, Vincent Laudet, and Jean Christophe Auffray. Developmental constraints revealed by co-variation within and among molar rows in two murine rodents. *Evolution and Development*, 11(5):590–602, sep 2009.

[27] Julien Roux, Marta Rosikiewicz, and Marc Robinson-Rechavi. What to compare and how: Comparative transcriptomics for Evo-Devo. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 324(4):372–382, jun 2015.

[28] I Salazar-Ciudad and J Jernvall. A gene network model accounting for development and evolution of mammalian teeth. *Proc Natl Acad Sci U S A*, 99(12):8116–8120, 2002.

[29] Isaac Salazar-Ciudad. Tooth patterning and evolution. *Current opinion in genetics & development*, 22(6):585–92, dec 2012.

[30] Isaac Salazar-Ciudad and Jukka Jernvall. How different types of pattern formation mechanisms affect the evolution of form and development. *Evolution and Development*, 6(1):6–16, jan 2004.

[31] Isaac Salazar-Ciudad and Jukka Jernvall. A computational model of teeth and the developmental origins of morphological variation. *Nature*, 464(7288):583–6, mar 2010.

[32] John R. True and Eric S. Haag. Developmental system drift and flexibility in evolutionary trajectories, 2001.

[33] S. Urdy, N. Goudemand, and S. Pantalacci. Looking beyond the genes: The interplay between signaling pathways and mechanics in the shaping and diversification of epithelial tissues. In *Current Topics in Developmental Biology*, volume 119, pages 227–290. Academic Press Inc., 2016.

[34] L M Van Valen. Serial homology : the crests and cusps of mammalian teeth. *Acta Palaeontologica Polonica*, 38(3/4):145–158, 1994.

[35] Günter P Wagner. The developmental genetics of homology. *Nature reviews. Genetics*, 8(6):473–9, jun 2007.

[36] Karl R. Wotton, Eva Jimenez-Guri, Anton Crombach, Hilde Janssens, Anna Alcaine-Colet, Steffen Lemke, Urs Schmidt-Ott, and Johannes Jaeger. Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly megaselia abdita. *eLife*, 2015(4), jan 2015.

[37] Itai Yanai, Dan Graur, and Ron Ophir. Incongruent Expression Profiles between Human and Mouse Orthologous Genes Suggest Widespread Neutral Evolution of Transcription Control. *OMICS: A Journal of Integrative Biology*, 8(1):15–24, jan 2004.

[38] Itai Yanai, Leonid Peshkin, Paul Jorgensen, and Marc W. Kirschner. Mapping Gene Expression in Two Xenopus Species: Evolutionary Constraints and Developmental Flexibility. *Developmental Cell*, 20(4):483–496, 2011.

## Supplementary figures

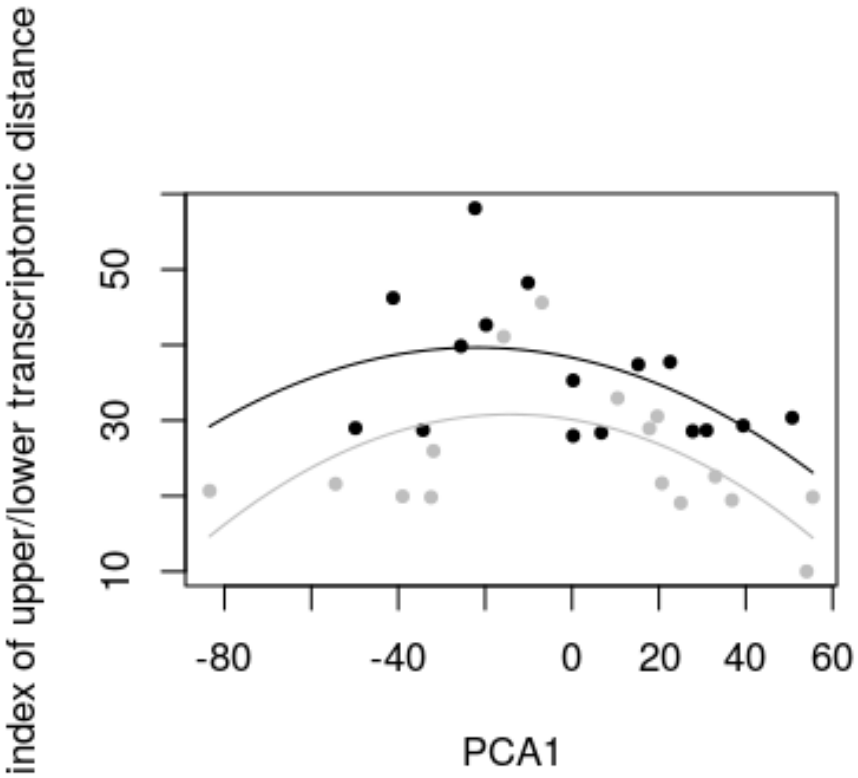



**Figure 7.** The distance between upper and lower transcriptomes peaks in early morphogenesis. Between multivariate analysis was used to extract an axis of variation associated to the difference between upper and lower tooth in the whole dataset. The distance between upper and lower samples from the same embryo on this axis is represented on the y-axis (in black for mouse embryos, in grey for hamster embryos). The x-axis is the second axis of the PCA (as in Figure 5-A) which is a developmental clock. The curves represent the best polynomial fit of this data (R2=0.51; quadratic term of the relation between y and x, p=0.00422). The index of upper/lower transcriptomic distance is significantly higher in mouse (p=0.0044) than in hamster.

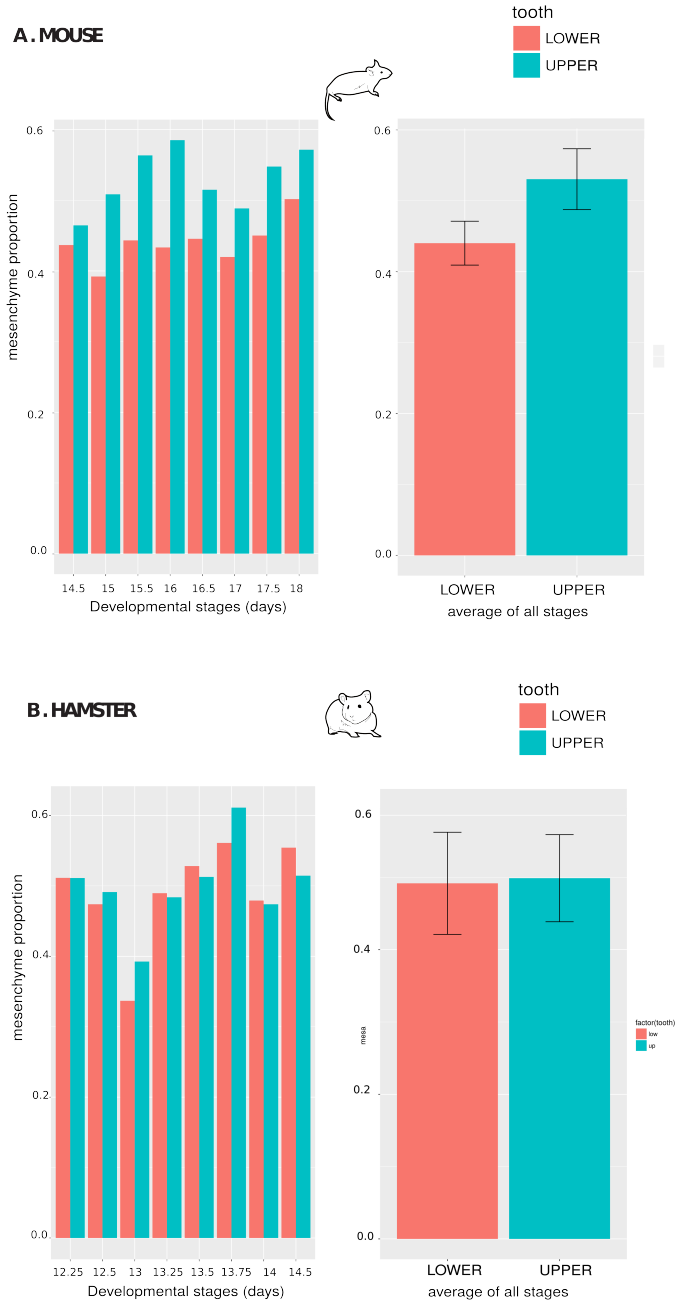



**Figure 8.** Mesenchyme proportions in mouse (A) and hamster (B) estimated from deconvolutions for 8 developmental stages. Markers for enamel knot were obtained from bite-it database (6 markers). Markers for mesenchyme and epithelium were obtained separately for each species from pure tissue transcriptomes. Left: The median proportion from 500 resampling of half of the markers is indicated for each time point. Right: the average of the time points is indicated.

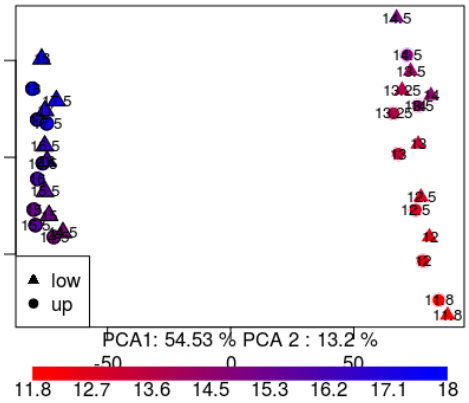



**Figure 9.** Principal component analysis with 32 samples from both species and tooth types. The location of the samples on the 2 first principal components is shown. All samples come from the same run of sequencing, and the libraries were prepared together. The mouse samples (left) and the hamster samples (right) are clearly separated. The upper/lower samples (respectively circles and triangles) are not separated on this map.

## A common upper-lower transcriptomic difference in mouse and hamster

An aspect of the comparative analysis between mouse and hamster that have not been presented in this article, are the specificities and conservation between upper and lower molars developmental program. However some analyses already enabled us to identify a shared characteristic between mouse and hamster molar development.

Multivariate analysis (*between* analysis for the upper-lower factor ) revealed a common upper-lower difference in both mouse and hamster molars (figure 9.3). Thus, in both species, there is a peak of upper-lower differences at mid-cuspidogenesis : at ED15.5-ED16 in mouse and ED13 in hamster. This period corresponds to the end of patterning of the second cusp and the beginning on the patterning of the third cusp. The intensity (amplitude or the height of the peak) and its duration (the period) are also conserved in both species. However, we noted that this peak occurred earlier ((time shifted)) in mouse molars. We also noted that, during this period there is a bigger number of genes differentially expressed between the upper and the lower molar in both species (figure 9.4). Finally, we also noted that there is a basal excess of upper-lower differences throughout crown morphogenesis in mouse molars compared to hamster, coherent with bigger morphological differences between mouse uppr and lower molars.

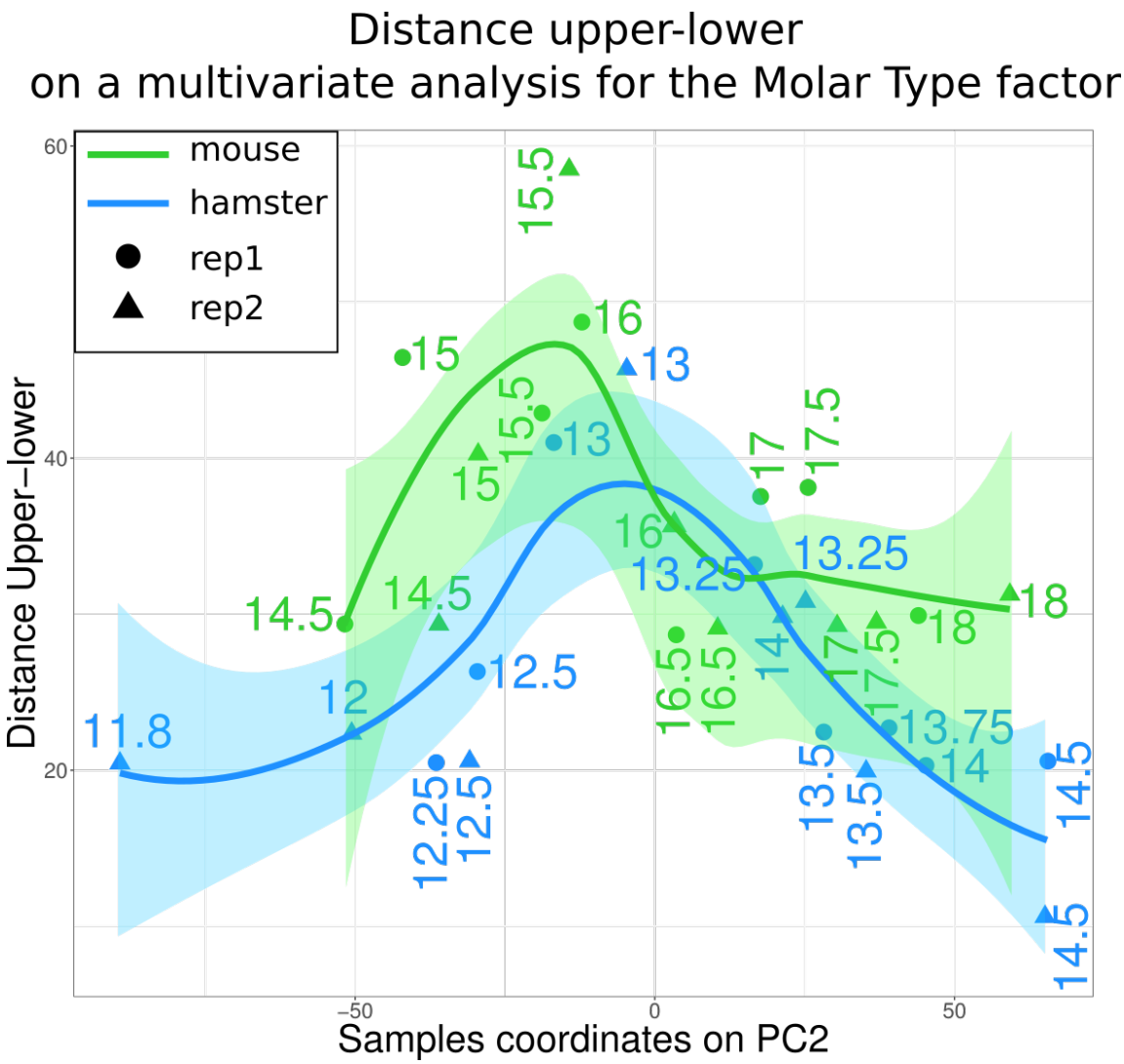



FIGURE 9.3 – **A shared peak of upper-lower differences**
Between analysis, a particular case of a multivariate analysis that enables to extract variation from one specific factor, performed on an upper-lower factor. Differences between upper and lower coordinates obtained with the between analysis is plotted on the y-axis. The x-axis are the coordinates on PC2 (of PCA on whole dataset, that is the shared temporal axis), to homologize developmental timec ourse between mouse and hamster. The curves represent the loess regression (or local regression) in each species.

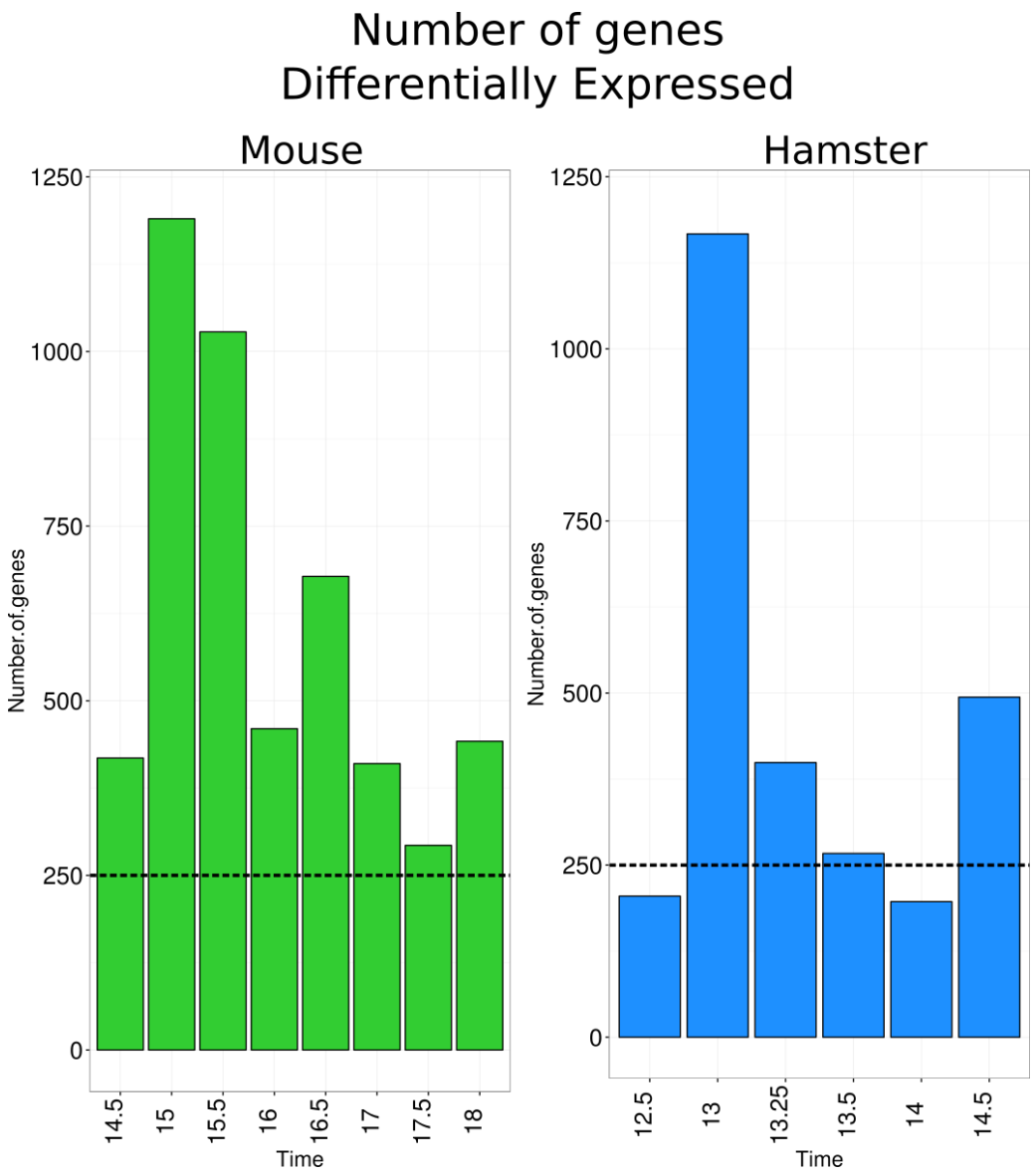



FIGURE 9.4 – **A shared increased in genes differentially expressed at *mid*-cuspidogenesis.** Number of genes differentially expressed between the upper and the lower molar of mouse (left, green) or hamster (right, blue) at a given time of development.

# 10 Characterizing bucco-lingual asymmetry during mouse molar development

Mouse first molars are very interesting because the upper and the lower molar have extreme divergent morphologies. Hence, the lower molar have 6 cusps distributed in two rows, whereas the upper molar develop two supplementary cusps in a third row lingually to the other two rows (see in the introduction the figure 3.9 *p. 57*). We would like to understand how the mouse upper molar is able to pattern two supplementary lingual cusps during its development. In this part, we focused on the buccal and lingual sides of upper and lower molars. We wondered if there is a different transcriptomic signature between buccal and lingual sides of upper and lower molar that correlates with differences in cusp patterning between both tooth types. Indeed, the literature lead us to believe that buccal and lingual side have their specificities.

**Asymmetric development of mouse teeth**

Mice only possess two kind of teeth, one ever growing incisor and three molars per jaw quadrant. The incisors and the molars both have a bucco-lingual asymmetric development.

**The mouse incisor development is highly bucco-lingual**

The incisor is intrinsically asymmetric. From stage ED14, during the cap stage, the first morphological manifestation of the bucco-lingual asymmetry is the rotation of the incisor germ toward the buccal side. Moreover the cervical loops (at the apical sides of the tooth germ) and the pool of stem cells located at the tip of the cervical loops are clearly different between the buccal and the lingual side [Harada *et al.* , 2002; Wang *et al.* , 2004a; Hu *et al.* , 2013]. Differentiation in ameloblast, cells secreting enamel, is asymmetric and only takes place in the buccal side. Thus the lingual side of the incisor only secrete dentin. The bucco-lingual asymmetry of the incisors is fairly well studied. From studies that identified genes with asymmetric bucco-lingual expression bias responsible for asymmetric enamel deposition [Kyrylkova *et al.* , 2012] to the characterization of the stem cells niche asymmetry [Harada *et al.* , 2002; Wang *et al.* , 2004a; Hu *et al.* , 2013; Balic & Thesleff, 2015].

**The molar development is also bias in the bucco-lingual axis**

Bucco-lingual asymmetry is also seen during mouse first upper and lower molar development [Cam *et al.* , 2002; Obara & Lesot, 2007; Lesot *et al.* , 2014]. However, this asymmetric development received very little attention and is poorly understood. Nevertheless, some studies highlighted the existence of a bucco-lingual bias in proteins [Fausser *et al.* , 1998], genes expression [Yoshiba *et al.* , 1998; Keränen *et al.* , 1998; Dassule *et al.* , 2000; Cho *et al.* , 2007; Zhang *et al.* , 2009; Juuri *et al.* , 2013] and processes [Obara & Lesot, 2007]. Gene-centered approaches identified some essential genes for the bucco-lingual asymmetric development in molars. For example, Zhang *et al.* [2009] found that *Osr2* is expressed with a lingual to buccal gradient in mouse molar. In mutant *Osr2* mice, there is the formation of lingual supernumerary teeth. The absence of *Shh* from the bud stage (ED13) in the mouse lower molar germ, results in the perturbation of the lingual asymmetry at stage ED14.5 [Dassule *et al.* , 2000]. Keränen *et al.* [1998] noted that "the buccal and lingual portions of the mesenchyme appear to differ in gene activity [(*eg.* *Bmp4,Lef-1* and *Msx2*)] at the time when the primary enamel knot is present", suggesting a different molecular identity of both sides.

Crown morphogenesis is the sequential patterning of cusps and it is intrinsically asymmetric. At stage ED15, the first sEK (secondary Enamel Knot), which is a signaling center responsible for cusps patterning [Jernvall *et al.* , 1994], is patterned buccally and the the second sEK is patterned lingually at stage ED15.5-ED16 (as shown in figure **??**) Cho *et al.* [2007] studied the fate of the pEK (primary Enamel Knot), the signaling center initiating crown morphogenesis. Does the cells of the pEK contribute to the sEK? They used DiI labeling, injected on pEK cells on molar germ slices, to follow the fate of these cells. They found that these cells contribute to the buccal sEK (the first patterned cusp). Based on these results and on histological data showing structural continuity between the pEK and the buccal sEK, the author proposed that the pEK maintained its original position and became the buccal sEK whereas the lingual sEK is newly form later on. Although the authors did not mention it, it seems obvious from these data that the pEK/buccal sEK in fact stay at the same position and that is the asymmetrical growth of the lingual side that enables the patterning of a new sEK (on the lingual side). The bucco-lingual asymmetry is transient in lower molar germ, whereas the asymmetry in the upper molar germs never resolves as it will pattern a second lingual row of two supplementary cusp.

**Molar evolution is characterize by the colonization of the lingual space**

In an evolutionary perspective, the buccal and lingual sides play a determinant role in molar evolution. As presented in part 3.1.3, the tribosphenic molar evolved from the patterning of cusps that explore the 3D space, exploiting the lingual space [Butler, 1990]. This exploitation of the 3D space allows the appearance of a grinding surface that in turn enables the appearance of mastication. Later in evolution, the lingual side is exploited again to pattern more cusps. For example, it happens in the Murinae lineage (the old World rat and mice), where two supplementary cusps are formed lingually, adjacent to the ancestral cusps.

**Tooth replacement in is biased toward the lingual side**

The lingual space is even exploited in animals other than mammals that don't have molars, like reptiles. Indeed, in tooth replacement, the lingual side is the side where the new tooth is patterned. From reptiles [Handrigan *et al.* , 2010; Handrigan & Richman, 2010] to mammals [Järvinen *et al.* , 2008, 2009], replacement teeth develop from the successional dental lamina, lingually to the developing tooth of the previous generation. Tucker & Fraser [2014] review the mechanisms and genetic networks of tooth replacement in various vertebrate species. I will detail here two signaling pathways that are implicated in tooth replacement. Constitutive activity of *Wnt* signaling in the oral epithelium of mouse, which does not develop replacement teeth, enable development of multiple teeth with simple conic shape, in place of the molars [Järvinen *et al.* , 2006]. Handrigan & Richman [2010] showed that *Wnt* canonical signaling is persistently active at the growing tip of the dental epithelium and promote cell proliferation in snakes. Juuri *et al.* [2013] identified that *Sox2*, expressed in the dental lamina with a lingual bias, marks the epithelial competence to generate teeth in mammals (ferret and human) and reptiles (corn snake, ball python, gecko, alligators and iguanas). These data suggest a conserved way to mark the lingual side to enable development of replacement teeth. Interestingly, Juuri *et al.* [2013] also identified a small budding, in the lingual side of mouse molar, expressing of *Sox2* at ED16, suggesting the reminiscence of a replacement tooth.

This literature suggest that the lingual side is specific and is different from the buccal side and that this is a very old property of the tooth as it is settled across all jaw vertebrates.

**What is this work about**

Together, these data suggest that the buccal and the lingual side of the tooth encompass different potentiality, suggesting a different molecular identity between these two sides enabling the teeth to differentiate the buccal from the lingual side. The novelty of the work presented below is to characterize the buccal and lingual identity of molar germs at the whole transcriptomic level. We also wondered if this buccal and lingual identity are shared between upper and lower molars. Moreover, knowing that the upper molar is more asymmetric than the lower molar the divergence of buccal and lingual transcriptomes could be exaggerated between upper and lower molar (heterometry) or the divergence could last longer in upper molar (heterochrony) or it could be a combination of the two. To answer these questions, we analyzed the transcriptomes of buccal and lingual sides of upper and lower first molars (in replicates coming from the same embryo) at stage ED15, the early breakdown of morphological symmetry of the molar germs. This analysis revealed that the buccal and the lingual sides have different transcriptomic identities. We also found that the bucco-lingual transcriptomic differences are bigger in upper molar compare to lower molar, in agreement with more pronounced bucco-lingual asymmetry in the adult morphology. This characterization of the developmental programs of buccal and lingual sides of mouse first upper and lower molar germs would help us to better understand how the mouse first upper molar is able to pattern two supplementary lingual cusps.

**Participation to this work**

The work presented in this section corresponds to one of the main projects during my thesis, together with chapter 8. I handled all the transcriptomic analysis (as detailed below), prepared the corresponding figures and drafted the first version of the manuscript. The text was subsequently corrected by my supervisors, and completed by the work of our collaborators. The writing is still in progress, notably to fully updated and integrate the new results provided by our collaborators in Prague.

I quantified abundance of transcripts from the RNA-seq data. I completed all the downstream analyses with *R* [R Core Team, 2016]. I carried out multivariate analyses of the subsequent data. More precisely I performed principal component analysis, together with *between* analyses,a particular case of a principal component analysis to extract variation from one specific factor. I used DESeq2 [Love *et al.* , 2014], and its GLM models, to identify genes that are differentially expressed in different pairwise comparisons of buccal/lingual sides and upper/lower molars. I also performed GO analysis, using the topGO package [Alexa & Rahnenfuhrer, 2016] and used a visualization tool that I designed, to visualize the results of my GO analysis (presented in chapter A). I coupled it with a pairwise comparison of GOTerms (based on $\chi^2$ test) to find if one GOTerm is more enriched in one side towards the other, or for a given side in one tooth type toward the other. I performed partial deconvolution analyses with the CellMix package [Gaujoux & Seoighe, 2013; Gaujoux, 2013], to estimate proportion of epithelium, mesenchyme and cusp tissue in each samples.

-DRAFT-

# Combined developmental and transcriptomic approaches reveals extensive bucco-lingual asymmetry in mouse molar development

December 18, 2016

Petit C.; Ledevin* R.; Lochovska* K.; Hovorakova M., Sémon M. and Pantalacci S.
*in prep.*

**Abstract**

The molar is an iconic innovation of mammals, and its development and evolution have been studied since the XIX century. Molar origination implied an enlargement of the tooth along the bucco-lingual direction, with buccal and lingual cusp rows, instead of a single longitudinal row. In some mammal lineages, including mice, the upper molar evolved a third cusp row. Despite important advances in tooth developmental and evolutionary developmental biology, we know relatively little about the bucco-lingual development of molars and its regulation. Here we combine a genetic, morphometric and transcriptomic approach to gain insights into this aspect. First, we performed lineage tracing to unambiguously demonstrate that the primary enamel knot, initially found in the center of the developing tooth germ, is one day later found buccally, suggesting early asymmetric morphogenesis of the germ. A morphometric analysis of 3D reconstructed tooth germs shortly after cap transition showed that bucco-lingual differences in morphogenesis were primarily found in the inner part of the epithelial component, as opposed to the external part, explaining how the enamel knot can be shifted on the buccal side while the overall growth of the tooth germ is only poorly asymmetric This analysis also emphasized differences between the lower and the upper molar, likely linked to final difference in number of cusp rows. Finally, we cut lower and upper molar germ in buccal and lingual halves and sequenced their transcriptomes. This revealed exceptionally marked differences between the buccal and lingual parts, largely shared by the two molars. With the help of tissue-specific transcriptomic data (epithelium, mesenchyme, enamel knot), we could interpret transcriptomic differences between the two halves and the two molars in terms of abundance of different cell populations. This highlighted the high complexity and regionalization of the epithelium, as well as unbalanced neural and endothelial cell populations in the mesenchyme. We also identified Wnt, Shh, BMP and RA signaling as likely key players of asymmetric development, as well as a small number of transcription factors including *Osr2* and *Sox2*. Finally, we focused on lower-upper molar differences. Bucco-lingual transcriptional differences were stronger for the upper molar, partly due to a more buccal positioning of the enamel knot in upper molar and a stronger deficit in epithelial tissue on the lingual side. We also identified candidate genes for these lower-upper molar differences. Taken together, our results provide extensive resources for dental studies and reveal unsuspected high level of bucco-lingual asymmetry, including in the lower molar. They open interesting hypothesis on the origins of molars and the 3 rowed upper molar of mice.

## Introduction

The mammalian molar has a long history of study under both evolutionary and developmental perspectives. Because it fossilizes extremely well and display complex characters, it has served taxonomy studies of mammal evolution, and the evolutionary history of molars in mammals has been extremely well documented. Comparative studies of molar development date back to the $XIX^{th}$ century [1], and more recently, work in molar has been pioneer in revealing how the developmental properties of an organ influence its evolution [2, 3]. Finally, molar and more generally tooth development have been studied from a developmental genetics point of view, as an interesting model to understand the formation of complex shape, human craniofacial genetic defects or more recently stem cells [4, 5, 6, 7]. Despite these impressive advances, one aspect of molar morphogenesis has remained elusive: the development of molars (more generally molariform teeth, *ie* molars and premolars) in the bucco-lingual dimension. This is surprising given that this aspect seems highly relevant for our understanding of molar development and

evolution, as we explain below. Molars show a large variety of shapes in mammals, but most often show individual main cusps, arranged in two rows: one buccal and one lingual. This arrangement is a specificity of the mammalian molar, first seen in the tribosphenic molar of the earliest mammals. Earlier in evolution, in reptilian ancestors and mammaliform species, the primitive condition was respectively conic teeth and multicusps teeth arranged in a single row. Thus the transition to tribosphenic molar, supposed to have participated to the evolutionary success of mammals [8], implied adding cusps on the lingual part of the tooth. Later, in some mammalian lineages, even more cusps were added on the lingual side, especially in the upper molar. This is seen in mice and rats, but also in other rodent lineages [9], in human with the carabelli cusp [10, 11]. It should be noted that a bucco-lingual asymmetry of the tooth and its development largely precedes the tribosphenic molar. The base of conic or single-row multicusps teeth in mammal ancestors were most often enlarged on their lingual side. Moreover, in fish and reptiles, whose teeth are continuously replaced, replacement proceeds from the lingual side of the tooth developing tissue (enamel organ). This is also the case in mammals, although teeth are at best replaced once. Of note, authors from the first part of the $XX^{th}$ century have pointed a possible link between limited tooth replacement and gain of molar complexity in mammals, up to proposing that molars are made from the fusion of 2 generation of teeth (*e.g.* L.Bolk, discussed by P.M. Butler in [1]). Without any need for invoking such an extreme scenario, the circuitry responsible for the asymmetry of the enamel organ in relation with tooth replacement could have been co-opted to make teeth further developing in the lingual direction, *i.e.* molars. However, at the moment, we know very little about the bucco-lingual development of mammalian molars, and the molecular determinants behind. Below, we present basics of molar development, and then focus on the relatively poorly known aspects of cusp formation.

The best studied model of molar development is the lower first molar of mouse. Tooth development relies on reciprocal epithelium-mesenchyme interactions. These drive the budding of the oral epithelium into the underlying condensing mesenchyme, and the formation of a signaling center called the primary enamel knot (PEK, ED14.0). The later is made of non-proliferating cells, packed in a very remarkable manner, and is though to induce epithelial downgrowth at a distance [12, 13]. Formation of the PEK and cap transition is a key step, many mutants are failing at this stage. Most recent research as shown the necessity of tightly controlling BMP and Wnt pathway for successful transition [14, 15]. From cap formation, cells of the enamel organ (epithelial part of the germ) progressively specialize, with on top of the PEK: the inner enamel epithelium (IEE) in contact with the capped ectomesenchyme, the outer enamel epithelium (OEE), and internally, the stellate reticulum made of star shaped cells spaced in a glycosaminoglycan-rich matrix (figure 1). As a consequence, the so-called cervical loops form and start capping the condensed mesenchyme (cap stage, from ED14.5), what will progressively outline the crown. Growth of these cervical loops is notably supported by FGF signals, coming from the PEK and the mesenchyme [16]. Moreover, it was recently shown that a BMP-Shh networks controls the growth and maintenance of cervical loops where resides a population of *Sox2* positive epithelial stem cells [17]. The relief of the crown results from further growth and folding of the interface between the internal epithelium of the cap and the mesenchyme. Secondary enamel knots, sitting at the future cusp tips, are thought to control this, in a manner similar to the PEK. These SEK form sequentially by activation-inhibition mechanisms as the tooth germ grows [18]. Finally, cells at the internal epithelium-mesenchyme interface will differentiate respectively in enamel and dentin-secreting cells. While many aspects of molar development have been well studied (*e.g.* tooth initiation, cell differentiation...), the bucco-lingual dimension of this whole process, is poorly known. And yet, asymmetric development of the molar had been repeatedly questioned in early $XX^{th}$ century studies, in relation with molar evolution. Studies in a number of mammals had consistently found that the first cusp to be formed was buccal (reviewed in [1]), and the histology of the tooth had been examined in great details, emphasizing the presence not only of the enamel knot but of a buccal structure called the enamel septum or enamel navel, bridging the enamel knot and the buccal cervical loop (figure 1). Bolk even saw in this structure a proof for its theory that molars are fused conical teeth (this would be the site of fusion), a theory largely denied since then (Bolk, 1926). Wakeford [19] proposed that this cells would be the last to transform into stellate reticulum cells, that fill the middle of the enamel organ (epithelial part of the germ). Latterly, Cho *et al.* [20] and more recently our lab [21] examined the dynamics of SEK formation with molecular markers, and showed that for both lower and upper molar, the first cusp patterned is a buccal cusp (*e.g.* paracone in upper molar and protoconid in lower molar), and the second is a lingual cusp (respectively, protocone and metaconid). However, when the cap forms, the PEK occupies a central position in the cap. This suggests that the PEK gives rise to the buccal SEK and lingual growth enables the formation of the lingual cusp. Alternatively, two SEK may form independently in the cap [22]. Cho *et al.* [20] have marked the tip of the molar germ bud by injecting DiI in tooth slice culture at ED13.5 (1 day before PEK and cap formation). Two days later, DiI-labeled cells were found on the buccal side of the germ, thus presumably at the tip of the first forming cusp (*e.g.* paracone in upper molar and protoconid in lower molar, figure 1). These data are thus in favor of the first hypothesis. An asymmetry of buccal and lingual cervical loop structure was also noticed by Fausser *et al.* [23]. They reported an asymmetric distribution of components of adherens junction, desmosomes and hemidesmosomes, and suggested that the lingual cervical loop may be less cohesive than the buccal one. The molecular processes and determinants behind asymmetric

morphogenesis of the mouse molar remain largely elusive. It is known that a number of genes show a bucco-lingual bias during early molar row development (*e.g. Osr2* [24]) or just before or during cap formation (*e.g. Bmp4*, *Msx2*; [20, 25]). A few mutants are known that affect bucco-lingual development (*e.g. Osr2* [24] or genes in the Shh pathway [26, 27] ). On top of that, the study of reptiles and mammals replacing their teeth (which is not the case for mouse) has revealed some molecular basis of tooth replacement, which proceeds from lingual side (reviewed in [28]). In the epithelium, *Sox2* marks the tooth forming region [29] whereas high Wnt/Lef1 marks the successional lamina [30].

In summary, we lack a clear description of molar asymmetric development, both morphologically and molecularly. Here we examined how BL asymmetry arises morphologically during early cap formation with a combination of morphological and transcriptomic data. Although lower molar is generally the model of choice for tooth studies, here we decided to compare lower and upper molar for two reasons. First, the above mentioned data suggested exaggerated lingual development of the lingual upper molar: this exaggeration could help us to identify relevant processes and genes. Second, identifying differences between the two molars may help us understanding an evolutionary innovation: the formation of additional lingual cusps in mouse upper molars. First, we traced genetically cells of the PEK, originally found at the tip of the molar bud, showing unambiguously that these cells are left on the buccal side of the tooth germ during cap formation. Second, we computed and mapped tissue shape changes on 3D-reconstructed lower and upper tooth germs from ED14.5 to ED15.5. This analysis suggested that the expansion of the inner enamel epithelium surface associated with its pressing in takes an important part in the buccal displacement of the PEK and lingual extension of the molar germ. Moreover, this phenomenon was especially marked in the upper molar germ, but was then followed by rapid growth of the cervical loops between 15.0 and 15.5. Third, we performed RNA-seq on buccal and lingual halves of lower and upper molar germs. Bucco-lingual differences were marked and largely shared by the two molars. In the line with a previous paper [21] and with help of tissue specific transcriptomic data (epithelium, mesenchyme, enamel knot), we could interpret transcriptomic differences between the two halves and the two molars in terms of abundance of different cell populations. We computed data from the literature mentioning bucco-lingual bias in expression or bucco-lingual defects from tooth initiation, and identified in our data Wnt, Shh, BMP and RA signaling as key players of asymmetric development, as well as a small number of transcription factors including *Osr2* and *Sox2*. Finally, we focused on lower-upper molar differences. BL transcriptional differences were stronger for the upper molar, and partly due to more buccal positioning of the enamel knot in upper molar and exaggerated depletion in epithelial tissue on the lingual side. We also identified genes possibly behind this differences. Our results provide new light on bucco-lingual development and strong resources for further studies.

# Material & Methods

## Mice breeding and embryo sampling

*CD1 (ICR) adult mice were purchased from Charles River (France). Females were mated overnight (night at 20:00 pm) and the noon after morning detection of a vaginal plug was indicated as embryonic day (ED) 0.5. Other breeding pairs were kept in a light-dark reversed cycle (night at 12:00 pm), so that the next day at 16:00 pm was considered as ED1.0. Pregnant females were killed by cervical dislocation and embryos were harvested on cooled Hank's or DMEM advanced medium, weighted as described previously and immediately decapitated. This study was performed in strict accordance with the European guidelines 2010/63/UE and was approved by the Animal Experimentation Ethics Committee CECCAPP (Lyon, France; ENS_2012_045).*

## Epithelium dissociations and in situ hybridizations

*Complete or hemi mandibles and maxillae were dissected in Hank's medium and treated with Dispase (Roche) 100% at 37°C for 1h30 to 2h20 depending on embryonic stage. Epithelium was carefully removed and fixed overnight in PFA 4% at 4°C. DIG RNA antisense Fgf4 probe were prepared from plasmids described elsewhere [12]. In situ hybridization was done according to a standard protocol. Photographs were taken on a Leica M205FA stereomicroscope with a Leica DFC450 digital camera (Wetzlar, Germany) or on a Zeiss LUMAR stereomicroscope with a CCD CoolSNAP camera (PLATIM, Lyon).*

## Immunolocalisation and 3D reconstructions

*Tooth germs dissected from litter-mate embryos of RNA-seq samples (same weight) were fixed overnight in PFA4% and dehydrated through a methanol series. In toto immunolocalisation protocol was adapted from [31]. Following incubation in H202 5%, DMSO 10% in methanol for 4 hours, they were rehydrated, blocked and incubated successively with a pan-cytokeratin antibody (overnight,1/200, Novus Biologicals) and a Dylight 549 conjugated Donkey Anti-rabbit antibody (overnight 1/200, Jackson immunoresearch) and finally with Hoechst (overnight , 50µg/ml). Following dehydration, they were clarified and mounted in BABB as described in [32]. They were imaged with a Zeiss LSM710 confocal microscope at the PLATIM (Lyon, France). The outline of the epithelium labeled by the pan-cytokeratin antibody and the outline of the tooth germ labeled with hoechst were delineated manually in the AMIRA software.*

## Visualization of deformations between developmental stages

*Using Amira, the surface of the first molar was segmented by one observer at three distinct developmental stages (days*

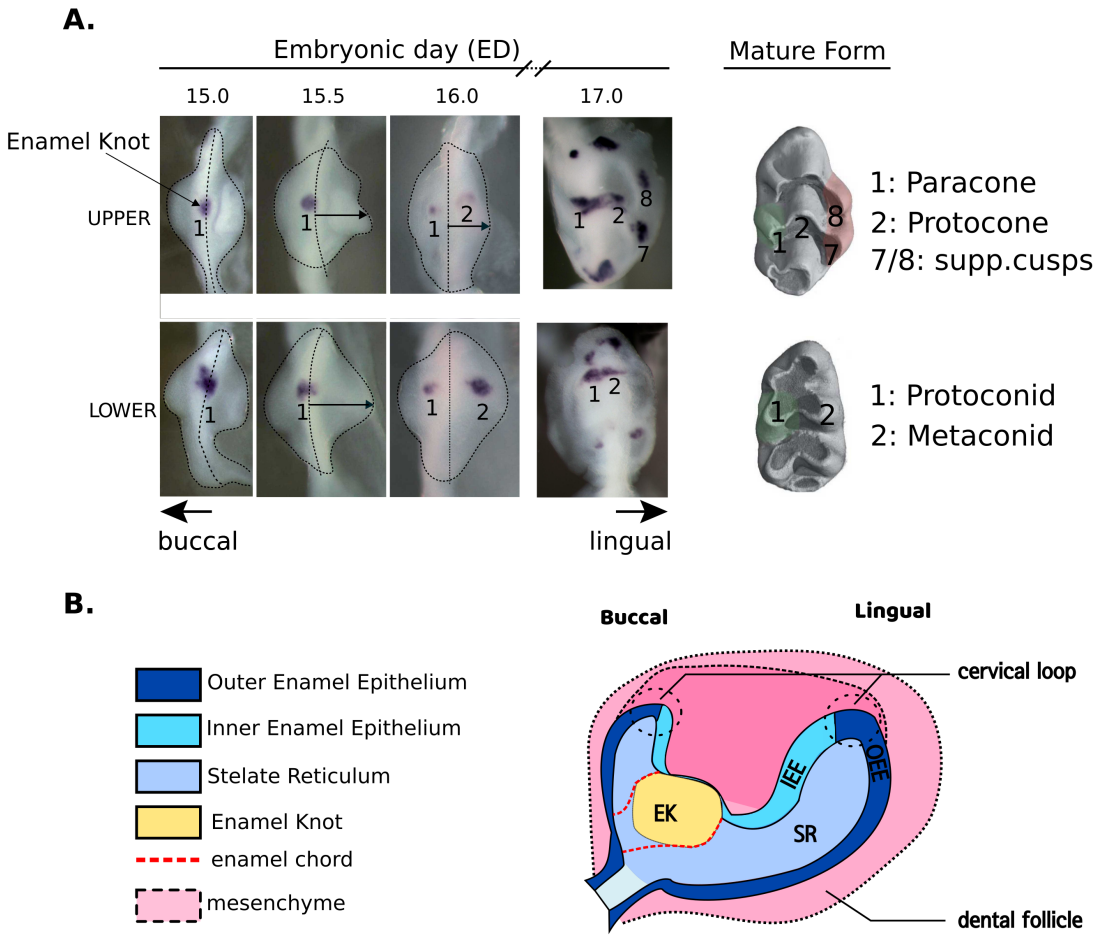



Figure 1 – **Morphology of mouse first upper and lower molar germs. A.** Top view. *In situ* hybridization of dissociated epithelia of mouse molar germs for *Fgf4*, a marker of secondary enamel knot (SEK). From ED15 to ED16, patterning of the first two cusps. Note that the EK is central at stage ED15. SEK are numbered according to their order of appearance. The supplementary cusps of the upper molar are the last two cusps to form (7 and 8). Mature form of upper and lower molars with the nomenclature of the first two cusp to form. Anterior toward the top and posterior toward the bottom. **B.** Transverse view. Schematic representation of the transverse section of an upper molar germ at stage ED15.5, showing the cellular regionalization of the molar germ. The section is located where the lingual cervical loop is the biggest.

*14.5, 15 and 15.5) for both maxillary and mandibular teeth. For each surface, 75 (lower) and 110 (upper) fixed landmarks were digitized using the landmark editor software (version 3.0). In order to visualize deformations between developmental stages, we applied a surface registration based on statistical shape models (SSM). The fixed landmarks were hence only used as priors to guide the registration process. Each mesh was cleaned, smoothed and decimated to 60000 faces using the 'Rvcg' package [33, 34]. Maxillary and mandibular templates were designed and used to compute a Gaussian process model, to control for allowed deformations in the elastic surface registration process. This is done by defining a synthetic covariance functions based on a Gaussian kernel. Additionally, isotropic variation was added to model size differences (this process made use of the functions statismoModelFromRepresenter and createBayes from the 'RvtkStatismo' package [35]). Finally, the resulting Gaussian process model was used to constrain the deformations obtained from an elastic ICP registration based on smoothed displacement fields (gaussmatch function from 'mesheR' package [36]). For both maxillary and mandibular teeth, the resulting meshes were finally rigidly aligned in space without scaling in order to visualize differences between developmental stages. Colormaps were used to visualize the distances between aligned meshes. To make the lower and upper visualizations comparable, the same scale of color was used in both cases. Animations were performed using the warpmovieMulti function from the mesheR package [36].*

### RNA-seq data analysis

*RNA-seq raw data: the 4 tooth germs for the first molar had been dissected from a unique mouse E15.0 embryo. Replicates were obtained by comparing right and left side of this same embryo. Tooth germs were cut in two halves to isolate buccal and lingual side, and total RNAs were extracted (fig. 1). IlluminaR TruSeqTM RNA v2 (PolyA+) libraries were prepared from 50-70 ng total RNAs, where an equal amount of AmbionR ERCC RNA Spike-In Mix1 had been added according to the AmbionR protocol (e.g. 1$\mu$L og a 1:1000 dilution for each tube). Sequencing was performed by the IGBMC Microarray and Sequencing platform, a member of the France Genomique program. 8 single-end libraries (1x 50bp reads) were obtained (table 1). The quality of the libraries was assessed with FastQC v0.10.1 (*`http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`*).*

*Expression levels were obtained by pseudo alignments using Kallisto [37] (version 0.43.0) and the mouse reference genome (GRCm38, release ensembl 85). Statistical analysis was performed using R (version 3.3.1, 2016-06-21), with ggplot2 for graphics (2.1.0 [38]). Differential expression analysis was performed with the DESeq2 package [39] (version 1.12.4 ) with following settings for "mult" analysis: ∼tooth + side; "interaction" analysis: ∼ tooth + side + tooth:side; and other settings are precised in the main text. Multivariate analyses (mainly PCA) were performed using ade4 packages [40] (version 1.7.4).*

### Deconvolution: estimating tissue proportions from RNA-seq data.

*Performing gene expression deconvolution, is estimating cell type proportions (and/or cell-specific gene expression signatures) from global expression data in heterogeneous samples. Here, we used the R package CellMix [41] (version 1.6.2), to estimate the proportions of the three tissue compartments in our tooth germ transcriptomes. CellMix implements in particular the method Digital Sorting Algorithm (DSA) proposed by Zhong et al. [42] which performs complete gene expression deconvolution using a set of marker genes only. We needed therefore to define 3 sets of markers, as specific as possible for each tissue compartment. We used markers obtained from microarray data [43] and RNA-seq data of pure tissue at ED15. The enamel knot: 16 genes, a pool from bite it database and microarray expression [43] such as ek > 1.7 epi and from RNA-seq data Epi > 2 Mes, the epithelium: 121 genes, ek< 1.1 epi from microarray data of pure tissues and Epi> 2 Mes from RNA-seq expression data in pure tissue and highest quartile [95%] of ratio between epithelium and mesenchyme from pure tissue expression) and the mesenchyme: 191 genes ek< 1.1 epi from microarray data of pure tissues and Mes > 2 Epi from RNA-seq expression data in pure tissue and highest quartile [95%] of ratio between epithelium and mesenchyme RNA-seq expression in pure tissue.*

## Results and Discussion

### Unbalanced bucco-lingual morphogenesis of the molar germ following cap transition

#### Increased growth of the lingual side of the molar germ during cap formation

In an effort to characterize cusp patterning in mouse [21], we noticed that the first SEK is clearly located on the buccal side of the lower and upper molar germs at ED15.0-15.5 (figure 1, *in situ* hybridizations with *Fgf4* SEK marker). If we assume a cellular continuity between PEK and the first buccal SEK, as was suggested by DiI labellings on molar slices cultures [20], it would suggest that the lingual side of the tooth germ grows more rapidly than the buccal side in both molars, while the PEK/first SEK remains on the buccal side. To check this assumption in an unambiguous manner, we used Shh-CreERT mice, that enables time restricted activation of the CRE recombinase in cells expressing Shh, following tamoxifen administration [44]. The recombinase induces irreversible activation of the lacZ gene, so that the progeny of cells expressing Shh at the time when tamoxifen is applied (and shortly after) can be revealed by X-gal staining. Since *Shh* is expressed in the PEK from ED14.0 [45] and later in SEK, we reasoned that tamoxifen application at ED13.5 or ED14.5 should mark the progeny of the PEK and possibly first SEK, enabling us to follow the latter during following developmental stages. The results (figure 2 and supplementary data confirmed that the PEK progeny sits on the buccal side of the ED15.5 molar germ, where the first SEK corresponding to the paracone and protoconid is also found (figure 1). This demonstrate that the inner enamel epithelium has formed or grown unequally on buccal and lingual side of the tooth germ following cap transition.

#### Reshaping of the inner enamel epithelium is a hallmark of cap formation and is stronger on the lingual side.

To better understand how this would be achieved, we generated 3D reconstructions of the enamel organ of lower and upper molar germs at ED14.5, ED15.0 and ED15.5 and modeled its deformation between successive stages. Then we mapped the extent of deformation between consecutive stages on the initial stage (*e.g.* ED14.5 $\longrightarrow$ ED15.0 on ED14.5 stage), with green color corresponding to outer expansion and red color to pressing in of the surface (white: no change; figure 3). Overall, this analysis showed that changes in tissue shape are highly regionalized: numerous color patches are seen, and different behaviors are seen both along antero-posterior and bucco-lingual axis. Between ED14.5 and ED15.0, deformations are consistent between the lower and upper molar germ (figure 3, a,a',a" and b, b',b"). While the outer enamel epithelium (OEE) expands towards the outside of the germ (green, see figure 3a,b and a",b"), the inner enamel epithelium (IEE) expands by being pressed in towards the

| | lower m1 | | | | upper M1 | | | |
|---|---|---|---|---|---|---|---|---|
| | left | | right | | left | | rifght | |
| | Buccal | Lingual | Buccal | Lingual | Buccal | Lingual | Buccal | Lingual |
| sample name | low1 B | low1 L | low2 B | low2 L | up1 B | up1 L | up2 B | up2 L |
| library size | 37.4 M | 32.8 M | 33.9 M | 39.8 M | 64.1 M | 44.2 M | 80.2 M | 56.4 M |
| Total Map | 27.8 M | 24 .7 M | 25.4 M | 30.2 M | 48.2 M | 33.8 M | 60.8 M | 43.2 M |
| % Total Map | 74% | 75% | 75% | 76% | 75% | 76% | 76% | 76% |
| Spike-In reads | 5500 | 12000 | 7300 | 9200 | 11800 | 7500 | 15900 | 14600 |
| % Spike-In reads | 0.02% | 0.05% | 0.03% | 0.03% | 0.02% | 0.02% | 0.03% | 0.03% |

Table 1 – Reads Statistics. (M= million reads)

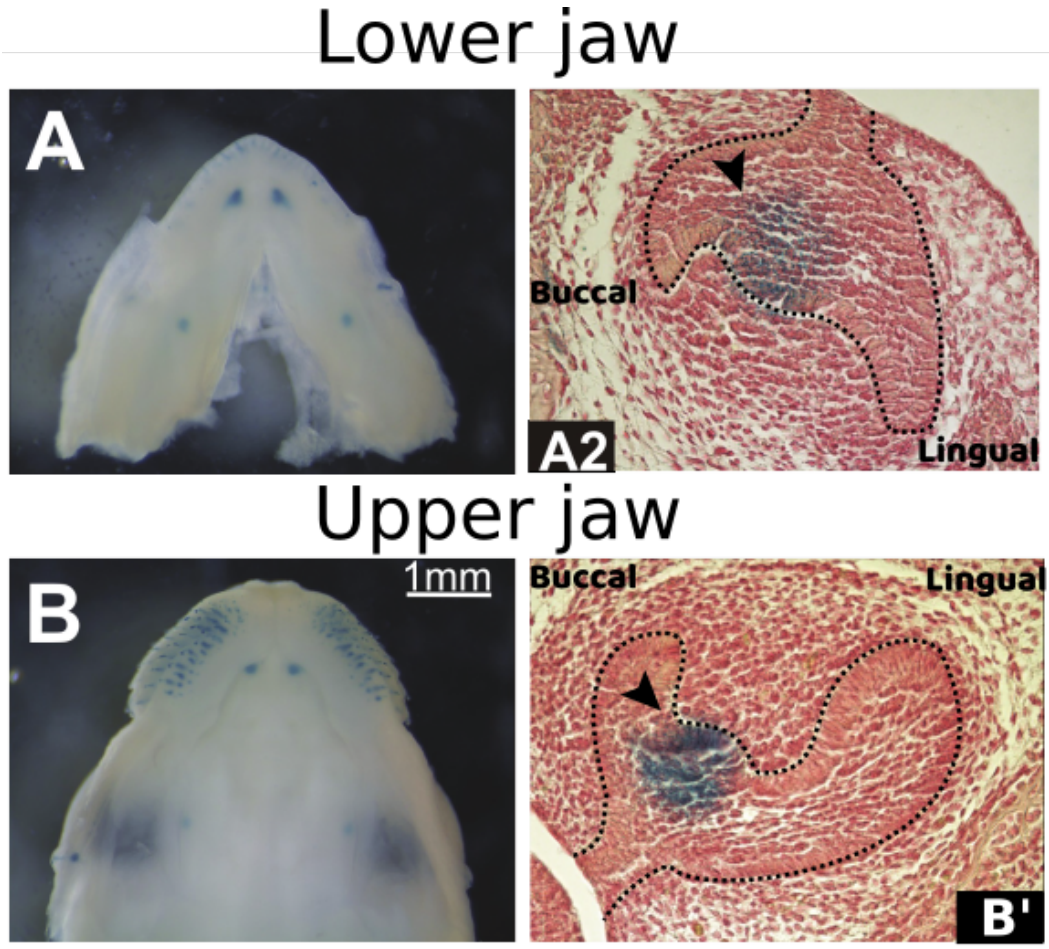



Figure 2 – **PEK tracing in the lower and upper first molar reveals unequal bucco-lingual growth of the inner enamel epithelium in the period following cap transition.** Pregnant *Shh*-CRE-ERT; Rosa mice were injected at ED13.5 and embryos were harvested 48h later, at ED15.5. Mandibles (**A**) and Maxilla (**B**) are shown, together with corresponding frontal sections. The PEK progeny clearly sits on the buccal side of the ED15.5 tooth germ (arrow). Similar results are obtained with ED14.5 injections and harvesting at ED15.5 (slightly larger X-gal positive domain, sup data(not yet provided)).The epithelium is is delimited by a dashed line.

base of the germ (red, see figure 3 a' and b'). Since cap transition and cervical loop formation is extremely fast at ED14.0-ED15.0, it seems reasonable that it involves mainly cell re-organization rather than cell proliferation. During this process, there was a clear asymmetry between buccal and lingual parts in the inner enamel epithelium, with pale color patches dominating the buccal part and intense color patches dominating the lingual part (this

concerns the central part of the tooth, where buccal and lingual cervical loops form; the formation of anterior and posterior cervical loops is delayed and moreover not timed in the two molars and will therefore be ignored here). The sharper tissue shape changes of the lingual part is also obvious in the morphing movie (*supplementary data*). However, the asymmetry is only slightly seen in the expansion of the outer enamel epithelium (compare figure 3a and 3a" or 3b and 3b": more light green or white versus more intense green, respectively). This is obvious only for upper molar. In contrast, the pressing in of the inner enamel epithelium is much stronger on the lingual side than on the buccal side (intense red versus light red or white, compare left and right in figure 3a' and 3b'). This suggests that the expansion of the inner enamel epithelium surface associated with this pressing in takes an important part in the asymmetrization of the tooth germ. This also explains how the buccal displacement of the PEK can be achieved despite limited asymmetry in cervical loop growth.

**Upper molar IEE expands at ED15.0, especially on the lingual side.**

In the next stage (ED15.0 to ED15.5), there was a marked difference between lower and upper molar germ. In the lower molar germ, the process mentioned before intensifies (figure 3d,d',d"), especially the pressing in of the IEE (except posteriorly where posterior cervical loop is forming). This peaks just posteriorly where the two first cusps are forming (stars indicate SEK position or future position). As a consequence, cervical loops have become thinner and the two first cusps are visible at ED15.5. Thus, in the lower molar, the morphological changes of the IEE suggest that cusp formation has already largely started in lower molar. In contrast, pressing in of upper germ IEE remained highly localized to the region where the first cusp is forming and a band posterior to it, while IEE mainly expanded out elsewhere. The asymmetry between buccal and lingual part was pronounced, with strong expansion on the lingual side leading in the next stage (ED15.5) to an especially large and thick lingual cervical loop. In light of our previous work showing early delay in cusp formation in the upper molar [21], these data suggest that cusp formation is limited to the first cusp (paracone) in the upper molar, and possibly inhibited in the rest of the germ.

## Early ED15.0 molar germs show marked transcriptional bucco-lingual regionalization

We designed a RNA-seq study to highlight transcriptional differences in the buccal and lingual part of the developing molar germ, for both lower and upper first molar. In the hope to highlight both the determinants of asymmetric development and the genes responsible for its realization, we chose an early ED15.0 stage (352-359 mg), when the asymmetric development of the germ has just started. With those transcriptomes, we not only aim to recover strongly biased genes that could be the determinants of buccal versus lingual side. Indeed, a previous study [21] has shown us that more global analysis of transcriptomes, taking into account small to moderate fold changes as well, can reveal differences in abundance of either cell types or cell behavior (*e.g.* mitotic versus resting for the same cell type).

To avoid the confounding effects of slight differences in developmental age (see Petit *et al.*, *in prep.* [46]), all bucco-lingual samples were extracted from a single CD1 mouse embryo, left and right germs providing replicates. Molar germs (lower and upper) were cut in a buccal and lingual halves, and total RNA was extracted to prepare RNA-seq libraries (figure 4A). Gene expression levels were generally well correlated between replicates. Moreover, when buccal and lingual libraries were pooled together and expression levels were obtained from these artificially reconstructed whole tooth germs , the resulting expression levels were well correlated with those obtained from ED15.0 whole tooth germs in a previous study [21]. This indicated that although our libraries were prepared from relatively small amount of RNAs, gene expression estimates are reliable. Global analysis of transcriptomes is strongly helped by tissue-specific datasets [21]. Therefore, we also generated ED15.0 transcriptomes for the epithelial and mesenchymal part of the tooth germ separately, after enzymatic separation of the two components (for both lower and upper molars, two replicates each). For Enamel knot tissue, we relied on a published dataset by O'Connell *et al.* [43].

**Marked bucco-lingual transcriptomic differences are shared by lower and upper first molars**

In order to get a global view of transcriptomic differences between our samples, we first performed a Principal Component Analysis (PCA) of our dataset (15,254 expressed genes , *i.e.* genes with basemean >= 10 in at least one condition). This analysis produces sample map whose axes represent the main components of the variation in the dataset. The first axis (PC1, 60%) separated buccal from lingual samples, and the second axis separated lower from upper samples (PC2, 18%) (figure 4B). Variation associated to a given axis of a PCA is not always associated to a single biological factor. The part of the variance explained by bucco-lingual and upper-lower factors were thus quantified by performing a "between analysis" (bca, package ade4 [40]). The bucco-lingual factor accounted for a total of 57.8% of total variation between transcriptomes, and the lower-upper factor for 17.8%, which is highly consistent with the percentage obtained from PCA axis. Thus a very large part of the variation in gene expression between our samples (almost 80%) is simply explained by taking into account the B/L identity and the jaw origin (upper or lower). Against, this argues for the reliability of our dataset, where other sources of variation (technical variability, developmental age,...) were kept minimal. Moreover, the fact that the variation attributed to the buccal/lingual origin (PC1, 60% - between analysis, 57.8%) is three times the variation attributed to the jaw origin

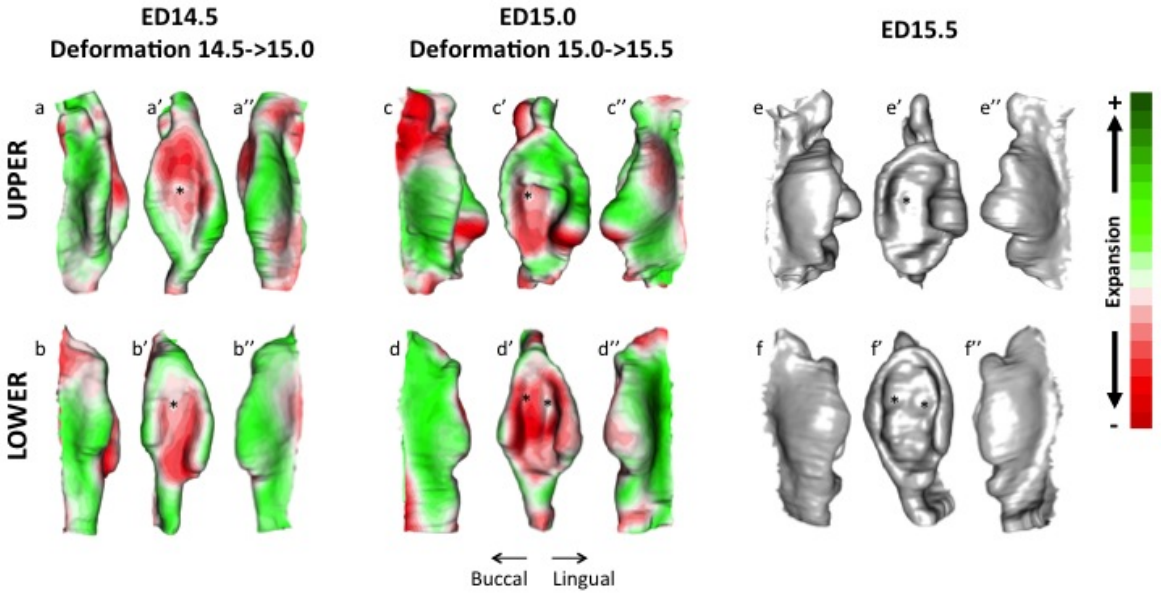



Figure 3 – **Quantifying enamel organ deformation reveals that asymmetric bucco lingual development is supported by unequal pressing in of inner enamel epithelium and highlights sharp difference between lower and upper molar at ED15.0-ED15.5** 3D reconstruction of the enamel organ (epithelial part of the tooth germ) shown for 3 consecutive stages (ED14.5: a and b; 15.0 : c and d; 15.5: e and f) in three views (x: buccal side; x': internal side; x": lingual side). The deformations between consecutive stages are shown with a color code mapped onto the starting stage (ie. ED14.5 for ED14.5 to 15.0 transition and ED15.0 for ED15.0 to 15.5): green for expansion, and red for compression. The morphing of the three consecutive stages is also available as a supplementary movie (not yet provided). Stars indicate position of the PEK or SEK (present or future). The physical and deformation scales were the same for all the reconstructed tooth germs.

(PC2, 18% - between analysis, 17.8%) shows that (1) the buccal and the lingual part of the molar germ are much more different than are the two molar germs and (2) a large part of the bucco-lingual transcriptional differences is shared by the two molars.

Analysis of Differentially Expressed (DE) genes with DEseq2 [39] confirmed these findings. Taking into account the molar origin as a factor in DEseq2, we found a huge number of bucco-lingually biased genes: about one third of the genes expressed in our samples (4750 DE genes over a total of 15200 genes expressed, FDR<0.1). For a list of genes known to be functionally relevant during molar development (*i.e.* ∼200 genes from the bite-it database (`http://bite-it.helsinki.fi/`)), the proportion of genes biased toward the bucco-lingual axis is of 61%. ($\chi^2$ test, pvalue<2.2e-16). Remarkably, there was three time more DE genes for the bucco-lingual factor than for the upper-lower factor (FDR<0.1, DEseq2), a similar ratio as obtained from PCA. All this confirmed that the buccal and lingual part of the tooth germ are markedly different, in line with the PCA. In addition, when the DE analysis was performed independently on lower and upper samples, we found an important overlap (1260 genes, 76% of lower DE genes and 30% of upper DE genes, figure 4C), in line with the idea that the two molars largely share the same bucco-lingual transcriptomic bias. We noted that the number of DE genes was higher for the upper samples than for the lower samples, and a closer inspection revealed that this could be partially attributed to a statistical artifact: a difference in RNA-seq library size increased the detection power (despite statistical correction by DEseq2, see supplementary text S1) and thus inflated the number of weakly biased genes in upper samples. We will come back later on other biological reasons for this difference.

### Buccal halves are enriched in epithelium and lingual halves in mesenchyme, what inflates transcriptomic differences between them

A proportion of biased genes as high as one third of the genes expressed may seem surprising from a first inspection, yet we know from a previous study that differences in tissue proportion can drive pervasive transcriptomic signatures, slightly but significantly biasing very large cohorts of genes [21]. Specifically, we have shown that a difference in epithelium versus mesenchyme proportion between the lower and upper molar germ, as modest as 10%, can deeply impact their transcriptomes. In the present

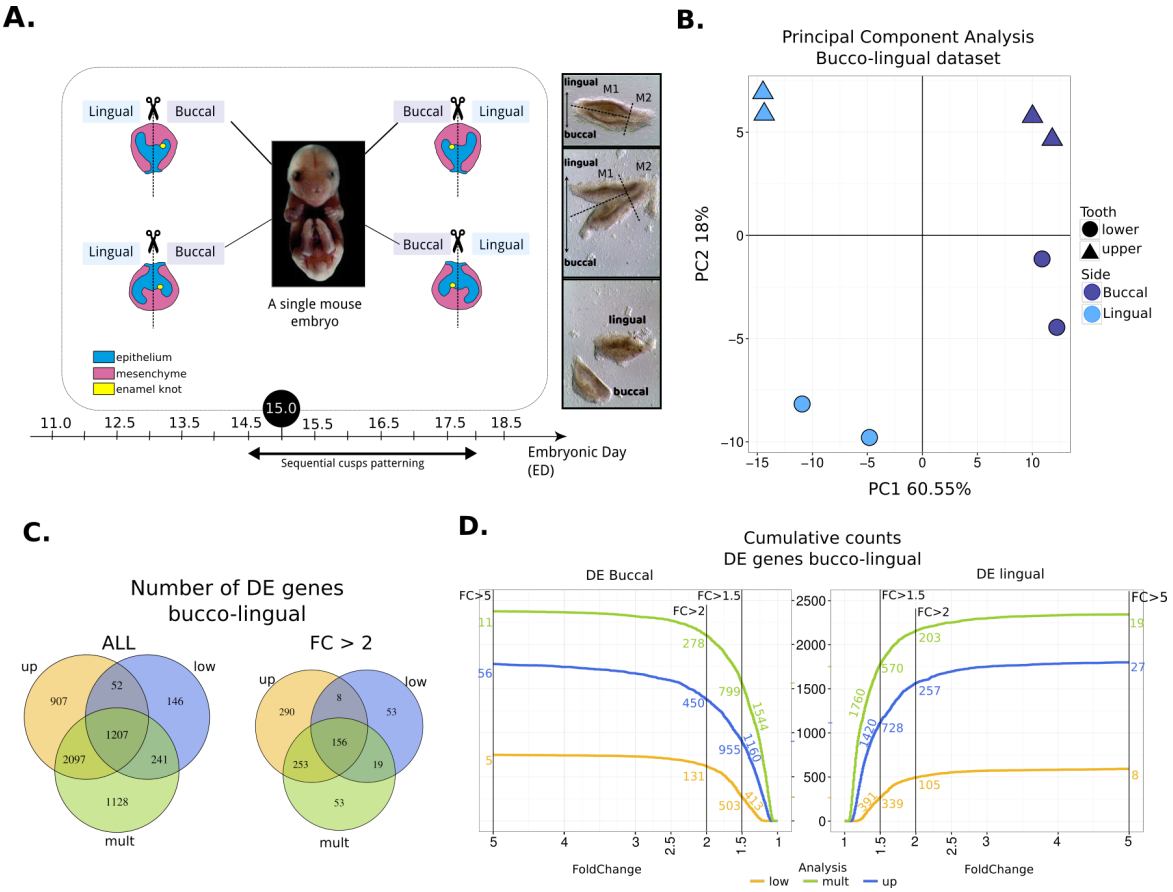



Figure 4 – **Large bucco-lingual differences in the transcriptomes are shared by lower and upper first molars. A.** Design of the RNA-seq dataset. The 4 molar germs of a single embryo were dissected at stage ED15 (embryonic day 15) and cut in half in order to separate buccal from lingual side. **B.** Principal Component Analysis of buccal and lingual RNA-seq libraries. The samples are mapped on the two main components (PC1 and PC2). **C.** Number of genes Differentially Expressed (DE, DESeq2 FDR<0.1) between buccal and lingual side for all DE genes (ALL, left) or for DE genes with a ratio >2 (right). Venn diagram representing the intersection of genes DE bucco-lingual in 3 different analyses: (green) mult: DE bucco-lingual (BL) in all samples taking into account the upper-lower effect, (orange) low: DE BL in the lower molar samples only, (blue) up: DE BL in the upper molar samples only. **D.** Number of differentially expressed genes with a bucco-lingual fold change below (or above) a given threshold (DE, DESeq analysis, FDR<0.1) in three different analyses: (green) mult: BL genes taking into account upper-lower differences, (orange) low: BL genes in lower molar samples only, (blue) up: BL genes in upper molar samples only. The number of DE BL genes with a fold change greater than 1.5, 2, and 5 is written for each type of test with the same color code.

case, a tissue proportion effect would be consistent with the large part of our 4750 DE genes showing a small fold change (2401 DE genes with enrichment <1.3, 51% of DE genes, figure 4D).

To search for such an effect, we first looked whether genes enriched in tooth epithelium (later designed *epithelial genes* for simplicity) and genes enriched in tooth mesenchyme (*mesenchymal genes*) are equally represented among genes with a bucco-lingual ratio below or over 1. We defined 2 classes of *epithelial genes*, based on the quartiles of the ratio epi/mes for epithelial and mesenchyme expression levels (FC obtained using DESeq2). The *highly specific epithelial genes* represent the 25% of more biased *epithelial genes* (ratio epi/mes > 1.9) and the *mildly specific* the subsequent 25% (ratio epi/mes between 1.3 and 1.9). Similarly we defined 2 classes of *mesenchymal genes* still based on quartiles of the ratio mes/epi. The *highly specific mesenchymal genes* represent the 25% of more biased mesenchymal genes (ratio mes/epi > 3.4) and the *mildly specific mesenchymal genes* the subsequent 25% (ratio mes/epi between 1.5 and 3.4. We conclude that the buccal part of the tooth germ is

made of more epithelial tissue while the lingual part is made of more mesenchymal tissue (figure 5. This obviously impacts the buccal and lingual transcriptomes, inflating the number of DE genes with a small BL ratio. This is true for both lower and upper molar and likely reflects the asymmetric morphogenesis of the germ, with three epithelial structures: the enamel cord, the enamel knot and to some extent the dental lamina being found more on the buccal side of the germ (figure 1).

**Buccal transcriptomes reveal putative new SEK genes.**

Among our list of DE genes, there is also a considerable number of genes with modest to large enrichment (1369 genes with FC > 1.5, 481 genes with FC > 2, 30 genes with FC > 5, see figure 4D). Among genes with BL ratio>2, there are significantly more genes biased buccally (278 genes, 58%) than lingually (203 genes, 42%) (Chisq test, *pvalue*=5.33e-7). What are those buccal genes? They are strongly enriched in epithelial genes (191 DE genes with Epi>2 Mes versus 69 DE genes with Mes>2 Epi, see also figure 6). Obviously, the enamel knot, with its buccally shifted position at this stage, could contribute to this pattern. Accordingly, out of 11 EK markers retrieved from the bite it database (*Dkk4, Edar, Edaradd, Fgf20, Fgf4, Fgf9, Ndrg1, Slit1, Wnt10a, Wnt10b, Wnt3* - specificity confirmed by published *in situ* hybridization data), 10 were significantly biased toward the buccal side at ED15.0 (only *Edaradd* was non significantly biased toward the buccal side and *Fgf4* was the only gene for which FC < 2 ). But to which extent the EK contributes and can we use our datasets (bucco-lingual and epithelium-mesenchyme) to extract a list of EK candidate genes?
To answer these questions, we take advantage of a previously published microarray profiling of ED14.5 enamel knot and ED13.5 epithelium [43]. For genes DE in epithelium versus mesenchyme in our own dataset (with a FC>2), we recovered an "EK index", as the ratio between expression in EK over expression in epithelium. Increasing the threshold, genes were rapidly less numerous but increasingly enriched in buccal genes (Supplementary figure **??**): 65% for a 1.35 ratio and 80% for a 1.7 ratio. We then used this property to extract a list of SEK enriched genes with 2 levels of confidence (high for 1.7; reasonable for 1.35, Supplementary Table 1 and figure 6). For reasonable-confidence SEK markers (113 genes), 42 genes had a FC>2 but only 8 are annotated as EK genes in bite it (7% of these "new" marker genes, these 8 genes are *Shh, Dkk4, Fgf9/20, Ndrg1, Wnt10a/b* and *Wnt3*). This suggested that many Enamel Knot genes are still uncovered. We noted that we recovered genes involved in the *Edar-Wnt* axis evidenced in other ectodermal organs [47] (with *Edar, Wnt10a, Wnt10b, Dkk4* and *Kremen2*, a receptor for DKK proteins that antagonizes Wnt signaling, suggesting that this signalization is at work during cusp patterning. We also found a number of genes involved in cell-cell adhesion (*Desmocolin2* and *3*, *Claudin1*, *Plakophilin1, Ladinin, Stratifin, Madcam1*), that may participate to the highly organized cellular organization of the enamel knot.

**Buccal transcriptomes reflect the presence of the Enamel Knot cells, but also the formation of the stellate reticulum and tooth innervation from the buccal side.**

Thus, a number of buccally biased genes are genes expressed in the enamel knot. However, this represents only 3.5% (high confidence) and 17% (reasonable confidence) of epithelial buccal genes with FC>1.5, that is a rather small proportion. *Epithelial genes* with an EK index below 1.1 still represent 50% of epithelial buccal genes with FC>1.5. Even if we take into account that a rapid maturation of the enamel knot between ED14.5 and ED15.0 could make us mis-annotate a number of genes as non-enamel knot genes, this likely leaves a great proportion of buccal DE genes biased for other reasons than the presence of EK on the buccal side. This strongly suggests that large number of genes, other than enriched in the enamel knot, show a buccal bias. We thus examined the list into more detail by performing functional analyses of buccal DE genes (FC>2) with topGO, Gorilla and String. To increase the specificity of these analyses, we looked separately at genes enriched in the epithelium or in the mesenchyme (FC>2; DE FDR<0.1), including or not putative enamel knot genes, and eventually focusing on genes with an EK index below 1.1, that have almost no chance to be enriched in EK.
First, we noted that buccal regionalization was mainly carried by the epithelium, yet some regionalization was also found in the mesenchyme (figure 6). This is not only consistent with the presence of epithelium, but also with the excess of epithelium and depletion in mesenchyme quantified by our deconvolutions (figure 5). Buccal epithelial genes were strongly enriched for genes involved in cell adhesion (homophilic and heterophilic), and cell junctions (desmosomes,adherens junction); but also chemotaxis and locomotion. Genes enriched in EK (EK index>3) were sufficient to recover these GO terms, suggesting that they can be attributed to the Enamel knot and possibly the enamel cord that share many characteristics with the enamel knot. Interestingly, GO terms associated with ion transport and water homeostasis are also found (*eg.* ion channels and regulators): this may be related to the formation of stellate reticulum, the internal tissue of the epithelial part composed of star shaped cells in a fluid-rich matrix (fig. 1 and fig. 7). Enrichment is even stronger in genes with EK index below 1.1, confirming that it concerned a different tissue compartment. Specifically for these genes, we recovered GO term associated with non canonical Wnt signaling *(Wnt4, 7a, 7b, Fdz3* and *5, Grhl3, Celsr*), G-protein coupled receptor signaling pathway, and epithelial cell differentiation (including columnar/cuboid cell differentiation). This could reflect inner/outer enamel epithelium and/or stellate reticulum formation (fig. 1 and fig. 7). Finally, buccal mesenchymal genes were mainly enriched for genes related to nerve function and formation (*eg.* potassium voltage dependent

channel or glutamate signaling). This is in line with nerve fibers colonizing the tooth germ at the cap stage, from its buccal side [48].
To summarize, buccal transcriptomes reflect the presence of the enamel knot, but also of other cell populations that differ in nature or abundance in buccal and lingual parts of the germ: this was known for nerve colonization, it is likely the case for forming stellate reticulum, and the specification of inner/outer enamel epithelium. Thus, our analysis provides lists of genes to be studied to gain insight into specification and functional specialization of the dental epithelium at the cap stage.

**Lingual transcriptomes are characterized by a population of migrating endothelial cells.**

As mentioned above, we found 203 buccal genes with a FC>2. Again this is an unexpected large number of "lingual genes" under a simple model where only the presence (or in this case the absence) of the enamel knot would structure buccal and lingual transcriptomes. This suggests that the lingual side is far from being a naive territory. In contrast with buccal genes, lingual genes are strongly enriched in mesenchyme-specific genes (see figure 6), suggesting that the lingual regionalization is essentially carried by the mesenchymal part of the tooth germ. In the epithelium, lingual genes are enriched for genes involved in cell cycle arrest (inc. *Cdkn2a, Cdkn2b* and *Nkx3.1*) and G-protein coupled receptor signaling pathway (*e.g. Ccr5*). It is noteworthy that the GO term G-coupled receptor signaling pathway was also found for non-EK buccal epithelial genes: different genes thus fulfill qualitatively or quantitatively different functions on the two sides of the germ. In the mesenchyme, there was a consistent enrichment for genes involved in regulation of angiogenesis, vasculature development and hematopoiesis. This includes enrichment for genes involved in leukocyte activation and proliferation, which have recently been shown to play a role in angiogenesis by secreting pro-angiogenic factors and assisting physically vessel sprouting [49]. We also noted strong enrichments for genes involved in chemotaxis/cell migration and locomotion (inc. G-protein coupled receptor *Cx3cr1* and a number of chemokines; *e.g. Cxcl1,5* and Ca2+ release involved downstream of G-protein receptors). Overall, this likely reflects the invasion of tooth mesenchyme by endothelial cell later forming tooth vasculature [50, 51]. It was however never noted that this process could start from the lingual part of the tooth germ or be stronger in the lingual part. Accordingly, in our data *Cd31* and *Cd34*, two early markers of blood vessel formation are only marginally biased (FC 1.2 to 1.4). This may suggest that only the earliest steps proceed from the lingual side of the tooth germ (*e.g.* endothelial cell attraction and migration), whereas later steps happen more or less homogeneously in the germ. Our results provide a list of candidate genes for regulating the overall process. Interestingly, it includes negative regulators of angiogenesis and hematopoietic lineage proliferation, suggesting that the process is tightly controlled (*e.g. Tgfbr2, Meox2*, the two partners *Hhex* and *Tal1, Lect1*...).

**Putative determinants of asymmetric molar development.**

With this global view of buccal and lingual transcriptomes in hands, we wanted to look for possible determinants of the asymmetry. Because data in the literature are scattered, we decided to review them, as a first step. Table 2 lists the genes for which a bucco-lingual bias of expression was mentioned in the literature from initiation up to the cap stage, and genes for which there is a perturbation of bucco-lingual development. This identifies BMP, Wnt and Shh pathways as key players, together with the upstream or downstream transcription factors *Msx1, Sox2, Osr2* and *Runx2*. Below, we examine these genes and pathways into more detail (genes DE with FC>2). Tight regulation of BMP and Wnt pathways has been recently shown to be critical for cap transition [15]. Interestingly, this tight regulation is organized in space bucco-lingually: at the bud stage, just before cap transition, both pathways are antagonized on the lingual side of the germ. *Osr2* antagonizes *Msx1* expression, and by this way constrains *Bmp4* expression on the buccal side [24, 14, 15]), and *Dkk2*, and *Sfrp2* antagonize Wnt activation, with expression of *Lef1* (a component of the pathway under positive feedback loop) being extended buccally. On top of that, tight regulation of Wnt pathway is also critical for inhibiting the formation of a tooth replacement bud on the lingual side. Regarding the BMP pathway in our bucco-lingual dataset, that is at ED15.0 after cap transition, the *bmp4* gene was no longer (lower molar) or only poorly (upper molar) biased, but *Osr2* was strongly biased (both results are consistent with data by Zhang *et al.* [24]). Moreover, we noted that the BMP modulator, *Bmper*, was expressed in mesenchyme with a lingual bias, and could possibly modulate the BMP pathway there. The *Bmp5* gene shows a similar lingual bias. In the buccal epithelium, we recovered *Bmp7* and *Msx2* as expected, given their known expression in the EK and buccal epithelium respectively.

Shh signaling seems especially relevant for B-L asymmetry, as mutants for two genes in the pathway (*Shh* [52] and *Evc* [27], a protein necessary for cilia, himself necessary for Shh signaling [53, 54]) display selective hypoplasy and hyperplasy, respectively, of their lingual side [52, 26, 27]. Moreover, at the bell stage, Shh signaling is necessary to sustain the epithelial stem cell marker *Sox2* in cervical loops (via *Gli1*). However, this effect is counteracted by the BMP pathway in the differentiating inner enamel epithelium [17]. Here, we recover -as expected for an enamel knot gene- biased *Shh* expression in the buccal epithelium. *Sox2* interestingly exhibited a clear lingual bias. The network identified at bell stage could thus be in place at cap stage. This could provide a mechanism for differential growth and/or competence maintenance of the dental epithelium, corresponding to higher *Sox2*

expression in the lingual part, if only BMP signaling is higher on buccal than on lingual side. *Bmp7*, which is expressed in EK (epithelial and buccal in our datasets) and later in the whole inner enamel epithelium except for cervical loops [55] could play this role, especially if it signals at shorter ranges than *Shh*. Alternatively, BMP modulators could limit BMP signaling on lingual side, such as the *Bmper* gene mentioned above.

Wnt signaling seems also relevant, as ectopic Wnt activity on lingual side results in formation of supernumerary, replacement-like tooth in *Lrp4* mutant [56]. A similar phenotype is seen in the BMP antagonist *Sostdc1* (*aka Wise, Ectodin or USAG-1*) mutant [57]. In fact, *Lrp4* and *Sostdc1* interact physically, and *Sostdc1* can modulate Wnt signaling [56, 58]. Wnt signaling is also found in successional dental lamina of reptiles [59]. In mouse molars, down regulation of Wnt signaling on lingual side may be necessary to suppress replacement tooth formation. In our data, we found that Wnt signaling was strongly represented buccally, with a number of Wnt ligands buccally biased (*Wnt10a,b, 16, 3* in epithelium and *Wnt2* and *9* in mesenchyme), as well as modulators (*Dkk4, Sfrp5* and *Kremen* in epithelium; *Notum, Wif1, Sfrp4* and the transcription factor *Lef1*). In the lingual part, biased Wnt pathway genes were only modulators (*Prickle1, Dkk2, Wif1* in mesenchyme). This is consistent with the idea that Wnt signaling is largely specific for the buccal side of the germ at cap stage, and possibly suppressed on lingual side.

In addition, in the course of our GO analysis, we noticed that the buccal part of the germ was associated with active Retinoic Acid (RA) signaling (*eg. RARb, RXRg, RORa, NR2F1, Aldh1a3, NR4a3* were buccal), whereas the lingual side of the tooth germ is protected from RA signaling by the stronger expression of two RA degrading enzymes (*Cyp26a1* and *c1*, both strongly lingual). This would suggest higher RA signaling in the buccal part of the germ. In molars, RA is necessary for cervical loop growth and RA excess at ED14.5 or ED16.5 inhibited specifically cusp formation but not cervical loop growth [60]. In absence of RA, withdrawal from cell cycle is anticipated in cervical loops, but surprisingly delayed in cusp inner enamel epithelium. Absence of RA signaling in the lingual part of the germ could thus in principle limit the growth of cervical loops but be favorable to that of inner enamel epithelium (maintenance of cell cycling). This is an interesting possibility regarding the results presented above: during ED14.5-ED15.5 stage, the lingual IEE expands rapidly, shifting the enamel knot on the buccal side, but surprisingly, a corresponding downgrowth of the lingual cervical loop was not that obvious. RA signaling is known to activate Fgfs and Shh. We note that *Fgf9* shows a pattern similar to Shh, with an expression in PEK first and then expanding in the IEE at bell stage [61]- in contrast with *Fgf4* or *Fgf20* that marked SEK-. Finally, we also discovered two transcription factors, specific for the lingual side in both lower and upper molar: *Foxl1*, in the mesenchyme, is a target of Shh signaling in some contexts [62]; *Sim2*, in the epithelium, is a target of Wnt signaling in some contexts [63].

## The exaggerated asymmetry of the upper molar transcriptomes has multiple sources and highlight.

As seen earlier, the asymmetric expansion of the lingual cervical loops is transient in the lower molar whereas it persists longer in the upper molar in relation with the formation of extra lingual cusps. Moreover, from visual inspection of the two molar germs, it seems that the asymmetry could be more pronounced in the upper molar as early as ED15.0 stage. Below, we looked for a confirmation of this early asymmetry in transcriptomes.

### Exaggerated asymmetry of the upper molar B/L transcriptomes.

First evidence came from the PCA analysis above. Indeed, the distance between buccal and lingual libraries on the PC1 axis is 1.33-time larger in the upper molar than in lower molars (figure 3B and supplementary figure A). To test the significance of this ratio, PCA were run multiple times on subsets of 1000 randomly chosen genes. The distribution of upper-lower ratio of bucco-lingual distance is shown in supplementary figure A following a normal distribution of mean=1.258 and standard deviation=0.0038 (Kolmogorov-Smirnov test, $pvalue$<2.2e-16). Thus, the higher distance between buccal and lingual upper transcriptomes is a robust property of these transcriptomes, carried by large number of genes, what we called in a previous study a transcriptomic signature [21]. The analysis of genes Differentially Expressed (DE) was congruent with this result. There are twice more genes DE between buccal and lingual side in upper molar. Even after removing statistical artifacts such as RNA-seq libraries size and variability between samples (see supplementary text S1), there were still more genes DE in upper molar. For instance, removing library size effect (by randomly choosing the same number of reads in all libraries, see material and method) still lead to 2 times more genes DE between buccal and lingual side in upper molar. Furthermore, we used a methodological trick to apply variability of upper molars to lower molars samples and vice versa (see material and method). These analyses decreased the excess of DE genes in upper molar, however there is still 1.25 to 1.65 more genes DE between buccal and lingual side in upper molar. Next, we investigated possible sources of this higher asymmetry of upper molar germ transcriptomes, looking at two factors we know from previous studies [21, 46] to generate transcriptomic signatures: developmental maturation and epithelium/mesenchyme proportions.

| Gene | bias | stage | tissue | pathway | regulation | remarks | ref |
|---|---|---|---|---|---|---|---|
| | | init/ bud/ cap | | | | | [55] |
| *Bmp4* | B | bud | Mes | BMPs | | | [64] |
| | | cap | | | | | [14] |
| *Msx1* | B | init/ bud/ cap/ early bell | Mes | BMP | induced by BMP4 | | [29] |
| *Sox2* | L | bud/ cap/ early bell | Epi | BMP | Repressed by BMPs at bell stage | Expressed in cervical loops at ED16. Link with replacement teeth in reptiles | [29] |
| *Osr2* | L | 13.5 | Mes | BMP | Antagonizes *Bmp4* expression | Formation of supernumerary lingual teeth $Osr2^{-/-}$ *mutant mice* | [24] |
| *Dkk2* | L | bud | Mes | Wnt | | expression in condensed MES | [15] |
| *Lef1* | B | bud/ cap/ early bell | Mes | Wnt | | | |
| *Sfrp2* | L | bud | Mes | Wnt | | Expressed in condensed MES | [15] |
| *Spry4* | L | bud | Mes | Wnt | | | [65] |
| *Fgf15* | L | init | | FGF | | | |
| *Fgf18* | B | init | | FGF | | | [66] |
| *Ptch1* | L | cap | Epi | Shh | Negative feedback of Shh pathway | Mesenchymal expression is not biased in the BL axis | [52] |
| *Tnfrsf19* | B | bud/ cap/ bell | | TNFRSF19 | | | [67] |
| *Msx2* | B | bud | Epi | TNFRSF19 | | Expressed in the enamel chord | [61, 68] |
| | | cap/bell | | | | | [20] |
| *Alcam* | L | cap | Mes | MPi | | Expressed in dental follicle | [69] |
| *Timp2* | L | cap | Mes | MPi | | | [70] |
| *Timp3* | B | cap | Mes | MPi | | | [70] |
| *Dlx2* | B | bud | Epi | | | Mesenchymal expression is not biased in the BL axis | [71] |
| *Pitx1* | L | cap | epi | | | | [52] |

Table 2 – **Gene biased bucco-lingually during mouse first molar development.** B: buccal bias; L: lingual bias; Epi : epithelium ; Mes : mesenchyme ; MPi: Metalloproteinases Inhibitors.

**Differences in maturation between buccal and lingual parts are more pronounced for the lower molar.**

As already shown elsewhere [46], the buccal transcriptome appears "older" than the lingual transcriptome when projected on a PCA of whole tooth germ transcriptomes from ED14.5 to ED18.0. This (together with further evidences) was interpreted as the buccal transcriptome being developmentally more mature than the lingual transcriptome, partly (but not only) because of the presence of the enamel knot. Therefore, one possibility to explain the excess of buccal-lingual differences in upper molar could be that the B/L difference in developmental maturation is exaggerated in the upper molar. To test for this hypothesis, bucco-lingual transcriptomes were projected on the whole tooth germ PCA (for both upper and lower molar) and the distance between buccal and lingual transcriptomes was measured on PC1 axis: in fact, the distance was shorter for the upper molar than for the lower molar (ratio of 0.79, supplementary figure A) We used the same method described above to estimate the statistical relevance of this ratio. The distribution was shifted toward 0.88 and was significantly different from a normal distribution centered around 1 (Wilcoxon test, *pvalue*=2.863e-08, supplementary figure A). Hence, differences in developmental maturation between buccal and lingual sides are more pronounced in the lower molar than in the upper molar. This cannot contribute to increase upper molar B/L transcriptome asymmetry. In addition, we noted that there is a sharper age difference between buccal transcriptomes than between lingual transcriptomes.

This suggests that the maturation of the buccal side is especially more advanced in the lower molar.

**B/L difference in epithelium-mesenchyme proportions exaggerate the asymmetry of upper molar transcriptomes.**

As shown above, a difference in Epi/Mes proportion is part of the bucco-lingual asymmetry. Could an exaggeration of this difference be responsible for the exaggerated asymmetry of upper molar transcriptomes? Our deconvolution analyses showed it is indeed the case (figure 5B), and the effect is essentially driven by the lingual side, being further depleted in epithelium and enriched in mesenchyme in the upper molar, as compared with the buccal side.

**The more buccal position of the enamel knot contributes to exaggerate the asymmetry of upper molar transcriptomes.**

Our morphological data suggested a faster lingual expansion in the upper molar as early as ED15.0. This could in principle influence the positioning of the enamel knot at ED15.0, more rapidly shifted to the buccal part of the tooth. We tested this hypothesis by performing deconvolution with enamel knot marker genes: indeed, the difference in EK proportion was stronger for upper molar. Moreover, we noted that enamel knot markers that have an extremely restricted area of expression (such as *Fgf4, Fgf20*) had virtually no expression in upper lingual samples, and small but significant expression levels in lower lingual samples. Thus, these findings were consistent with the enamel knot having a slightly more central position in the lower molar, and a slightly more buccal position in the upper molar at ED15.0, which is again consistent with exaggerated lingual development in the upper molar from a very early stage.

**Putative determinants of bucco-lingual differences between the two molars.**

To further understand possible differences between the two molars, we looked for functional categorization of 1) genes that are DE only in one type of molar and 2) genes with exacerbated fold changes in one type of molar. We took advantage of the GLM method (from DESeq2 package) to extract the interactions between the side factor (lingual vs. buccal) and the tooth factor (upper vs. lower) with the following formula: $\sim$ tooth + side + tooth:side. This method enabled us to identify genes that have a different bucco-lingual bias between tooth types, called here *exacerbated genes*, such as differences in the intensity of the bias (*e.g.* same BL effect but stronger in one tooth type than the other).

Genes specifically DE in the buccal part of the lower molar (25 DE genes with FC>2 and 105 genes with FC>1.5) were strongly enriched for genes involved in bone development (*Fgfr3, Runx3, Foxc1...*). This could suggest that the process of encapsulation of the tooth germ by the bone starts from the buccal side. Genes specifically DE in the lingual part of the lower molar (6 DE genes with FC>2 or 41 genes with FC>1.5) showed no remarkable enrichment. Examination of genes with exacerbated FC in lower molar (96 genes) revealed genes involved in branching morphogenesis. For the buccal side, these were genes involved in mesenchymal cell proliferation: *Fgfr2, BMP2, Dachs2, Sox10, Eya1*. For the lingual side, these were *Fgf10* and a gene involved upstream of *Fgf10* expression in kidney, *Fras1*. *Fgf10* is thought to control cervical loop growth during bell formation, and this may thus suggest that a lingual burst of *Fgf10* could sustain lingual cervical loop growth at stage ED15.0 in the lower molar. Given that *Fgf10* levels, overall start increasing in the tooth germ at ED15.0 (data not shown), with some delay in the upper molar, it seems plausible that the burst is just delayed in upper molar (and indeed, stronger expression is seen on lingual side, although not as sharp as in lower molar). Also, in lower molar the *Dusp6* and *Etv4* FGF target genes are slightly lingual (FC=1.5 and 1.2) whereas they are clearly buccal in the upper molar (FC=1.8), suggesting that FGF signaling is indeed delayed in upper molar, i.e. still mainly buccal. This finding also suggests that this burst would only be implicated in a late phase (ED15.0 to ED15.5, even further for upper molar), and not for initiating cervical loop formation (ED14.5-ED15.0). This is also consistent with our 3D data showing active cervical loop and cusp formation in the lower molar at ED15.0 as compared with the upper molar. A critical local growth regulator, *IGF-1* (mainly expressed in the mesenchyme), was also twice higher in the lingual part of the lower molar only.

Genes specifically DE in the upper molar reflected the especially unbalanced composition of its buccal and lingual halves demonstrated above. Thus, not surprisingly, the buccal genes (23 DE genes with FC>2, 495 DE genes with FC>1.5) were enriched for epithelial genes involved in cell adhesion, cell junction assembly and epithelial differentiation, and genes expressed in EK and enriched for mesenchymal genes involved in neural differentiation. The lingual genes were enriched for genes involved in endothelial cell migration and differentiation. *Exacerbated genes* in the upper molar again mainly reflected unbalanced tissue composition, with similar results as for upper-molar specific genes. Closer inspection of these lists of genes reveal a number of interesting candidates for further studies (FC>3). In the upper molar only, *Osr1* transcription factor showed a strong buccal bias, in contrast with its paralogue *Osr2* found lingually. The paralogue of *Barx1*, *Barx2*, is found in the epithelium with a very strong buccal bias. Signaling molecules with strong bias include *Activin$\beta$B* (aka *Inhbb*), its downstream target *Foxi3*, *Endothelin-1* and several members of the Wnt pathway not described yet in the tooth: *Wnt7a, Wnt3a, Wnt3* and *Kremen* (epithelium). Finally, the Wnt regulator *Rspo4* in the mesenchyme also showed a strong buccal bias.

*Alx1*, expressed in mesenchyme, showed a strong lingual bias in upper molar only (FC>4). It is especially in-

teresting as this transcription factor has been involved in craniofacial defects in burmese cats [72] (including tooth defects to be characterized) and epithelium-mesenchyme transition in sea urchin [73]. *Nos1* was in the top of exacerbated genes (and GO terms associated with NO signaling were recovered on lingual side): it is largely specific for the lingual part of the molar, and twice more expressed in the lingual upper molar. It is interestingly involved in convergence-extension of craniofacial cells in zebrafish [74]. Convergence-extension is a plausible mechanism for rapid surface increase of the lingual IEE in early steps of cervical loop formation (ED14.5 $\longrightarrow$ ED15.0).

Finally,*Engrailed-1* was also among the top *exacerbated genes*, specific for the lingual part of the molar, and twice more expressed in the lingual upper molar. The transcription factor *En-1* act in the limb with a complementary pattern to *Wnt7a*, repressing in the ventral limb the differentiation induced by *Wnt7a* dorsally [75]. A similar antagonism may be at work in the molar, as we identify *Wnt7a* as a DE buccal gene and *En-1* as a lingual gene. High levels of *En-1* might protect the lingual side from Wnt activation and therefore cusp formation.

## Concluding remarks

Our combination of morphological data and extensive transcriptome analysis greatly expands our view of the ED15.0 tooth germ and its bucco-lingual development (summed up in figure 7). High regionalization of the molar germ is obvious in both analysis of 3D shape changes and buccal/lingual transcriptomes. The germ appears a much more complex than histological sections in the early cap would suggest. Innervation and vascularization are already ongoing. The buccal epithelium is highly regionalized, with formation of stellate reticulum, inner/outer enamel epithelium, differentiation of IEE but the buccal epithelium is very different: most likely because the patterning of these cell populations is delayed. Indeed the lingual side seems to be actively protected from this differentiation (Wnt and RA signaling antagonism, stronger sox2 expression). Obviously Wnt, Shh, BMP and RA signaling are key players in this asymmetric regionalization We also identify other strongly biased transcription factors, such as *Alx1*, *Foxl1*, *Sim2*, whose role remain to be determined. The question of the primary determinant is still open. *Osr2* is an obvious candidate: it makes sense both developmentally and evolutionary speaking, because *Osr2* it plays a role in patterning dental lamina along the bucco-lingual axis (*e.g.* in both mammals and reptiles. This bucco-lingual position information could have been easily re-used to pattern molars in the bucco-lingual direction. Based on our data suggesting that the buccal and the lingual part of the germ are much more different than could be expected, we thought to revise our view on the lingual epithelium, starting to think that at 15.0, it might not be equivalent to the buccal epithelium, but closer to a dental lamina or a successional lamina, before a tooth is induced (no BMP nor Wnt activity). Adopting this view, we intrigued that the Wnt suppression we see in the lingual part (with the interesting *En-1/Wnt7* antagonism) could have two roles: limiting the number of tooth generations to 1 (defective in mutants for Wnt antagonists, exhibiting a small replacing-like tooth) and repressing the formation of the lingual part of the tooth from the successional/dental lamina (partly defective in caspase 3 mutants, where a small loop is forming). As a result, the formation of a *bona fide* lingual cervical loop could be postponed, which would result in enlarging the tooth germ in the lingual direction. A secondary SEK could then be patterned in this lingually enlarged tooth germ. In the lower molar, the process would cease, with definitive loss of competence under the influence of BMP and Wnt signaling from this SEK (except at the tip of cervical loops). A rudimentary successional bud could however form at the base of the germ. In the upper molar, this process could go on further to enable further lingual growth and supplementary cusp patterning. We find this mechanisms interesting because it could explain concomitantly molar evolution (from triconodont teeth having already evolved SEK, but in the antero-posterior direction) and the reduction of the regenerative potential, what is consistent with evolutionary evidences linking both trends. Further work in mammals and reptiles on the genes identified in this study could help to test this model.

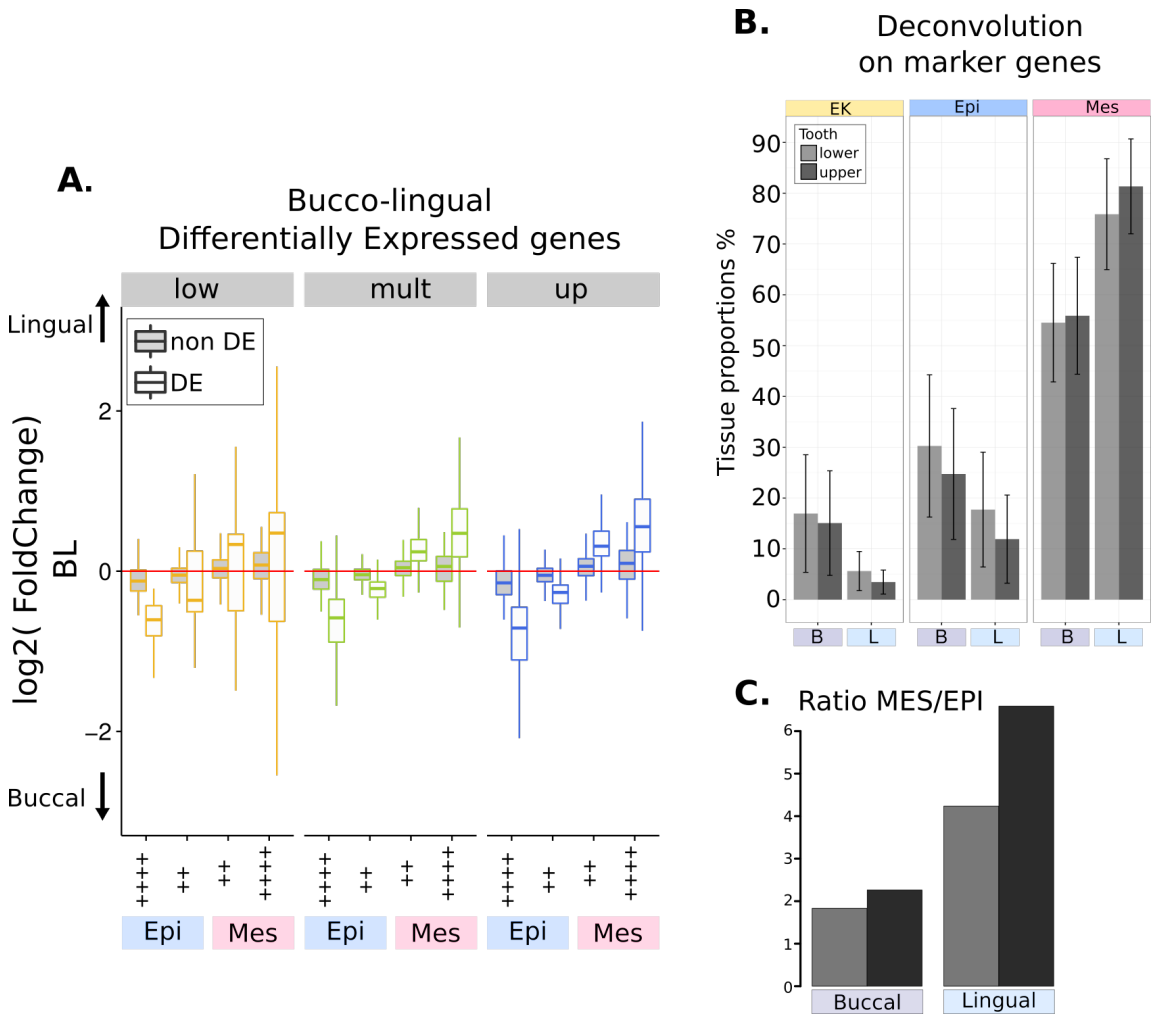



Figure 5 – **A difference in epithelium-mesenchyme composition of buccal and lingual part of the germ contributes to BL transcriptomic differences. A.** Log2 Fold change of genes differentially expressed between buccal and lingual side (BL genes) in three different analyses (mult: BL genes taking into account upper-lower differences; low: BL genes in lower molar samples only; up: BL genes in upper molar samples only). On the x-axis, genes are indexed according to their expression levels in pure tissue transcriptomes, the threshold have been defined according to the quartiles of the ratio between epithelial and mesenchymal expression levels: epi ++++: Epi /Mes > 1.9; epi ++: 1.3< epi/mes < 1.9; mes ++: 1.5< mes/epi < 3.4; mes ++++: mes/epi > 3.4. On the y-axis the positive log2FC correspond to genes biased toward the lingual side and negative log2FC correspond to genes biased toward the buccal side. Note that genes specific to the epithelium (from highly to mild) are significantly biased toward the buccal side (Wilcoxon test, *p-value*<2.2e-16, for the 3 analyses) whereas genes specific to the mesenchyme (from highly to mild) are significantly enriched toward the lingual side (Wilcoxon test, *p-value*<2.2e-16, for the three analyses). **B.** Deconvolution of bucco-lingual samples for the three tissues (based on marker genes for each tissue): (i) the enamel knot (orange, 16 genes, a pool from bite it database and microarray expression [43]such as ek > 1.7 epi and from RNA-seq data Epi > 2 Mes), (ii) the epithelium (blue, 16 genes randomly chosen from a set of 121 genes, ek< 1.1 epi from microarray data of pure tissues and Epi > 2 Mes from RNA-seq expression data in pure tissue and highest quartile [95%] of fold change between epithelium and mesenchyme RNA-seq expression in pure tissue) and (iii) the mesenchyme (pink, 16 genes randomly chosen from a set of 191 genes ek< 1.1 epi from microarray data of pure tissues and Mes > 2 Epi from RNA-seq expression data in pure tissue and highest quartile [95%] of fold change between epithelium and mesenchyme RNA-seq expression in pure tissue). **C.** Ratio of mesenchyme proportion over epithelium proportion (obtained with deconvolution. Lingual side is enriched in mesenchyme (over epithelium) compared to the buccal side. Moreover, there are no differences in relative proportion of mesenchyme over epithelium in buccal side between upper and lower molars, whereas for the lingual side the upper molar is enriched in mesenchyme over epithelium.

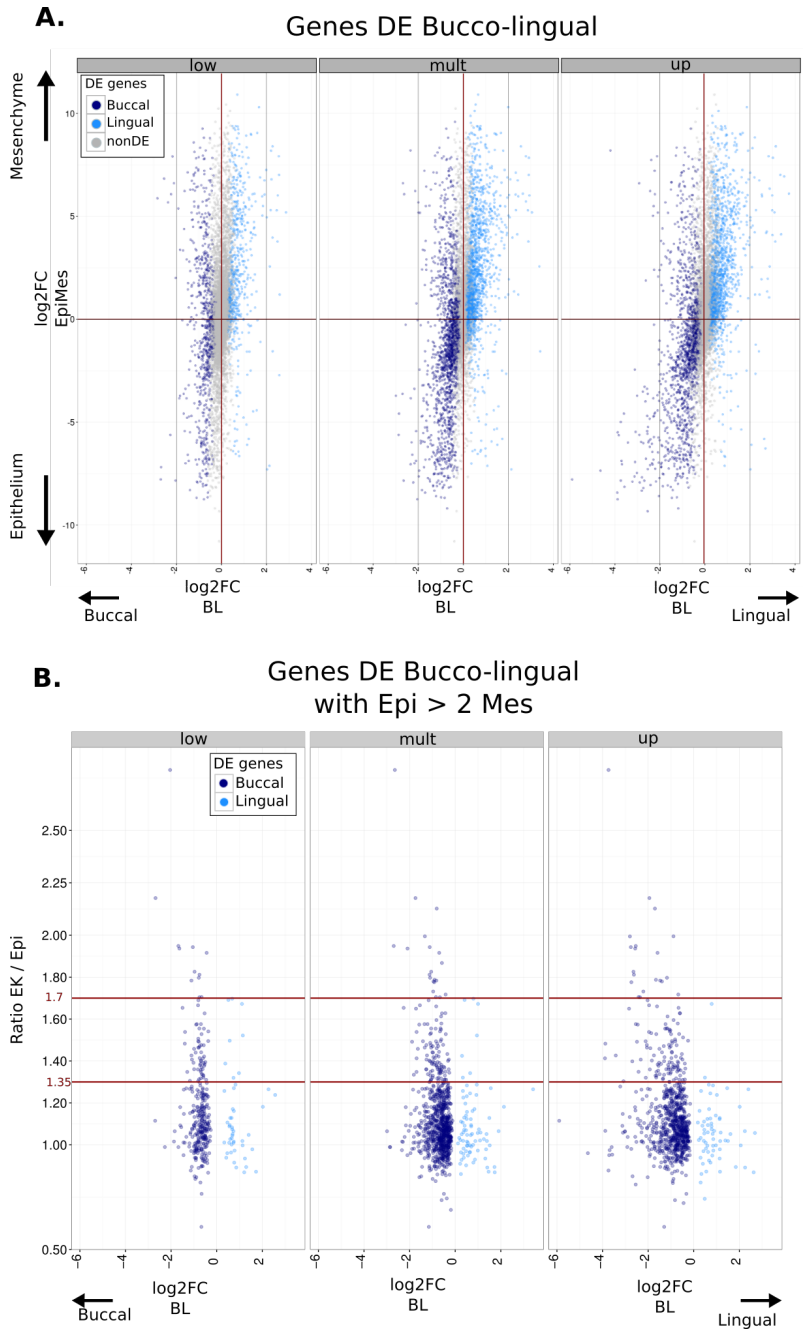



Figure 6 – **Transcriptional regionalization on buccal and lingual side are mainly carried by the epithelium and mesenchyme, respectively. A.** x-axis bucco-lingual log2FoldChange for buccal genes (darkblue), lingual genes (light blue) and non DE genes (gray). Y-axis epithelium-mesenchyme log2FoldChange for genes DE between epithelium and mesenchyme in the molar germ at ED15. Epithelial genes log2FC<0 and mesenchymal genes log2FC>0. Note that buccal genes are enriched in epithelial genes and lingual genes are enriched in mesenchymal genes. **B.** x-axis bucco-lingual log2FoldChange for buccal genes (darkblue)and lingual genes (light blue) and y-axis ratio EK/epi (from microarray data [43] for genes DE bucco-lingually and for which Epi>2 Mes in pure tissue transcriptomes. Note the enrichment toward bucaal genes. Moreover, we defined two ek/epi threshold to define enamel knot genes in our BL dataset. The highly specific ek genes: ek> 1.7 epi and epi>2 mes and log2FC BL <0 (*i.e the buccal genes*). The reasonable ek genes: ek> 1.35 epi and epi>2 mes and log2FC BL <0.

Combined developmental and transcriptomic approaches reveals extensive bucco-lingual asymmetry in mouse molar development

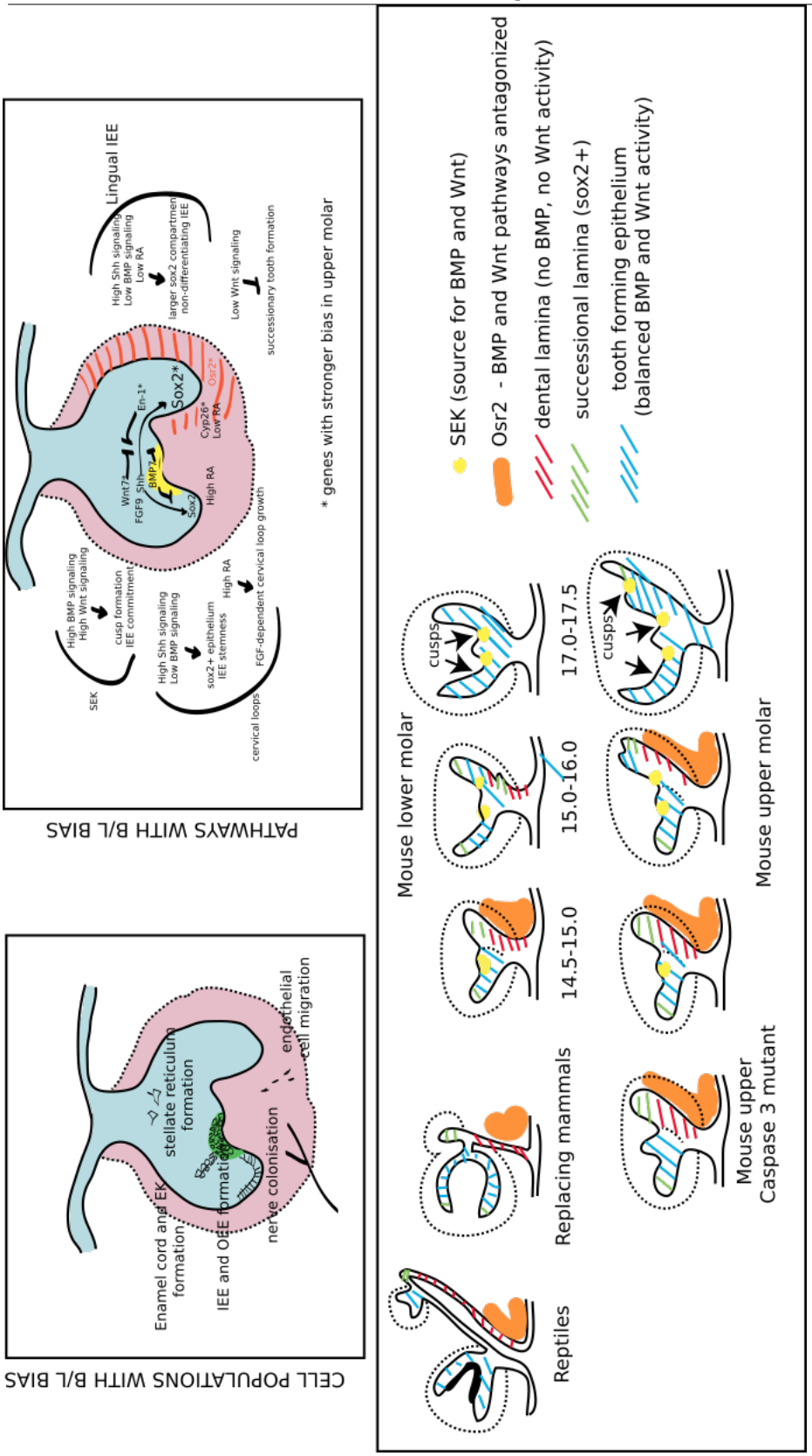



Figure 7 – **Summary of findings and proposed model for bucco-lingual development and evolution. Top, left**:. bucco-lingual differences in cell populations **Top, right**: bucco-lingual differences in molecular pathways and their possible role. **Bottom**: a putative model for the evolution of bucco-lingual development from reptiles, replacing mammals to mouse. In reptilian ancestors, several tooth generations arise from the bucco-lingually patterned dental and successional lamina (*Osr2* dependent). In replacing mammals (conical teeth), the three kinds of epithelial populations remain. In the mouse early cap, only the buccal epithelium may really correspond to a tooth germ epithelium, while the lingual part would rather correspond to the dental lamina/successional lamina. This rapidly resumes in the lower molar as the second SEK is patterned, and the lingual part of the germ becomes a true tooth epithelium (later a successionaltooth buds forms at the base of the germ). In the upper molar, this resumes only later on, enabling the extended lingual growth. In both molars, the tips of the cervical loops maintain high proliferative capacity (*Sox2* expression) as the successional lamina.

# References

[1] P. M. Butler, "The ontogeny of molar pattern," *Biological Reviews*, vol. 31, pp. 30–69, feb 1956.

[2] J. Jernvall, *Mammalian molar cusp patterns*. Finnish Zoological and Botanical Publishing Board, 1995.

[3] J. Jernvall and I. Salazar-Ciudad, "The economy of tinkering mammalian teeth," in *Tinkering: The Microevolution of Development: Novartis Foundation Symposium 284*, pp. 207–224, Wiley Online Library, 2007.

[4] J. Jernvall and I. Thesleff, "Tooth shape formation and tooth renewal: evolving with the same signals," *Development*, vol. 139, pp. 3487–3497, sep 2012.

[5] A. Brook, J. Jernvall, R. Smith, T. Hughes, and G. Townsend, "The dentition: the outcomes of morphogenesis leading to variations of tooth number, size and shape," *Australian Dental Journal*, vol. 59, pp. 131–142, mar 2014.

[6] A. Balic and I. Thesleff, "Chapter seven - tissue interactions regulating tooth development and renewal," in *Craniofacial Development* (Y. Chai, ed.), vol. 115 of *Current Topics in Developmental Biology*, pp. 157 – 186, Academic Press, 2015.

[7] P. T. Sharpe, "Dental mesenchymal stem cells," *Development*, vol. 143, pp. 2273–2280, jul 2016.

[8] Z.-X. Luo, "Transformation and diversification in early mammal evolution," *Nature*, vol. 450, pp. 1011–1019, dec 2007.

[9] V. Lazzari, C. Charles, P. Tafforeau, M. Vianey-Liaud, J.-P. Aguilar, J.-J. Jaeger, J. Michaux, and L. Viriot, "Mosaic convergence of rodent dentitions," *PLoS ONE*, vol. 3, p. e3607, oct 2008.

[10] J. P. Hunter, D. Guatelli-Steinberg, T. C. Weston, R. Durner, and T. K. Betsinger, "Model of tooth morphogenesis predicts carabelli cusp expression, size, and symmetry in humans," *PLoS ONE*, vol. 5, p. e11844, jul 2010.

[11] V. Lazzari, *Origine et évolution du plan dentaire murin chez les muroida (rodentia, mammalia). Apports de la paléontologie, de la morphologie fonctionnelle et du développemet dans l'étude d'une innovation évolutive.* phdthesis, Université Montpellier II, December 2006.

[12] J. Jernvall, T. Aberg, P. Kettunen, S. Keranen, and I. Thesleff, "The life history of an embryonic signaling center: Bmp4 induces p21 and is associated with apoptosis in the mouse tooth enamel knot," *Development*, vol. 125, no. 2, pp. 161–169, 1998.

[13] I. Salazar-Ciudad, "Tooth morphogenesis in vivo, in vitro, and in silico," *Current Topics in Developmental Biology*, pp. 341–371, 2008.

[14] S. Jia, J. Zhou, Y. Gao, J.-A. Baek, J. F. Martin, Y. Lan, and R. Jiang, "Roles of bmp4 during tooth morphogenesis and sequential tooth formation," *Development*, vol. 140, pp. 423–432, dec 2012.

[15] S. Jia, H.-J. E. Kwon, Y. Lan, J. Zhou, H. Liu, and R. Jiang, "Bmp4-msx1 signaling and osr2 control tooth organogenesis through antagonistic regulation of secreted wnt antagonists," *Developmental Biology*, vol. 420, pp. 110–119, dec 2016.

[16] P. Kettunen, J. Laurikkala, P. Itäranta, S. Vainio, N. Itoh, and I. Thesleff, "Associations of FGF-3 and FGF-10 with signaling networks regulating tooth morphogenesis," *Developmental Dynamics*, vol. 219, no. 3, pp. 322–332, 2000.

[17] J. Li, J. Feng, Y. Liu, T.-V. Ho, W. Grimes, H. A. Ho, S. Park, S. Wang, and Y. Chai, "BMP-SHH signaling network controls epithelial stem cell fate via regulation of its niche in the developing tooth," *Developmental Cell*, vol. 33, pp. 125–135, apr 2015.

[18] I. Salazar-Ciudad and J. Jernvall, "A gene network model accounting for development and evolution of mammalian teeth," *Proceedings of the National Academy of Sciences*, vol. 99, pp. 8116–8120, jun 2002.

[19] W. CA, "Some observations on the tooth band and on the enamel organs of the human deciduous teeth," *Proceedings of the Royal Society of Medicine*, vol. 20, no. 4, pp. 411–426, 1927.

[20] S.-W. Cho, H.-A. Lee, J. Cai, M.-J. Lee, J.-Y. Kim, H. Ohshima, and H.-S. Jung, "The primary enamel knot determines the position of the first buccal cusp in developing mice molars," *Differentiation*, vol. 75, pp. 441–451, jun 2007.

[21] S. Pantalacci, L. Guéguen, C. Petit, A. Lambert, R. Peterkova, and M. Sémon, "Transcriptomic signatures shaped by cell proportions shed light on comparative developmental biology," *Genome Biology*, Dec. 2016.

[22] E. Matalova, G. Antonarakis, P. Sharpe, and A. Tucker, "Cell lineage of primary and secondary enamel knots," *Dev. Dyn.*, vol. 233, no. 3, pp. 754–759, 2005.

[23] J.-L. Fausser, O. Schlepp, D. Aberdam, G. Meneguzzi, J. V. Ruch, and H. Lesot, "Localization of antigens associated with adherens junctions, desmosomes, and hemidesmosomes during murine molar morphogenesis," *Differentiation*, vol. 63, pp. 1–11, apr 1998.

[24] Z. Zhang, Y. Lan, Y. Chai, and R. Jiang, "Antagonistic actions of msx1 and osr2 Pattern mammalian teeth into a single row," *Science*, vol. 323, pp. 1232–1234, feb 2009.

[25] S. Keränen, T. Åberg, P. Kettunen, I. Thesleff, and J. Jernvall, "Association of developmental regulatory genes with the development of different molar tooth shapes in two species of rodents," *Development Genes and Evolution*, vol. 208, pp. 477–486, nov 1998.

[26] A. Gritli-Linde, M. Bei, R. Maas, X. M. Zhang, A. Linde, and A. P. McMahon, "Shh signaling within the dental epithelium is necessary for cell proliferation, growth and polarization," *Development*, vol. 129, pp. 5323–5337, dec 2002.

[27] M. Nakatomi, M. Hovorakova, A. Gritli-Linde, H. J. Blair, K. MacArthur, M. Peterka, H. Lesot, R. Peterkova, V. L. Ruiz-Perez, J. A. Goodship, and H. Peters, "Evc regulates a symmetrical response to shh signaling in molar development," *Journal of Dental Research*, vol. 92, pp. 222–228, jan 2013.

[28] J. A. Whitlock and J. M. Richman, "Biology of tooth replacement in amniotes," *International Journal of Oral Science*, vol. 5, pp. 66–70, jun 2013.

[29] E. Juuri, M. Jussila, K. Seidel, S. Holmes, P. Wu, J. Richman, K. Heikinheimo, C.-M. Chuong, K. Arnold, K. Hochedlinger, O. Klein, F. Michon, and I. Thesleff, "Sox2 marks epithelial competence to generate teeth in mammals and reptiles," *Development*, vol. 140, pp. 1424–1432, mar 2013.

[30] M. Gaete and A. S. Tucker, "Organized emergence of multiple-generations of teeth in snakes is dysregulated by activation of wnt/$\beta$-catenin signalling," *PLoS ONE*, vol. 8, p. e74484, sep 2013.

[31] J. Ahnfelt-Ronne, M. C. Jorgensen, J. Hald, O. D. Madsen, P. Serup, and J. Hecksher-Sorensen, "An improved method for three-dimensional reconstruction of protein expression patterns in intact mouse and chicken embryos and organs," *Journal of Histochemistry and Cytochemistry*, vol. 55, pp. 925–930, may 2007.

[32] T. Yokomizo and E. Dzierzak, "Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos," *Development*, vol. 137, pp. 3651–3661, sep 2010.

[33] S. Schlager, "Soft-tissue reconstruction of the human nose: population differences and sexual dimorphism," *PhD diss., Albert-Ludwigs-University, Freiburg*, 2013.

[34] S. Schlager, "Rvcg: Manipulations of triangular meshes (smoothing, quadric edge collapse decimation, im- and export of various mesh file-formats, cleaning, etc.) based on the vcglib api. r package avail. at: `http://github.com/zarquon42b/Rvcg`."

[35] S. S., "Rvtkstatismo: Integrating statismo and r using the vtkstandardmeshrepresenter. avail. at: `http://github.com/zarquon42b/RvtkStatismo`."

[36] S. Schlager, "mesher: Meshing operations on triangular meshes. avail. at: `https://github.com/zarquon42b/mesheR`."

[37] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," *Nature Biotechnology*, vol. 34, pp. 525–527, apr 2016.

[38] H. Wickham, *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media, 2009.

[39] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biol*, vol. 15, dec 2014.

[40] S. Dray and A.-B. Dufour, "The ade4 package: implementing the duality diagram for ecologists," *Journal of statistical software*, vol. 22, no. 1, pp. 1–20, 2007.

[41] R. Gaujoux and C. Seoighe, "CellMix: a comprehensive toolbox for gene expression deconvolution," *Bioinformatics*, vol. 29, pp. 2211–2212, jul 2013.

[42] Y. Zhong, Y.-W. Wan, K. Pang, L. M. Chow, and Z. Liu, "Digital sorting of complex tissues for cell type-specific gene expression profiles," *BMC Bioinformatics*, vol. 14, no. 1, p. 89, 2013.

[43] D. J. O'Connell, J. W. Ho, T. Mammoto, A. Turbe-Doan, J. T. O'Connell, P. S. Haseley, S. Koo, N. Kamiya, D. E. Ingber, P. J. Park, *et al.*, "A wnt-bmp feedback circuit controls intertissue signaling dynamics in tooth organogenesis," *Science Signaling*, vol. 5, pp. ra4–ra4, jan 2012.

[44] S. Hayashi and A. P. McMahon, "Efficient recombination in diverse tissues by a tamoxifen-inducible form of cre: A tool for temporally regulated gene activation/inactivation in the mouse," *Developmental Biology*, vol. 244, pp. 305–318, apr 2002.

[45] J. Prochazka, M. Prochazkova, W. Du, F. Spoutil, J. Tureckova, R. Hoch, T. Shimogori, R. Sedlacek, J. L. Rubenstein, T. Wittmann, and O. D. Klein, "Migration of founder epithelial cells drives proper molar tooth positioning and morphogenesis," *Developmental Cell*, vol. 35, pp. 713–724, dec 2015.

[46] C. Petit, S. Pantalacci, and M. Sémon, "Tracking temporal signal in developmental transcriptomes.." in prep.

[47] M. Voutilainen, P. H. Lindfors, E. Trela, D. Lönnblad, V. Shirokova, T. Elo, E. Rysti, R. Schmidt-Ullrich, P. Schneider, and M. L. Mikkola, "Ectodysplasin/NF-$\kappa$b promotes mammary cell fate via wnt/$\beta$-catenin pathway," *PLoS Genet*, vol. 11, p. e1005676, nov 2015.

[48] S. Løes, K. Luukko, I. H. Kvinnsland, M. Salminen, and P. Kettunen, "Developmentally regulated expression of netrin-1 and -3in the embryonic mouse molar tooth germ," *Developmental Dynamics*, vol. 227, pp. 573–577, jun 2003.

[49] S. Nucera, D. Biziato, and M. D. Palma, "The interplay between macrophages and angiogenesis in development, tissue injury and regeneration," *The International Journal of Developmental Biology*, vol. 55, no. 4-5, pp. 495–503, 2011.

[50] M. Rothova, J. Feng, P. T. Sharpe, R. Peterkova, and A. S. Tucker, "Contribution of mesoderm to the developing dental papilla," *The International Journal of Developmental Biology*, vol. 55, no. 1, pp. 59–64, 2011.

[51] G. Yuan, L. Zhang, G. Yang, J. Yang, C. Wan, L. Zhang, G. Song, S. Chen, and Z. Chen, "The distribution and ultrastructure of the forming blood capillaries and the effect of apoptosis on vascularization in mouse embryonic molar mesenchyme," *Cell and Tissue Research*, vol. 356, pp. 137–145, jan 2014.

[52] H. R. Dassule, P. Lewis, M. Bei, R. Maas, and A. P. McMahon, "Sonic hedgehog regulates growth and morphogenesis of the tooth.," *Development*, vol. 127, pp. 4775–4785, nov 2000.

[53] A. Ohazama, C. J. Haycraft, M. Seppala, J. Blackburn, S. Ghafoor, M. Cobourne, D. C. Martinelli, C.-M. Fan, R. Peterkova, H. Lesot, B. K. Yoder, and P. T. Sharpe, "Primary cilia regulate shh activity in the control of molar tooth number," *Development*, vol. 136, pp. 897–903, feb 2009.

[54] J. A. Caparros-Martin, M. Valencia, E. Reytor, M. Pacheco, M. Fernandez, A. Perez-Aytes, E. Gean, P. Lapunzina, H. Peters, J. A. Goodship, and V. L. Ruiz-Perez, "The ciliary evc/evc2 complex interacts with smo and controls hedgehog pathway activity in chondrocytes by regulating sufu/gli3 dissociation and gli3 trafficking in primary cilia," *Human Molecular Genetics*, vol. 22, pp. 124–139, oct 2012.

[55] T. Åberg, J. Wozney, and I. Thesleff, "Expression patterns of bone morphogenetic proteins (bmps) in the developing mouse tooth suggest roles in morphogenesis and cell differentiation," *Developmental Dynamics*, vol. 210, pp. 383–396, dec 1997.

[56] A. Ohazama, E. B. Johnson, M. S. Ota, H. J. Choi, T. Porntaveetus, S. Oommen, N. Itoh, K. Eto, A. Gritli-Linde, J. Herz, and P. T. Sharpe, "Lrp4 modulates extracellular integration of cell signaling pathways in development," *PLoS ONE*, vol. 3, p. e4092, dec 2008.

[57] Y. Kassai, P. Munne, Y. Hotta, E. Penttilä, K. Kavanagh, N. Ohbayashi, S. Takada, I. Thesleff, J. Jernvall, and N. Itoh, "Regulation of mammalian tooth cusp patterning by ectodin," *Science*, vol. 309, pp. 2067–2070, sep 2005.

[58] A. Ohazama, T. Porntaveetus, M. S. Ota, J. Herz, and P. T. Sharpe, "Lrp4: A novel modulator of extracellular signaling in craniofacial organogenesis," *American Journal of Medical Genetics Part A*, vol. 152A, pp. 2974–2983, apr 2010.

[59] G. R. Handrigan and J. M. Richman, "A network of wnt, hedgehog and BMP signaling pathways regulates tooth replacement in snakes," *Developmental Biology*, vol. 348, pp. 130–141, dec 2010.

[60] M. Mark, A. Bloch-Zupan, and J. Ruch, "Effects of retinoids on tooth morphogenesis and cytodifferentiations, in vitro.," *International Journal of Developmental Biology*, vol. 36, no. 4, pp. 517–526, 1992.

[61] P. Kettunen and I. Thesleff, "Expression and function of FGFs-4, -8, and -9 suggest functional redundancy and repetitive use as epithelial signals during tooth morphogenesis," *Developmental Dynamics*, vol. 211, pp. 256–268, mar 1998.

[62] B. B. Madison, L. B. McKenna, D. Dolson, D. J. Epstein, and K. H. Kaestner, "FoxF1 and FoxL1 link hedgehog signaling and the control of epithelial proliferation in the developing stomach and intestine," *Journal of Biological Chemistry*, vol. 284, pp. 5936–5944, dec 2008.

[63] D. J. Epstein, L. Martinu, J. L. Michaud, K. M. Losos, C. Fan, and A. L. Joyner, "Members of the bhlh-pas family regulate shh transcription in forebrain regions of the mouse cns," *Development*, vol. 127, no. 21, pp. 4701–4709, 2000.

[64] Y. Chen, M. Bei, I. Woo, I. Satokata, and R. Maas, "Msx1 controls inductive signaling in mammalian tooth morphogenesis," *Development*, vol. 122, no. 10, pp. 3035–3044, 1996.

[65] T. Åberg, X.-P. Wang, J.-H. Kim, T. Yamashiro, M. Bei, R. Rice, H.-M. Ryoo, and I. Thesleff, "Runx2 mediates FGF signaling from epithelium to mesenchyme during tooth morphogenesis," *Developmental Biology*, vol. 270, pp. 76–93, jun 2004.

[66] T. Porntaveetus, Y. Otsuka-Tanaka, M. A. Basson, A. M. Moon, P. T. Sharpe, and A. Ohazama, "Expression of fibroblast growth factors (fgfs) in murine tooth development," *Journal of Anatomy*, vol. 218, pp. 534–543, feb 2011.

[67] J. Pispa and I. Thesleff, "Mechanisms of ectodermal organogenesis," *Developmental Biology*, vol. 262, pp. 195–205, oct 2003.

[68] M.-H. Chen, C. W. Wilson, and P.-T. Chuang, "SnapShot: Hedgehog signaling pathway," *Cell*, vol. 130, pp. 386–386.e2, jul 2007.

[69] S. Fraboulet, K. Kavvadia, O. Pourquié, P. T. Sharpe, and T. A. Mitsiadis, "BEN/DM-GRASP/SC1 expression during mouse facial development: differential expression and regulation in molars and incisors," *Gene Expression Patterns*, vol. 3, pp. 255–259, jun 2003.

[70] N. Yoshiba, K. Yoshiba, C. Stoetzel, F. Perrin-Schmitt, Y. Cam, J. V. Ruch, and H. Lesot, "Temporospatial gene expression and protein localization of matrix metalloproteinases and their inhibitors during mouse molar tooth development," *Developmental Dynamics*, vol. 228, pp. 105–112, aug 2003.

[71] X. Zhao, Z. Zhang, Y. Song, X. Zhang, Y. Zhang, Y. Hu, S. H. Fromm, and Y. Chen, "Transgenically ectopic expression of bmp4 to the msx1 mutant dental mesenchyme restores downstream gene expression but represses shh and bmp2 in the enamel knot of wild type tooth germ," *Mechanisms of Development*, vol. 99, pp. 29–38, dec 2000.

[72] L. A. Lyons, C. A. Erdman, R. A. Grahn, M. J. Hamilton, M. J. Carter, C. R. Helps, H. Alhaddad, and B. Gandolfi, "Aristaless-like homeobox protein 1 (ALX1) variant associated with craniofacial structure and frontonasal dysplasia in burmese cats," *Developmental Biology*, vol. 409, pp. 451–458, jan 2016.

[73] C. A. Ettensohn, "Alx1, a member of the cart1/alx3/alx4 subfamily of paired-class homeodomain proteins, is an essential component of the gene network controlling skeletogenic fate specification in the sea urchin embryo," *Development*, vol. 130, pp. 2917–2928, jul 2003.

[74] Y. Kong, M. Grimaldi, E. Curtin, M. Dougherty, C. Kaufman, R. M. White, L. I. Zon, and E. C. Liao, "Neural crest development and craniofacial morphogenesis is coordinated by nitric oxide and histone acetylation," *Chemistry & Biology*, vol. 21, pp. 488–501, apr 2014.

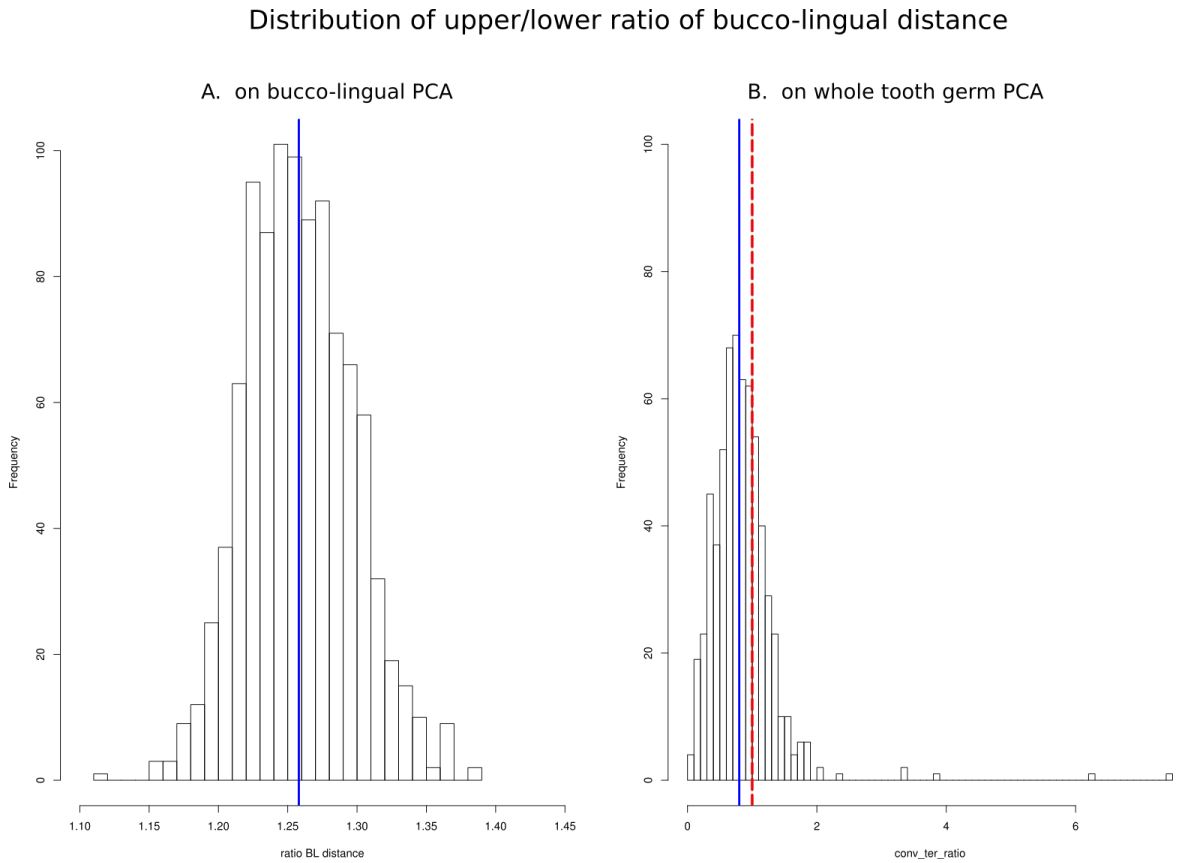
[75] C. A. Loomis, E. Harris, J. Michaud, W. Wurst, M. Hanks, and A. L. Joyner, "The mouse engrailed-1 gene and ventral limb patterning," *Nature*, vol. 382, pp. 360–363, jul 1996.

## Supplementary material

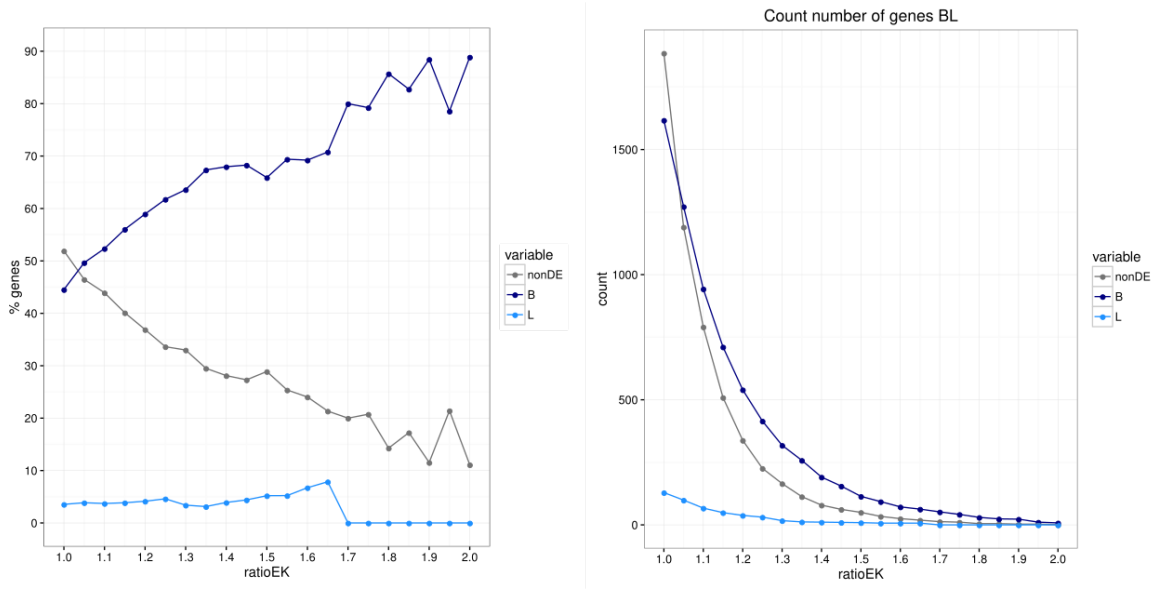
**Sup. Text S1: Rule out statistical artifacts**

The excess of bucco-lingual differences in upper molar could be caused by statistical artifacts. We tested the following artifacts: the library size and the variability among samples. We tested the influence of libraries size because upper libraries (total 244M) tend to be bigger than lower libraries (total 144M). To test the influence of libraries size on the number of genes Differentially Expressed and on the bucco-lingual distance on the PCA, we re-run the entire analysis on 20M reads randomly chosen for each libraries. We repeat the process twice. We still found 2 times more genes DE and an excess of BL distance in upper molar.

Looking at the bucco-lingual PCA we saw that the the lingual libraries are more variable / dispersed and the upper libraries are more alike. The differences in dispersion directly impact the statistical power for detecting genes DE. We artificially increased the dispersion in upper libraries using dispersion law of lower molars (from DESeq2 analysis) or we artificially decreased dispersion in lower library using dispersion law of upper molar (from DESeq2 analysis), in both cases we still found more genes DE in upper molars. Applying dispersion of lower libraries to upper libraries we still found more genes DE in upper molar (1.62-time more). Reciprocally, applying dispersion of upper libraries to lower libraries, we still found more genes DE in upper molar (1.25-time more).




**Supplementary Figure A.** Ratio of distance between buccal and lingual samples in upper and lower molars. A. Bucco-lingual distance measured on the first axis of the bucco-lingual PCA (on 1000 genes chosen randomly, analysis repeated 1000 times). The distribution follow a normal law of mean equal 1.258 (blue line) and standard deviation of 0.038 (Kolmogorov-Smirnov, *pvalue*<2.2e-16). Thus the bucco-lingual distance is significantly bigger in upper molar around 30%. B. Bucco-lingual distance measured on the first axis of the whole tooth germ PCA on 1000 genes randomly chosen repeated 1000 times. The distribution is significantly shifted toward 0.8 (blue line) (Wilcoxon test compared to a normal distribution centered around 1, *pvalue*=2.863e-08 ). The bucco-lingual distance is not bigger in upper molar (when BL libraries are projected to the whole tooth germ PCA). The excess of bucco-lingual differences in upper molar is not caused by timing differences.




**Supplementary Figure B.** (Left). Number of bucco-lingual (BL) genes: buccal genes (darkblue), lingual genes (light blue) and non DE genes (grey) for different value of $EKratio$, a ratio of expression in the EK at ED14.5 over expression in the epithelium and the mesenchyme at ED13/13.5, from microarray data [43]. (Right). Cumulative number of BL genes along the $EKratio$ space. Theses data enabled us to determine $EKratio$ threshold for which the specificity and the sensitivity toward the definition of new enamel knot genes is enough. That is, new enamel knot genes are genes expressed with a strong buccal bias (enrichment >2) and an $EKratio$ large enough to assure the belonging of the gene to the EK and not another putative structure of the epithelium.

# 11 *Bmper* : a candidate for the bucco-lingual asymmetry of tooth development

## Introduction

The molars are serial organs, as such they develop through iteration of a similar gene network deployed at different positions in the body. In the case of upper and lower molar they deployed in the upper and the lower jaws. As already discussed in the introductory part of this manuscript (see chapter 3, *p. 41*), the upper and the lower jaw identities are defined around ED8.25-ED9 by homeobox genes. Such as *Dlx5/6* are essential in the first pharyngeal arch to determine the identity of the lower jaw [Depew, 2002; Beverdam *et al.* , 2002]. Afterwards, the odontogenic homeobox code is known to regulate the type of the tooth [Qiu *et al.* , 1997; Thomas *et al.* , 1997; Tucker, 1998; Cobourne & Sharpe, 2003; Catón & Tucker, 2009; Miletich *et al.* , 2005].

An interesting feature about mouse first molars is that they present extreme divergent morphologies between the upper and the lower jaw. Indeed, the mouse first upper molar develops a supplementary lingual row with two supplementary cusps on a it, *ie.* n=8 cusps distributed in 3 rows vs. 6 cusps distributed in 2 rows for the lower molar. How to increase the number of cusps patterned during development? Experiments of tissue recombination show that cusp patterning is tightly linked to mesenchyme identity and proportion [Schmitt *et al.* , 1999; Hu *et al.* , 2006; Cai *et al.* , 2007; Ishida *et al.* , 2011]. Gene-centered studies have already identified some genes and pathways responsible for the patterning of supplementary cusps [Tucker *et al.* , 2004; Kassai *et al.* , 2005; Charles *et al.* , 2009a; Harjunmaa *et al.* , 2012] (see chapter 3, *p. 52* ). However, most of these gene-centered studies were done in mouse lower molars only and were not interested in identifying the developmental bases responsible for the increase number of cusps in mouse upper molar.

In this part we adopted a gene centered approach in order to identify genes responsible for the development of the two supplementary lingual cusps in mouse upper molar. To do so, we need to identify genes that are differentially expressed (DE) in the mouse upper molar. However, the

differences between mouse upper and lower developmental program are two fold : (i) there are the intrinsic differences between upper and lower developmental program and (ii) they are the differences that allow the mouse upper molar to form these two supplementary cusp. On the contrary, the hamster has the same number of cusps on its first upper and lower molars. Thus we used the hamster as a control species to identify the genes responsible for an expected upper-lower differences in a case where the same number of cusps are formed. Nevertheless, hamster and mouse molars have a certain amount of specific features, such as the presence of enamel free zone in mouse molars that are not present in hamster. We thus needed to remove the genes responsible for these features, for example we can do so by identifying genes differentially expressed between mouse and hamster lower molars and removed these genes from the one identified to be differentially expressed between mouse and hamster upper molars. Moreover, as the supplementary cusps of the upper molar are patterned lingually and we have shown that there is an early specification of territory that is linked to patterning of cusps that are even added lately, good candidate genes would also have an early bucco-lingual asymmetric bias.

## Results

### *Bmper* is differentially expressed between upper and lower molars of mouse but not of hamster

The analysis of genes differentially expressed (DE) between the upper and the lower molar at each stages (for duplicates), enriched at least 1.5-times in the upper molar germs at one stage, lead us to identify 166 genes that are biased toward the upper molar in the mouse and that are not biased in the hamster. Among the genes differentially epressed between the mouse upper and lower molar (formula : $\sim stage + toothType$), they are 7 genes that have an enrichement bigger than 2 : *Bmper, Cldn11, Foxp2, Gap43, Pcdh20, Pde8b,*, and *Rtn4rl1*, and except for *Gap43* they are not enriched in hasmter molars. We focused our attention on *Bmper* specifically because *Bmper* is known to interact with *Bmp4* in an other biological system [Kelley *et al.* , 2009].

*Bmper* is significantly enriched in mouse upper molar all along cuspidogenesis (FDR<9.6e-12) except at stage ED17.5. On the contrary, *Bmper* is not significantly enriched in hamster upper molar except at stage ED13 (FDR=0.002). Figure 11.1 shows the value of the enrichment of *Bmper* toward the upper molar in both mouse and hamster molars. In mouse, we noted that the upper molar is enriched in *Bmper* more than two fold between stages ED15 to ED16.5, then the enrichment level drops back to the level it was at cap transition (*ie.* 1.6, red line on figure 11.1), except at stage ED17.5 when *Bmper* seems no longer to be biased toward the upper molar, which do not seems to be particularly related to a bigger variation between replicates at these stage (data not shown). In hamster *Bmper* is enriched 1.44-times in upper molar at stage ED13 then its no longer enriched during cuspidogenesis.

Figure 11.2 shows the expression profile of *Bmper* in mouse (for both replicates) and hamster

molar germs (for only one replicate) during cuspidogenesis. *Bmper* expression decreases all along cuspidogenesis in both species, until it reaches a plateau : at ED13 in hamster and at ED16 for mouse lower molar and around ED17-17.5 for mouse supper molar. Interestingly, the inflection of the curve corresponds to the end of the asymmetrical growth of the germ in the bucco-lingual axis. This period is extended in mouse upper molar in relation with patterning of the supplementary lingual cusps on a third row, as explained in the introductory part of the manuscript (*p.* 3.1.3.2 and figure 3.4, *p.* 40). In mouse lower molar and in hamster lower and upper molars, there are only two rows of cusps, one buccal and one lingual. The first row forms together with the first cusp in the buccal side, the second row forms just afterward with the patterning of the second cusp in the lingual side (from ED15.5 to ED16 in mouse and from ED13 to ED13.25 in hamster, see chapter 9, figure 2). However, in the mouse upper molar, there is the formation of a second lingual row, with the formation at the end of cuspidogenesis of two supplementary cusps from stage ED17.5 (*cf.* chapter 9, figure 2).

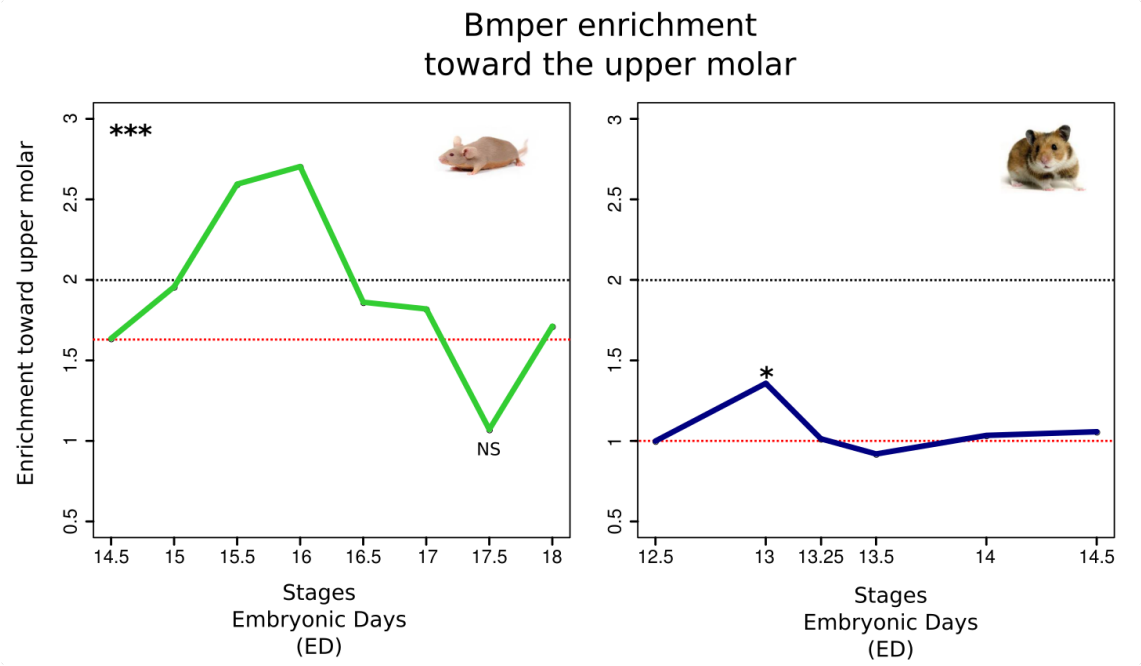



FIGURE 11.1 – **Enrichment of *Bmper* in upper molars of mouse and hamster**
In mouse, *Bmper* is significantly enriched toward the upper molar (FDR<9.6e-12) except at stage ED17.5. In hamster, *Bmper* is not differentially expressed in the upper molar, except at stage ED13 (FDR=0.002). The dashed red line represent the enrichment at stage cap transition (ED14.5 for mouse and ED12.5 for hamster). NS : statistically non-significant.

### *Bmper* is biased buccally during mouse molar development

In addition, the analysis of buccal and lingual side transcriptomes of mouse upper and lower molar germs at ED15 (*cf.* chapter 10) identified *Bmper* as being strongly enriched toward the

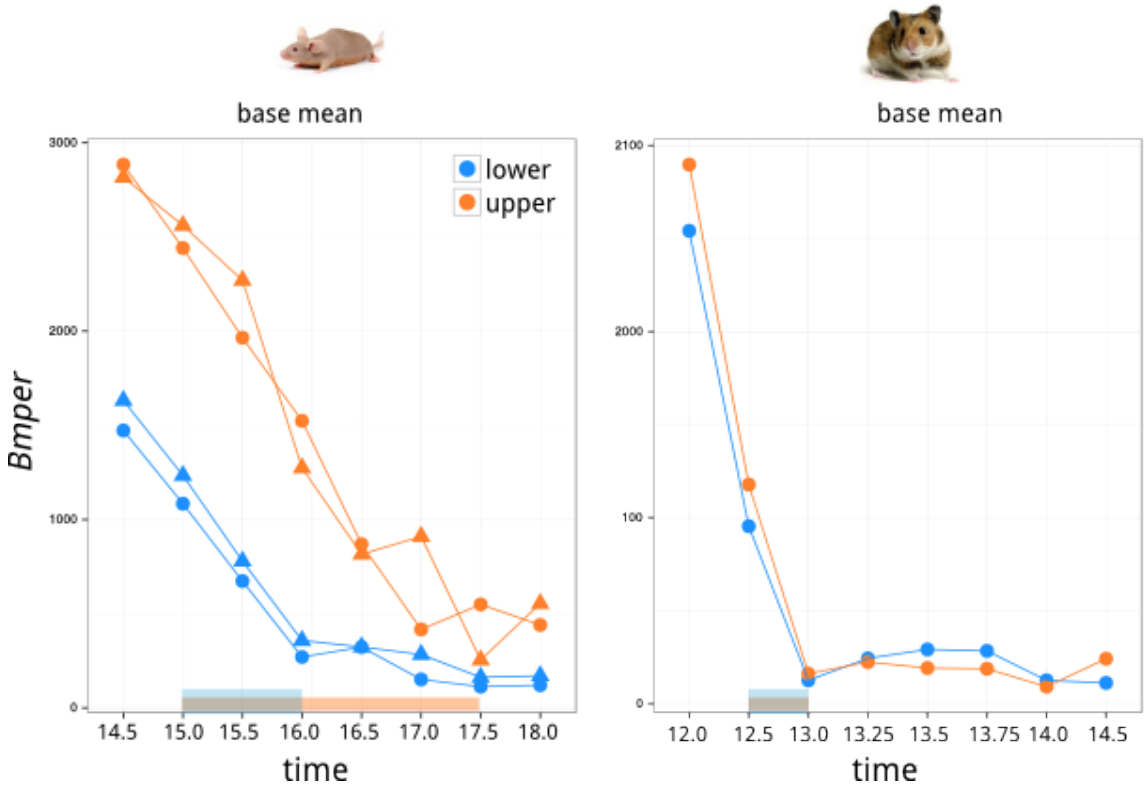



FIGURE 11.2 – ***Bmper* expression profile in mouse and hamster**
Expression of *Bmper* during upper and lower molar crown morphogenesis in mouse and hamster. The highlighted stages on the x-axis represents the period of asymmetrical bucco-lingual growth of the tooth germ.

buccal side. This BL bias is a supplementary argument to consider *Bmper* as a candidate gene. We noted that the buccal bias of *Bmper* is similar in upper and lower molar (3-times more in the buccal side, FDR<7.7e-32 ), however expression levels are higher in upper molar, on average it is 3.8-times more expressed in the upper molar (3.4-times more on the lingual side and 4.2-times more on the buccal side). Moreover, from our dataset of pure tissues transcriptomes (at stage ED15), we could more precisely pinpoint *Bmper* in the mesenchyme. Even though there seems to be non negligible residual expression in epithelial tissue of mouse molar germs (on average base mean = 620 ), *Bmper* expression is strongly biased toward the mesenchymal compartment (enrichment>10, FDR<6.9e-42).

### *Bmper* is known to interact with *Bmp4*

*Bmper*, for bone morpho-genetic protein [BMP]-binding endothelial cell precursor-derived regulator, is a secreted protein that was shown to directly interacts with BMP2/4/6 [Moser *et al.* , 2003; Kelley *et al.* , 2009]. *Bmper* was proposed to have a pro-BMP activity [Moser *et al.* , 2003; Ikeya *et al.* , 2006; Rentzsch *et al.* , 2006], despite it was sometimes show to inhibit BMP

activity [Ambrosio *et al.* , 2008]. In fact, it was shown that regulation of the BMP pathway by *Bmper* is a bit more complex; its action, -activation or inhibition-, is dependent upon *Bmper* concentration (where high levels of *Bmper* inhibit BMP signaling and low levels activate it) and BMPs concentration [Serpe *et al.* , 2008; Kelley *et al.* , 2009]. However, *Bmper* expression has never been documented in molar development

As mentioned above, *Bmper* (aka *Cv2*) is known to interact with *Bmp4* in other biological systems (*e.g.* the *Drosophila* [Serpe *et al.* , 2008], the *Xenopus* [Moser *et al.* , 2003] and the mouse in kidneys [Ikeya *et al.* , 2006] and lungs [Kelley *et al.* , 2009]). As already presented in the introductory part of this manuscript (*cf.* section 3.3.2, *p.* 46), *Bmp4*, and more broadly the BMP pathway, is one the key regulator of the mouse molar development. Indeed, it plays an important role during the cap transition [Chen *et al.* , 1996; Bei *et al.* , 2000; Jia *et al.* , 2012, 2016] and the maintenance of cervical loops [Li *et al.* , 2015].

### The characterization of the spatial pattern of expression of *Bmper* during mouse molar development

We then wanted to characterize *Bmper* expression pattern in molar germs during cuspidogenesis. To do so, we began to perform whole mount *in situ* hybridizations (WISH) of mouse molar germs. Thus, at stage ED15.5 we found that *Bmper* is expressed in the mesenchyme with a buccal bias. We noted that *Bmper* is strictly restricted toward the buccal side whereas it extends more lingually in upper molar (figure 11.3). As there are not differences in the buccal bias of upper and lower molars at stage ED15, this pattern of expression suggest that *Bmper* could could gradually retreat towards the buccal side.

This stronger asymmetric expression of *Bmper* is interesting since the tooth seems to grow faster on the lingual side than on the side (*cf.* chapter 10) and this is especially true for the mouse first upper molar which patterned 2 supplementary cusps on its lingual side. We also noted that *Bmper* is expressed in the mesenchyme surrounding the tip of both cervical loops. Together this pattern of expression pinpoints the asymmetric outcome of the *Bmper* action in the buccal and the lingual side of the mouse molar germs.

### The analyses of *bmper* mutant mice revealed a potential role in mouse molar development

In order to investigate the role of *Bmper* in mouse molar morphogenesis, we decided to study the phenotype of molar germs lacking *bmper*. In such mutants we expected to see a developmental defect in the bucco-lingual axis. *Bmper* homozygous mutant mice die at birth, due to respiratory failures caused by generalized skeletal abnormalities and more specifically caused by the extremely reduced lumen of the trachea [Ikeya *et al.* , 2006; Zakin *et al.* , 2008; Kelley *et al.* , 2009]. Among the four knock out mutant available, we worked with the $bmper^{tm1Cpat}$ null allele

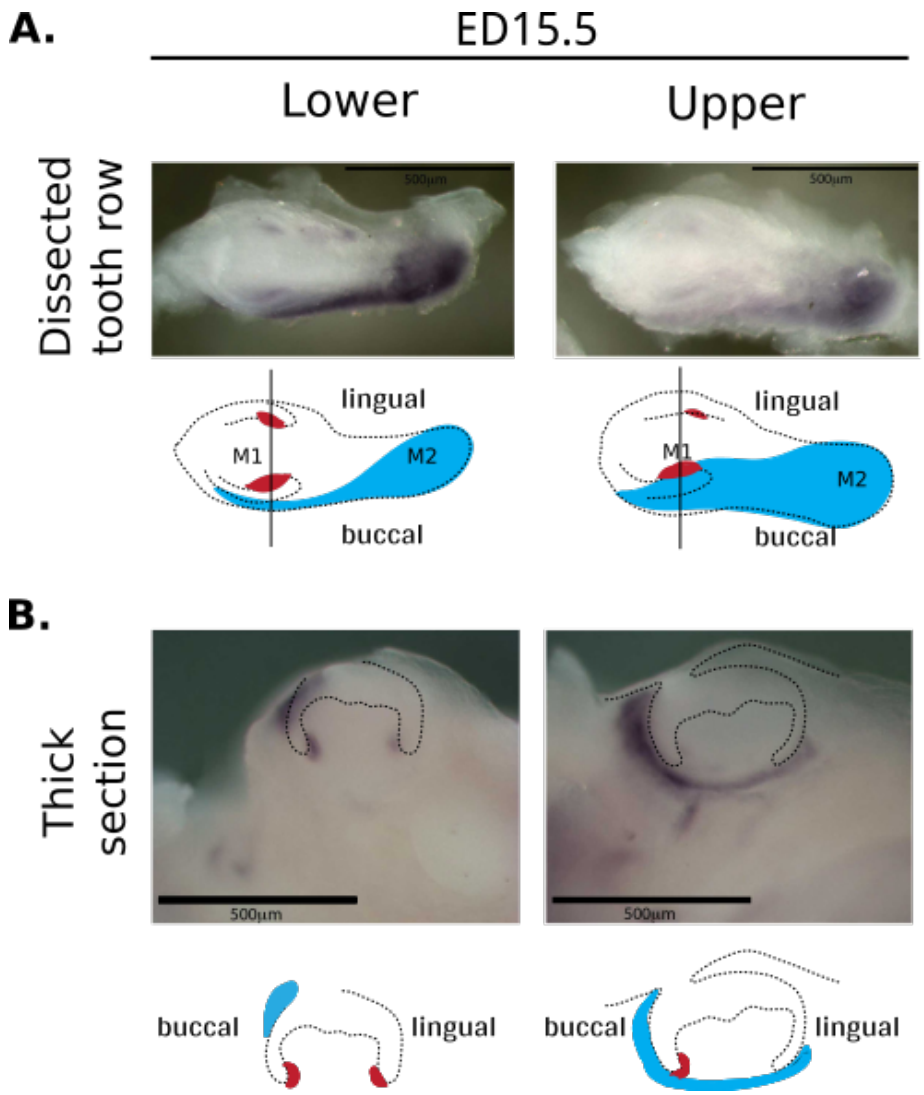



FIGURE 11.3 – ***Bmper* expression in mouse molars at ED15.5**.
**A. Dissected tooth row :** Whole mount *in situ* hybridization (WISH) on dissected upper and lower mouse molar germs at stage ED15.5. Represented below is the schematic pattern of expression obtained from WISH data on *Bmper*. The vertical lines on the schematics represent the section plan used below. **B. Thick sections :** WISH on thick sections of mouse first molar germs at ED15.5 together with the schematic representation of the area of *Bmper* expression. Note that *Bmper* is expressed in the mesenchyme of the first molar (M1) with a buccal bias in both the upper and the lower molar (represented in blue). However, *Bmper* expression is strictly buccal in lower molar whereas in upper molar is more widely distributed toward the lingual side. Moreover, *Bmper* is expressed in both cervical loops in lower molar and in the buccal cervical loop in upper molar (represented in red). Finally, *Bmper* is expressed in the forming second molar (M2) The dashed line outline the epithelium.

(in collaboration with Cam Patterson lab for the mice supply).

As molars erupt some days to weeks after birth (depending on the strain consider), we could not assess the molar morphology of mature forms of *Bmper* homozygous mutant mice. We thus decided to analyze the mature forms of *Bmper* heterozygous mutant and compare it to the wild type. However, no striking differences were noted between $bmper^{+/-}$ mice and $bmper^{+/+}$ mice.

We also looked at molar germs at ED18.5, when all cusps are patterned, in mice lacking *bmper* (both heterozygous and homozygous) and compared it to the ED18.5 wild type molar germs. Preliminary results on histological sections suggested that *Bmper* affect the formation of the supplementary cusps in mouse upper molars (figure 11.4, analyses of 5 $bmper^{-/-}$ / 3 $bmper^{+/+}$ and 1 $bmper^{+/-}$). Indeed, in $bmper^{-/-}$ molar germs, the supplementary cusps seemed smaller and sometimes even absent. More precisely, cusp number 8 (aka cusp L1 in Gaunt [1955] nomenclature, figure 11.4), which is the last cusp to form (*cf.* figure 3.9, *p.* 57) seemed to be absent from homozygous mutant $bmper^{-/-}$ molar germs and cusp 7 (aka cusp L2), which is the penultimate cusp to form, is shallower. Together, these data could indicate that the mouse upper has not the time to pattern the correctly the last to cusps to form (L1 and L2) or it could be that the germ is smaller than in the wild type, thus there is not enough space to pattern correctly the last two cusps. However, the results on sections were hard to analyzed and we could not clearly conclude on the action of *Bmper* during crown morphogenesis. That is why we plan to use X-ray tomography (see below discussion section of this chapter). Furthermore, we could not exclude that this phenotype was caused by a simple delay of development, or even that due to random sampling we had only delayed embryos, but we think it is unlikely because the rest of the tooth does not seem delayed (*e.g.* the cervical loops seem as much if not more advanced than in the wild type). Anyhow, this ambiguity would be removed by analyzing later stages.

## Discussion on the progress of the project

In conclusion to this project, *Bmper* seems a fairly interesting candidate on the formation of two supplementary lingual cusps in the mouse upper molar, regarding that it is more expressed in the mouse upper molar (and not in hamster), it is expressed with a buccal bias and it is known to regulate BMP activity. However, this project is at a very preliminary state. First, the characterization of *Bmper* expression pattern by WISH came to an halt after I tried to perform histological sections embedded in paraffin, because *in situ* signal was washed away by alcohol during the embedding process. To overcame this problem we will need to develop a technique to perform *in situ* hybridizations in cryostat sections. Second, as mentioned in the results, the characterization of the role played by *Bmper* during crown morphogenesis was not very easy on histological sections, due notably to orientation of the section and section plan, we decided to overcome this problem by using microtomograpghy X-ray imaging to have 3D reconstructions of the mutant molar germs. However, this part of the project was delayed because our collaborators moved the mice to another facility and we were not able to have access to new samples yet.

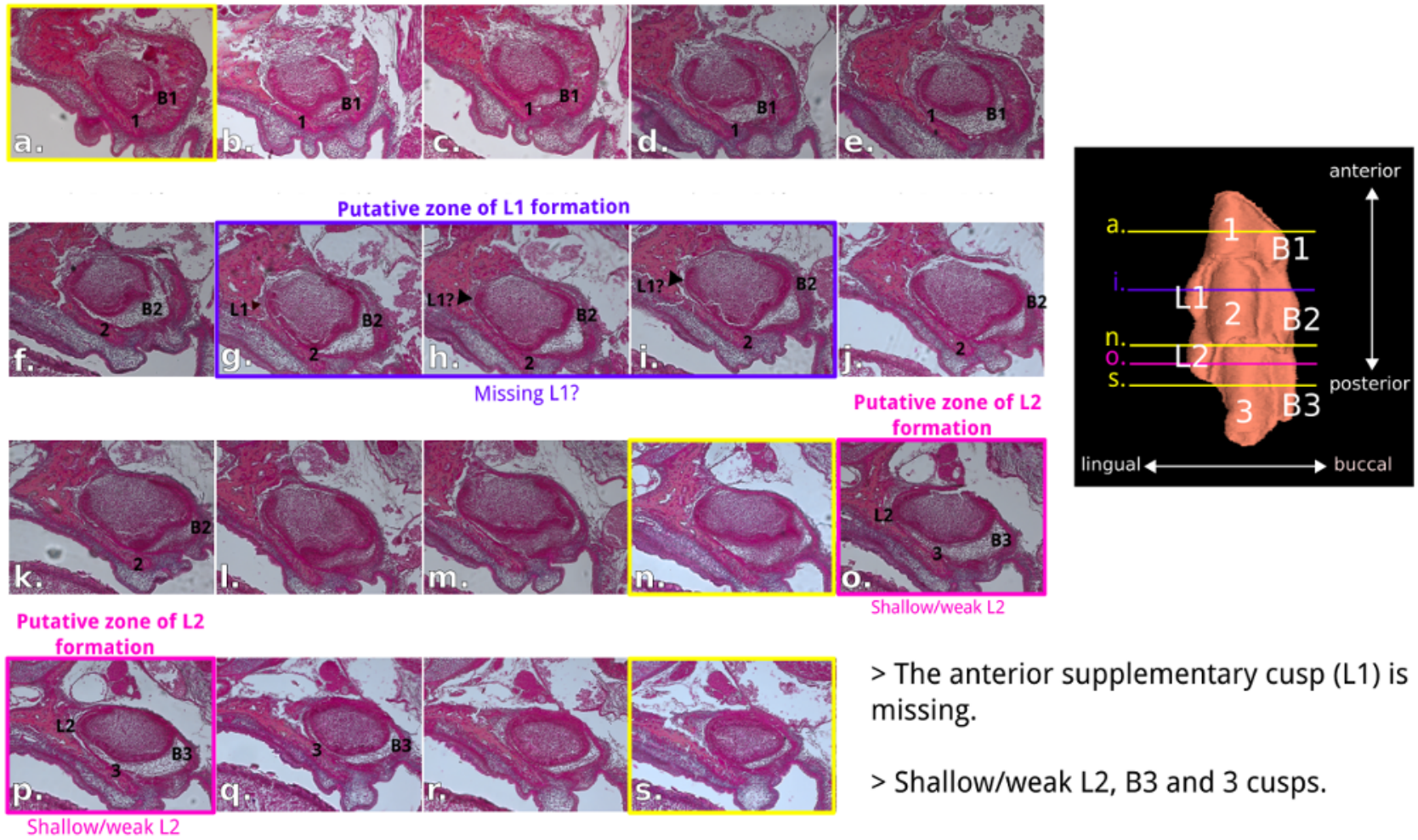



FIGURE 11.4 – **Molar morphology of *bmper* homozygous mutant mice at staged ED18.5**. Histological sections of *Bmper*$^{-/-}$ upper molar germ at stage ED18.5 every 35-40$\mu$m. On the right is a top view of 3D reconstructed epithelium, where the names of the cusps (according to Gaunt [1955] nomenclature) are described and some section plan are highlighted corresponding to section a,i, n, o and s. L1 and L2 are the supplementary cusps, L1 (the last cusp to form) seems to be absent while L2 (the penultimate cusp to form) seems shallower than expected.

# Discussion Part IV

# 12 Discussion

## 12.1 Conclusion

Les travaux de cette thèse ont porté sur la comparaison des programmes de développement à l'aide de transcriptomes. Notre modèle d'étude a été un organe sériel :la molaire, chez deux espèces de rongeurs, la souris et le hamster. Nous nous sommes demandés comment les différences morphologiques des premières molaires inférieures et supérieures se reflètent dans les programmes de développement entre espèces et entre types de molaire (inférieure vs. supérieure). Ainsi, nous pouvons découpé notre approche en deux grandes parties qui seront les parties structureront cette discussion. (1) Tout d'abord nous nous sommes intéressés aux signatures transcriptomiques présentes dans ces transcriptomes, ce qui nous a permis de mettre en évidence une signature temporelle majoritaire ainsi d'une signature identitaire des molaires. De plus, une analyse transcriptomique de molaires coupées en deux nous a permis de caractériser les signatures transcriptomiques présentes dans un demi-organe (côtés buccal et lingual). (2) Dans un deuxième temps, nous avons étudié l'évolution des programmes de développement entre la hamster (dont les molaires inférieures et supérieures mettent en place le même nombre de cuspides au cours du développement), ainsi que la molaire de souris dont la molaire supérieure met en place deux cuspides linguales supplémentaires. Couplé à notre étude sur les transcriptomes buccaux et linguaux des molaires de souris nous avons pu proposer des pistes pour comprendre comment les molaires, qui sont des dents qui ne se remplacent pas, ont pu évoluer à partir de dents à remplacement continu, type reptiles.

Nos travaux montrent l'intérêt des études transcriptomiques comparées du développement pour comprendre l'évolution du développement et comme outils précieux pour établir la carte génotype-phénotype. Je pense que cet outil, pour lequel nous avons développé une méthodologie d'analyse qui est présentée dans ces travaux, sera fortement utilisé dans les études d'évo-dévo afin de répondre à des questions d'évolution des programmes de développement entre espèces. Mais aussi, l'étude comparée des transcriptomes au cours du développement va apporter un

éclairage nouveau sur les contraintes qui ont lieu au cours du développement.

## 12.2 Approche transcriptomique comparée de l'étude des programmes de développement

Au cours de ces travaux, nous avons expérimenté une nouvelle façon de comparer des transcriptomes, que ce soit entre stades de développement (chapitre 8 et annexe A), entre différentes parties de la dent (chapitre 10 ou encore entre espèces (chapitre 9). Pour cela, nous nous sommes basés sur des comparaisons globales de transcriptomes, notamment par des analyses multivariées. De là, nous nous sommes aperçus que (i) les transcriptomes nous fournissaient des informations temporelles que nous pouvions utiliser pour classer les échantillons selon leurs stades de développement ; (ii) les différences de composition tissulaire, notamment l'abondance relative des types cellulaires, influençaient majoritairement les transcriptomes, ce qui nous a permis d'extraire des informations intéressantes sur les populations de cellules en présence ainsi que leur dynamique au cours du développement. De plus, après un traitement approprié des données, il nous a été possible de comparer les niveaux d'expression entre espèces, où nous avons pu exploiter un signal temporel commun qui s'est révélé très utile pour obtenir des stades les plus homologues possibles entre espèces. Pour finir, j'ai appliqué ce nouveau cadre méthodologique et conceptuel à la comparaison des transcriptomes d'une dent coupée en une partie buccale et une partie linguale. L'analyse de ces transcriptomes selon l'angle des populations cellulaires doublée d'une analyse plus gène spécifique a révélé des différences extrêmement marquées entre partie buccale et partie linguale. Combinée à une analyse de la littérature assez disparate sur cet aspect, notre travail soulève des hypothèses inattendues sur le développement bucco-lingual de la dent.

### 12.2.1 Les difficultés de la comparaison des niveaux d'expression entre espèces

La transcriptomique comparée du développement permet de questionner l'évolution des programmes développementaux entre espèces. Cependant, la comparaison de l'évolution d'expression entre espèce comporte encore des défis techniques [Roux *et al.* , 2015]. Plus particulièrement, deux obstacles majeurs rendent cette analyse difficile. Tout d'abord, il faut être capable d'estimer des niveaux d'expressions comparables entre espèces. En effet, toutes analyses comparatives de données d'expression se heurtent à des biais de méthodes pour l'estimation de niveaux d'expression. Or, si les niveaux d'expression ne sont pas estimés correctement et/ou avec les mêmes biais de méthode pour toutes les espèces traités, la comparaison des ces niveaux d'expression entre espèce n'aura pas de sens biologique. Pour estimer des niveaux d'expression, il faut connaitre les séquences des transcrits ou des gènes de références. Or, tous les génomes/-transcriptomes disponibles n'ont pas le même qualité d'annotation. De plus, dans le cas des espèces non modèles, pour lesquels il n'y a pas de données génomiques, aucune séquence de

référence n'est disponible. Même dans les cas où des séquences de références sont disponibles ou reconstruites par assemblage *de novo*, la difficulté est d'identifier les séquences qui sont orthologues entre espèces afin de de s'assurer que l'on compare bien le niveau d'expression des séquences orthologues. C'est pourquoi Carine Rey, une étudiante en thèse du laboratoire, a développé une méthode nommé Amalgam qui permet dans un premier temps de reconstruire les séquences de référence pour chaque espèce (avec le module apytram). Puis dans un deuxième, les positions phylogénétiques des séquences de références de toutes les espèces comparées sont vérifiées. Puis, comme préconisé par Brawand *et al.* [2011], seulement les portions orthologues des séquences de références sont conservées, ce qui implique que la séquence de référence a la même taille dans toutes les espèces traitées. Enfin, les lectures RNA-seq sont alignées puis comptées sur ces séquences de référence. De la sorte, les niveaux d'expression qui sont estimés pour chaque espèce sont le plus comparables possibles, et si cette estimation est biaisée par la méthode on s'attend à ce que ce biais soit le même pour la totalité des espèces traitées. C'est cette méthode que nous avons utilisé dans le chapitre 9 afin de comparer les programmes de développement des molaires entre souris et hamster.

D'autre part, afin de comparé le développement de deux espèces ou plus, il faut être capable de définir des stades de développement homologues qui pourront être comparés. Cette partie sera traitée plus bas (paragraphe 12.2.2.2) car la méthode que nous avons proposé se base sur une signature transcriptomique que notre analyse a permis de mettre en évidence aux niveaux des transcriptomes d'organes au cours du développement.

### 12.2.2 Caractérisation des signatures transcriptomiques au cours du développement des molaires inférieures et supérieures

#### 12.2.2.1 Le signal temporel est une signature transcriptomique majoritaire

Une des signatures présente dans les germes de molaire au cours du développement est un signal temporel qui ordonne les transcriptomes de molaires selon leur stade de développement (voir chapitre 8). Cette signature transcriptomique avait déjà été mise en évidence à l'échelle des embryons en entier et pour des temps de développement beaucoup plus grands, couvrant la totalité de l'embryogenèse [O'Connell *et al.* , 2012; Levin *et al.* , 2012; Anavy *et al.* , 2014; Tschopp *et al.* , 2014]. Notre étude a montré que (1) cet ordonnancement temporel est aussi vrai dans des organes;(2) ce signal temporel est présent à des échelles de temps très fines allant de la demie journées et même des variations de quelques heures de l'état d'avancement du développement entre réplicats biologiques. Enfin nous avons aussi montré que ce signal temporel est présent dans des molaires coupées en deux. La nouveauté de notre étude a été la caractérisation de ce signal temporelle. En effet, nous avons montré que ce signal est porté par des changements subtils d'expression entre stades consécutifs, notamment pour les gènes dont l'expression croit/décroit linéairement au cours du temps. Comme mentionné plus haut, ce signal temporel est très robuste, on le retrouve à différentes échelles de temps : (1) d'une large

échelle temporelle, la morphogenèse de la couronne, (2) à une très fine échelle temporelle, entre réplicats biologiques, (3) ainsi qu'à l'échelle du demi organe, partie buccal et linguale du germe (voir chapitre 8) et (4) à l'échelle tissulaire, l'épithélium et le mésenchyme composant le germe (données non présentées). Nous avons aussi mis en évidence que le signal temporel dans les germes de molaire en développement est causé par des différences de proportions de tissus et des différences de maturation des tissus composant le germe : l'épithélium, le tissu de cupides et le mésenchyme (voir annexe A). Notamment nous avons proposé un modèle qui permet de récapituler des différences entre la molaire supérieur et inférieure de souris simplement par des différences de proportion et de maturation du tissus de cupide entre les deux germes. Bien que ce modèle soit un modèle simpliste il permet de montrer que le facteur proportion de tissu, qui est souvent peu considéré, participe de façon non négligeable à l'identité de l'organe (figure 12.1).

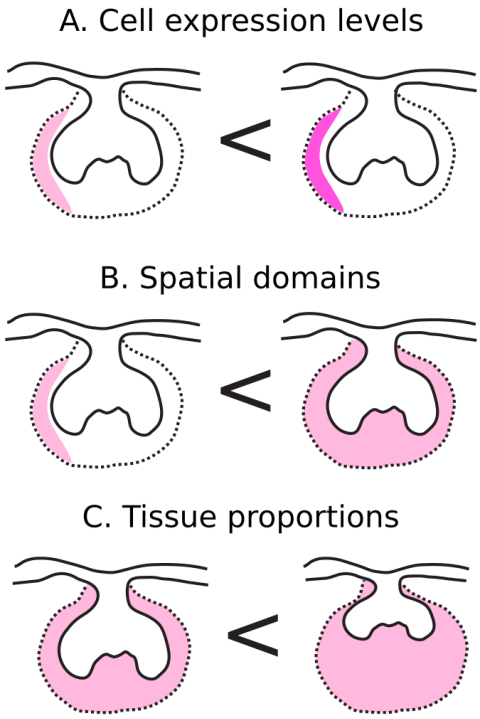



FIGURE 12.1 – **Comment les différences de niveau d'expression se traduisent-elles dans un organe ?**
Les différences d'expression entre deux organes (ici les germes de molaires) peuvent résulter de différentes situations schématiquement représentées ici. Prenons un gène hypothétique dont l'expression est représentée en rose. Dans la situation (**A.**) la même quantité de cellules exprime le transcrit, la différence résulte de la quantité de transcrits exprimée par ces cellules. Dans la situation (**B.**) le niveau d'expression par cellule reste constant, cependant il y a plus de cellules qui expriment le transcrit. Enfin, dans la situation (**C.**) le patron d'expression est similaire et les cellules expriment la même quantité de transcrit, cependant il a a proportionnellement plus de cellules mésenchymateuses, donc le transcrit est plus exprimé. Ligne continue : épithélium ; ligne pointillée : mésenchyme. (D'aprèsPantalacci & Sémon [2014]).

#### 12.2.2.2 Utilisation du signal temporel pour rendre homologues des stades de développement entre différentes espèces et comparer leurs profils d'expression

L'une des difficultés à comparer les programmes de développement entre espèces est de pouvoir définir quels stades de développement sont homologues et peuvent être comparés. Anavy *et al.* [2014] ont proposé une une méthode appelée BLIND pour ordonner des transcriptomes provenant de séries temporelles. Dans un article récent, Levin *et al.* [2016] ont utilisé cet ordonnancement de stades consécutifs au sein d'une espèce pour ordonner les transcriptomes et pouvoir calculer un profil d'expression moyen, sur des fenêtres coulissantes basées sur cet

ordonnancement, afin de pouvoir comparer la dynamique d'expression entre 10 espèces différentes. Nous proposons ici une approche un peu différente mais toujours basée sur l'utilisation du signal temporel. Ainsi, nous sommes partis de l'observation que la signature temporelle est partagée entre les germes de molaires de souris et de hamster, ce qui signifie que les changements à grande échelle dans le transcriptome au cours du développement responsables de cette signature transcriptomique sont partagés dans les deux espèces. C'est pourquoi, nous avons décidé d'utiliser cette propriété de signature temporelle partagée pour rendre homologue les stades de développement entre la souris et le hamster. Pour rendre homologue les stades de développement entre les espèces, nous avons effectué une analyse de composante principale (PCA) des données transcriptomiques des deux espèces, qui a identifié un axe de variation commun qui sépare les transcriptomes en fonction de leur stade de développement. Ensuite, nous avons utilisé les coordonnées sur cet axe de composante principale pour définir des points de temps homologues entre les espèces (voir chapitre 9, figure 4). Cette approche nous permet d'obtenir des profils d'expression comparables entre espèces.

Par la suite nous souhaitons être capable d'aligner ces profils entre espèces afin de pouvoir estimer les changements de ces profils entre espèces. Pour cela, nous avons commencé un travail en collaboration avec le LBBE et notamment avec Laurent Guéguen et Lucas Michon. La méthodologie développée a deux finalités. La première a pour but d'estimer à l'échelle de l'ensemble des gènes du transcriptomes, le niveau de ressemblance entre des profils d'expressions, soit des différents types de molaire au sein d'une espèce, soit entre le même type de molaire mais entre deux espèces. Pour cela, les profils d'expression doivent être correctement alignés (ici avec l'algorithme d'alignement DTW (Dynamic Time Warping)) afin de pouvoir estimer le niveaux de ressemblance entre les profils d'expression. La deuxième finalité est cette fois de modéliser l'évolution des profils d'expression dans les différentes espèces comparées afin de pourvoir identifier les gènes dont l'évolution de l'expression est neutre ou sous sélection (aussi positive que négative). Néanmoins, l'utilisation de cette approche nécessitera des données d'expression pour un plus grand nombre d'espèces afin de pouvoir détecter statistiquement les profils de variations entre espèces. Cette méthode est actuellement en cours de développement pour être appliquée à un autre jeu de donnée, cependant, nous essayerions de l'utiliser dans notre analyse des germes molaires qui sera finalement composé de données d'expression pour quatre espèces de rongeurs : la souris, le hamster, la souris épineuse de la gerbille. Je pense que cette méthodologie, qui vise à comprendre comment les patrons d'expression ont évolué, présente un grand intérêt pour le domaine de l'évo-dévo. Tout d'abord, elle permettrait de mieux comprendre comment les patrons d'expression évoluent entre espèce et au sein d'une espèce entre organes. Deuxièmement, elle permettrait de corréler les modèles d'expression sous la sélection adaptative avec l'adaptation morphologique. Cette méthode est applicable à n'importe quelles données transcriptomiques dans la mesure où il y suffisamment d'espèces et où il y suffisamment de points temporels.

#### 12.2.2.3 L'identité de l'organe est la seconde signature transcriptomique

##### 12.2.2.3.1 Signature transcriptomique de l'identité de l'organe au niveau de l'organe entier

La seconde signature transcriptomique que nous avons mis en évidence est une signature des organes (voir annexe A et chapitre 9). Nous avons montré que l'identité des organes, dans notre cas la première molaire inférieure et la première molaire supérieure, est en partie porté par des différences relatives de proportions de tissus (figure 12.1). Ainsi, une partie de de l'identité inf/sup est causée par une hétérochronie entre ces deux molaires. Nous avons émis l'hypothèse que cet hétérochronie est le résultat de différences dans la proportion de tissus de cuspide entre la molaire supérieure et la molaire inférieure, ce que nous avons confirmé par modélisation de différentes proportions de tissu cuspide entre les molaires supérieures et inférieures. Cette modélisation permet de récapituler les différences temporelles observées entre les deux types de dents (voir annexe A). Des déconvolutions partielles des données d'expression, basées sur un ensemble de gènes spécifiques du tissu épithélial ou mésenchymateux, permettent également de déceler une différence dans les proportions relatives de tissus entre les deux dents. Ainsi, la molaire supérieure est enrichie en tissu mésenchymateux alors que la molaire inférieure est enrichie en tissu épithélial. Fait intéressant, l'identité de l'organe, molaire inférieure ou supérieure, peut encore être tracée au niveau des demi-organes. En effet, une analyse en composante principale des côtés buccal et lingual de premières molaires inférieure et supérieure de souris discrimine les transcriptomes de la molaire supérieure de ceux de la molaire inférieure. Quant au germe entier, nous avons identifié des différences dans les proportions relatives de tissus entre les côtés buccaux et les côtés linguaux. Ainsi nous avons montré que le côté lingual est enrichi en mésenchyme alors que le côté buccal est enrichi en épithélium. De la sorte, ces résultats mettent l'accent sur la contribution des proportions relatives des tissus dans l'identité transcriptomique des organes, facteur qui est souvent négligé lorsqu'on tente de démêler comment les différences de niveaux d'expression se traduisent aux niveaux des organes (figure 12.1) [Pantalacci & Sémon, 2014]. Ce résultat est généralisable à d'autres organes, en effet, nous avons montré que le rôle de ces proportions relatives de tissus dans la participation à l'identité des organes, est aussi observé au cours du développement des bourgeons de membre (*cf.* annexe A, figure 7) où l'on a montré que que les différences de quantité de cellules chondrogéniques corrèlent à la signature transcriptomique des membres postérieurs et antérieurs pendant le développement [Pantalacci *et al.* , 2016].

##### 12.2.2.3.2 Signature transcriptomique de l'identité de l'organe au niveau des tissus composants l'organe

Une autre étape pour caractériser la contribution des tissus à l'identité des organes consiste à étudier directement les transcriptomes au niveau des tissus. Grâce aux méthodes de déconvolution [Zhong *et al.* , 2013; Gaujoux & Seoighe, 2013], on peut estimer les proportions relatives des tissu dans un échantillon. Dans ces travaux, j'ai utilisé la déconvolution partielle qui permet d'estimer des proportions de tissus à partir de groupes de gènes marqueurs pour chaque tissus,

que j'ai défini à partir des données d'expressions de puces à ADN de tissus pures [O'Connell *et al.* , 2012] (voir chapitre 10). La déconvolution est une méthode très sensible aux types de données, par exemple j'ai déconvolué des données d'expression normalisées en *base mean* ou en *rld* (log transformées), ce qui m'a permis d'obtenir des proportions relatives mais ne me permettait pas d'estimer de proportions pour l'autre. C'est pourquoi, nous n'avons pu utiliser des déconvolutions totales, nos données d'expression provenant de la technologie RNA-seq et nos données d'expression de tissus pures de puces à ADN [O'Connell *et al.* , 2012]. Cependant, nous avons depuis peu à notre disposition des données RNA-seq de tissus purs (épithélium et mésenchyme) dont nous réalisons actuellement l'analyse. Nous projetons ainsi d'exécuter des déconvolutions complètes sur l'ensemble des données de germes molaires entiers et de demi-germes. Les avantages d'une déconvolution complète par rapport à la déconvolution partielle sont non seulement l'estimation des proportions, qui cette fois ne seront plus biaisées par le groupe des gènes marqueurs, mais aussi l'estimation de la contribution de chaque tissu au niveau d'expression global. Cela nous permettrait d'identifier quel tissu, du mésenchyme ou de l'épithélium, participent à l'identité inférieure/supérieure.

#### 12.2.2.3.3 Signature transcriptomique de l'identité de l'organe au niveau des cellules même de l'organe

Le séquençage en cellule unique est le nouveau venu de la famille des omics. Cette technologie permet d'accéder directement au transcriptome d'une cellule et non au transcriptome moyen d'une population de cellules. D'abord les cellules sont triées (par FACS par example), puis le contenu en ARN des cellules peut être obtenu soit par analyse RT-qPCR à haut débit [White *et al.* , 2011; Sanchez-Freire *et al.* , 2012], puces à ADN [Kurimoto, 2006] ou encore séquençage RNA-seq (*scRNA-seq*) [Picelli *et al.* , 2014; Grün & van Oudenaarden, 2015]. Pour le moment, le séquençage en cellule unique est principalement utilisé dans deux champs disciplinaires de la biologie : dans le domaine des études sur le cancer pour déterminer et détecter l'hétérogénéité cellulaire dans la population de cellules cancéreuses (revue dans Qian *et al.* [2016]). Ainsi que dans le domaine de la différentiation cellulaire et de l'engagement cellulaire (cell fate commitment) [Perié & Duffy, 2016].

Ce deuxième champ disciplinaire est étroitement lié à des études de développement. Par exemple, certaines études se sont ainsi concentrées sur le développement à des stades très précoces pour déterminer l'engagement du destin cellulaire. Récemment, Scialdone *et al.* [2016] ont examiné la formation du mésoderme précoce lors de la gastrulation des mammifères. Ils ont suivi la spécification de la lignée mésodermique précoce chez la souris à quatre points de temps différents, ils ont découvert que les profils de transcription étaient corrélés à l'emplacement des cellules dans l'embryon. Par conséquent, ces données suggèrent que le devenir cellulaire est en partie régulé par des informations de position plutôt que par des informations codées en dur dans la cellule.

Pour aller plus loin dans l'étude de l'identité transcriptomique des organes et le rôle joué par

les proportions relatives de tissus, le séquençage en cellule unique serait d'un grand intérêt. En effet, bien que le séquençage de tissus purs permette en partie de répondre à cette question, il reste une approche au niveau de la population de cellule. La technologie séquençage en cellule unique permettrait de questionner l'organisation du tissus en différentes sous-populations qui structure /régionalise le tissu. À partir de nos données transcriptomiques des côtés buccal et lingual des molaires supérieures et inférieures de souris, nous avons constaté que le côté lingual a une identité propre différente de l'identité buccale. Comme les gènes qui sont enrichis côté lingual sont aussi des gènes fortement mésenchyme, cela pose la question d'une structuration du mésenchyme, qui aurait pu échappé aux analyses histologiques. Par conséquent, l'utilisation d'un séquençage en cellule unique permettrait de remettre en question l'hétérogénéité du tissu, en identifiant (ou non) différents groupes de cellules partageant une signature transcriptomique similaire. De plus, il permettrait de quantifier le rôle relatif joué par les différences intrinsèques de transcription et les proportions relatives des tissus dans l'établissement de l'identité transcriptomique de l'organe.

### 12.2.3 Application de l'approche transcriptomique à l'étude du développement bucco-lingual

Le développemment des premières molaires de souris est intrinsèquement biaisé dans l'axe bucco-lingual. Or très peu d'études ce sont intéressées à caractériser cet aspect du développement de la molaire [Cam *et al.* , 2002; Obara & Lesot, 2007; Cho *et al.* , 2007; Lesot *et al.* , 2014]. Cependant, cette asymétrie bucco-linguale se retrouve dans plusieurs aspect du développement et de l'histoire évolutive des dents, notamment pour le remplacement dentaire, ce qui laisse présager que c'est un aspect important. Ainsi, les dents de remplacement, des reptiles [Handrigan *et al.* , 2010; Handrigan & Richman, 2010] aux mammifères [Järvinen *et al.* , 2008, 2009] (sauf quelques exceptions chez les mammifères), se mettent en place côté lingual. De plus, l'histoire évolutive des molaires est étroitement liée au côté lingual, puisque c'est la mise en place d'une deuxième rangée de cuspide du côté lingual qui offre une dimension 3D à la molaire tribosphénique [Butler, 1990]. L'ensemble de ces données suggèrent l'existence d'un "mécanisme" permettant de différencier les côtés buccal et lingual du germe dentaire, et que cette différenciation des côtés buccal et lingual est une propriété très ancienne de la dent car elle est présente chez l'ensemble des vertébrés à mâchoire.

Surtout, nous espérons que découvrir de ce qui rend le côté lingual différent du côté buccal nous permettra de mieux saisir pourquoi la molaire supérieure de souris met en place deux cuspides linguales supplémentaires. Nous avons choisi d'étudier un stade relativement précoce de la cuspidogenèse, le stade ED15, qui correspond au moment ou le germe de première molaire devient asymétrique dans l'axe bucco-lingual (*cf.* chapitre 10). Les nouveautés de notre approche ont été (i) d'étudier et de comparer les transcriptomes de demi-organe : le côté buccal et lingual des molaires supérieures et inférieures chez la souris et (ii) de combiner cette approche trans-

criptomique à une approche morphologique, à la fois basée sur des données expérimentales (collaboration à Prague avec Renata Peterkova) et sur des reconstructions 3D de germes de molaires (collaboration avec Ronan Ledevin).

#### 12.2.3.1 Découverte d'une forte identité buccale et linguale

D'une part, l'analyse des transcriptomes des côtés buccal et lingual a révélé une forte disparité transcriptomique entre les deux côtés (*cf.* chapitre 10). Comme mentionné plus haut, les données de la littérature suggéraient en effet l'existence d'une telle identité, cependant nous avons été surpris de trouver que le tiers des gènes exprimés dans la dent à ce stade ont une expression biaisée selon l'axe bucco-lingual. Par la suite, grâce à des analyses d'ontogénies génétiques (GO), nous avons cherché à caractériser les processus biologiques / cellulaires ainsi que les voies de signalisation qui différencient chaque côté. Nous avons couplé nos listes de gènes différentiellement exprimés entre les côtés buccal et lingual avec des données d'expression dans le mésenchyme et dans l'épithélium au même stade (ED15), ce qui nous a grandement facilité l'interprétation des analyses de GO. Ainsi, nous avons pu définir que malgré la présence d'un nœud d'émail du côté buccal, l'identité buccale ne dépend pas principalement de cette structure, elle est aussi caractérisée par la formation du réticulum en étoile (dans l'épithélium), la colonisation des nerfs (dans le mésenchyme) et la structuration de l'épithélium (IEE, OEE et boucles cervicales). Le côté lingual quant à lui n'est pas un côté "naïf"/indifférencié qui attend la formation d'un nœud d'émail, au contraire il a une forte identité spécifique, plus précisément nos données suggèrent une forte régionalisation du mésenchyme lingual. Notamment nous avons pu définir qu'au moins une partie de ces cellules mésenchymateuses étaient des cellules en migration active, qui participent, en autre, à l'angiogenèse. En outre, nous demandons si cette structuration pourrait refléter une activité particulière des cellules écto-mésenchymateuses en relation avec formation des cuspides. L'analyse de ces cellules, par séquençage en cellules uniques, pourrait nous aider à démêler ces aspects.

D'autre part, l'analyse des données morphologiques de reconstruction 3D de germes de molaires, en collaboration avec Ronan Ledevin, montre une croissance différentielle des boucles cervicales buccales et linguales des germes de molaires des stades ED14.5 à ED15.5, avec notamment une plus forte expansion du côté lingual. Ceci est cohérent avec des observations mentionnées dans la littérature [Cam *et al.* , 2002; Obara & Lesot, 2007; Cho *et al.* , 2007; Lesot *et al.* , 2014] mais qui n'avait jamais été quantifiée *per se*. De plus, des données récentes de nos collaborateurs à Prague (qui seront intégrées dans une prochaine version du manuscrit bucco-lingual) montrent que la croissance dans les boucles cervicales buccales et linguales sont différentes (figure 12.2). Du côté buccal les cellules restent solidaires de la boucle cervicale, mais pas du tout côté lingual. Cela suggère que côté buccal, la structure est relativement figée, alors que côté lingual il doit y avoir un remaniement très actif. Cependant il faudra certainement descendre au niveau cellulaire pour vraiment comprendre ce qui se passe côté lingual. Ces résultats sont cohérents avec des gènes comme *Ccr5* impliqué dans la migration cellulaire, qui sont sur-exprimés dans épithélium

lingual. Ensemble ces résultats nous permettent de proposer l'hypothèse que les côtés buccal et lingual (au moins aux stades précoces de la cuspidogenèse) sont deux structures dentaires différentes, dont je discuterai l'identité et le rôle dans le développement des molaires plus bas (voir section 12.3.4, *p.* 196).

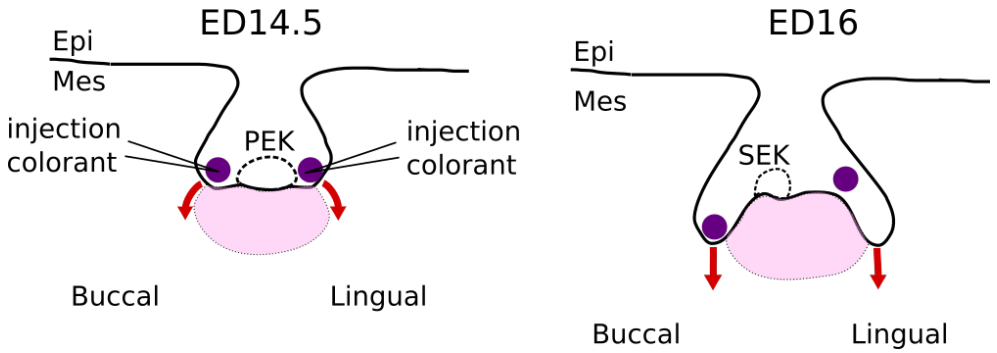



FIGURE 12.2 – **Différents mécanismes de croissance entre les côtés buccaux et linguaux.** Représentation schématique de résultats obtenus récemment par nos collaborateurs tchèques et qui seront intégrés dans une prochaine version du manuscrit. Au stade ED14.5, juste après la transition coiffe, du colorant (diI) est injecté dans les cellules au sommet de boucles cervicales. Quelques jours plus tard, au stade ED16, on peut voir le devenir des cellules qui ont été marquées.

## 12.3 Évolution du développement

Au cours de ce travaux, nous avons aussi examiné l'évolution des programmes développementaux de premières molaires inférieures et supérieures chez deux espèces de rongeurs : la souris et le hamster. Nous nous sommes demandés si les différences dans les signatures transcriptomiques des molaires supérieures et inférieures de souris et de hamster corrèlent avec les ressemblances/différences de la morphologie finale de la molaire. Nous avons identifié une conservation de la dynamique temporelle chez les deux espèces ainsi qu'une conservation des différences entre les programmes développementaux de molaire inférieure vs. supérieure. Cependant, nous avons aussi retrouvé un effet majoritaire dû à l'identité des espèces. Bien que les molaires inférieures de souris et de hamster partagent des caractères morphologiques, par exemple la mise en place de six cuspides alors que les molaires supérieures de souris mettent en place deux cuspides de plus que celles de hamster, elles sont transcriptomiquement aussi différentes entre elles que les molaires supérieures. Ce fort effet espèce entre les transcriptomes et l'absence de corrélation entre les différences au niveau morphologique et les différences dans les programmes de développement implique une dérive des programmes développementaux chez la souris et le hamster. Puis nous nous sommes demandés comment le programme de développement de molaires supérieures de souris a évolué pour être capable de modeler deux cuspides supplémentaires dans son côté lingual.

### 12.3.1 Conservation d'une différence entre programmes développementaux de molaires inférieures et supérieures

### 12.3.2 Une dynamique temporelle commune entre le hamster et la souris

À partir de ces données d'expression comparables, nous avons d'abord effectué une analyse multivariée afin d'identifier les axes de variations majeures présent dans les transcriptomes de premières molaires inférieures et supérieures de souris et de hamster au cours du développement. Cette analyse a révélé que les premières molaires inférieures et supérieures de souris et de hamster partagent une dynamique temporelle commune. De plus, nous avons mis en évidence que les différences entre les programmes développementaux des molaires supérieures et inférieures sont conservées entre nos deux espèces. En effet, les variations associées au type de dent, *ie.* molaires inférieure vs. molaire supérieure, représentent 4,6 % de la variation totale entre les transcriptomes de molaires de souris et de hamster. En outre, nous avons trouvé un chevauchement significatif des gènes différentiellement exprimés entre les molaires supérieures et inférieures chez la souris et le hamster, comprenant un ensemble de facteurs de transcription biaisés depuis le développement de la mâchoire précoce (données non présentées).

Le fait qu'il existe des différences entre les programmes développementaux des molaires supérieures et inférieures n'est pas surprenant. D'abord, d'un point de vue évolutif, dès l'apparition de la molaire tribosphénique il y a 200 millions d'années [Butler, 1990], il y a une très nette différence morphologique entre la molaire supérieure et la molaire inférieure. Cette différence morphologique est d'autant plus prononcée dans la première molaire supérieure de souris car elle met en place deux cuspides linguales supplémentaires. De plus, au cours du développement molaire, nous savons que les molaires supérieures et inférieures sont déployées dans différents «environnements de mâchoires». Ainsi, la spécification de l'identité inférieure-supérieure des mâchoires est opéré par une combinaison de gènes homéotiques, tel les *Dlx* [Depew, 2002; Beverdam *et al.* , 2002; Minoux & Rijli, 2010]. Finalement, les approches classiques de développement de souris mutantes ont souvent noté que la pénétrance d'une mutation diffère entre les molaires supérieures et inférieures (*eg.* $Bmp4^{-/-}$ [Jia *et al.* , 2012], le double mutant $Msx1^{-/-}$ $Osr2^{-/-}$ [Zhang *et al.* , 2009]), suggérant différents réseaux de développement en jeu dans les molaires supérieures et inférieures.

#### 12.3.2.1 Conservation d'un excès de différences entre les programmes développementaux des molaires supérieures et inférieures

Nous avons aussi mis en évidence que les différences entre les programmes développementaux de molaires inférieures et supérieures sont conservées entre la souris et le hamster (voir chapitre 9. Nous avons identifié un pic de différence entre la molaire supérieure et inférieure au début/milieu de la morphogenèse de la couronne, lorsque le deuxième cuspide fini d'être mise en place et que la troisième cuspide est en train d'être mise en place. Ce pic a la même amplitude et la même durée dans les deux espèces (chapitre 9, figure 9.3, *p. 143*). Par conséquent, les diffé-

rences des niveaux d'expression entre les molaires supérieures et inférieures sont plus faibles au début et à la fin de la morphogenèse de la couronne, alors qu'elles atteignent leur maximum au mi-cuspidogenèse (voir chapitre 9, *p. 143*). Ces résultats peuvent être simplement expliqués par une contrainte fonctionnelle au début, qui est la nécessité d'initier le germe dentaire et à la fin, qui est la sécrétion de l'émail et de la dentine. Ces deux contraintes fonctionnelles sont communes à chaque type de dent. Les étapes entre ces deux périodes contraintes pourraient différer entre les types de dents, car elles pourraient être liées à des caractéristiques spécifiques d'un type de molaire qui n'est pas en commun avec l'autre type. Ainsi, la différence entre les programmes de développement des molaires supérieures et inférieures a une forme de sablier inversé ou un double entonnoir.

### 12.3.3 L'évolution du programme développemental des molaires n'est pas corrélée avec leur évolution morphologique

#### 12.3.3.1 L'identité entre espèces est plus forte que l'identité entre molaires

Malgré des signatures transcriptomiques spécifiques à chaque type de molaires (inférieure ou supérieure), nous avons montré que l'effet espèces est le facteur majoritaire qui explique le mieux la variation entre les transcriptomes de premières molaires de souris et de hamster. Ces résultats peuvent être perçu comme inattendus parce que des études comparatives sur différents organes adultes chez différentes espèces [Liao & Zhang, 2005; Brawand *et al.* , 2011; Gilad & Mizrahi-Man, 2015; Breschi *et al.* , 2016] ont montré que la signature transcriptomique majoritaire entre tous ces échantillons est un fort effet organe qui prédomine sur l'effet espèce. Cependant, notre étude diffère dans les points suivant : (1) nous étudions des organes en développement alors que les études comparatives mentionnées plus haut étudiaient des organes adultes, (2) ces études examinaient des organes très différents, tel que le cœur, le cerveau et les reins qui ont des fonctions principales extrêmement différentes et qui sont donc censés exprimer des ensembles de gènes très différents, alors que nous travaillions uniquement sur les molaires. De plus, l'identité sérielle des molaires doit induire des contraintes spécifiques sur l'évolution du transcriptome (voir section 12.3.3.2 plus bas, *p. 195*). Malgré la nette divergence morphologique entre les molaires supérieures et inférieures, qui est plus ancienne (~200 MYA) que la divergence entre souris et hamster (~24 MYA) [Steppan *et al.* , 2004], l'effet espèce est la signature transcriptomique dominante entres les transcriptomes des premières molaires supérieures et inférieures de souris et de hamster. Les différences transcriptomiques entre la souris et le hamster, même dans les molaires inférieures qui partagent une morphologie semblable, suggèrent que les programmes de développement des molaires supérieures-inférieures ont subi une dérive du système de développement.

#### 12.3.3.2 Dérive du programme développemental

Il est tacitement admis qu'une morphologie conservée sous-tend une conservation du programme de développement responsable de la mise en place de cette morphologie. Cependant, comme nous l'avons montré dans ces travaux et comme il a déjà été démontré dans la mouche [Wotton *et al.* , 2015] et dans le nématode [Kiontke *et al.* , 2007], à la fois au niveau morphologique et au niveau transcriptomique [Yanai & Hunter, 2009], une morphologie conservée n'implique pas forcément une conservation au niveau de programmes de développement qui permettent la mise en place de cette morphologie. Ce phénomène est connu sous le nom de dérive du système de développement, en anglais Developmental System Drift (DSD). Pavlicev & Wagner [2012] proposent un modèle de pléiotropie et de compensation afin d'explique l'existence du DSD. Ce modèle stipule que tout changement adaptatif dans un trait est associé à un effet délétère dans d'autres traits en raison de la nature pléiotrope des gènes de développement. Par conséquent, on s'attend à ce que des modifications compensatoires soient sélectionnées pour diminuer ces effets pléiotropes négatifs. Il en résulte donc des changements du programme de développement qui ne sont pas liés à la morphologie finale mais qui atténuent la contrainte pléiotrope.

Comme les dents sont des organes sériels, il y a une pléiotropie extrêmement forte entre les molaires supérieures et inférieures. Pour cette raison, la contrainte pléiotropique entre les deux types de molaires pourrait être responsable de la dérive des programmes de développement. Ainsi, le découplage entre l'évolution de la morphologie molaire et l'évolution des programmes de développement des molaires dans chaque espèce, que nous avons mis en évidence dans ces travaux, nous a amené à proposer cette dérive des programmes de développement comment étant le mécanisme en jeu chez ces deux espèces. En outre, le processus de DSD est encore plus évident pour les molaires inférieures qui divergent fortement au niveau de leurs patrons d'expression malgré une faible proportion de changements morphologiques. Á terme, nous allons disposer des transcriptomes de molaires supérieures et inférieures au cours du développement pour deux espèces de rongeurs supplémentaires : la gerbille et la souris épineuse. Les molaires de gerbille ont, comme les molaires de hamster, le même nombre de cuspides sur la molaire supérieure ei inférieure (n=6) alors que comme la souris, la souris épineuse possède six cuspides sur sa molaire inférieure mais développe deux cuspides linguales supplémentaires sur sa molaire supérieure (figure 12.4). Ainsi, si la dérive des programmes de développement est effectivement en jeu dans les molaire on s'attend à trouver un fort effet espèce entre les transcriptomes des premières molaires de ces quatre rongeurs (*cf.* section 12.3.6). C'est effectivement ce nous avons observé dans une analyse pilote des transcriptomes de molaires provenant de ces quatre espèces pour un stade de développement. Cependant, ces résultats devront être confirmés sur l'ensemble du jeux de données et avec les données d'expressions obtenues avec le traitement Amalgam. Enfin, nous souhaiterions estimer l'intensité de la DSD entre les molaires supérieures et inférieures, néanmoins il nous manque une estimation de l'évolution neutre entre les niveaux d'expression des espèces considérées, dans notre cas la souris et le hamster, auquel viendront s'ajouter plus tard la gerbille et la souris épineuse. Bientôt, le laboratoire aura accès

à des données transcriptomiques sur les reins de divers rongeurs. Nous prévoyons d'utiliser ces données pour estimer le taux neutre d'évolution de l'expression entre le transcriptome du hamster et de la souris dans un autre organe. Nous voulons tester si la divergence d'expression entre les molaires chez la souris et le hamster est plus que prévu dans un modèle d'évolution neutre, suggérant ainsi que les programmes de développement des molaires sont sous sélection positive pour compenser l'effet délétère de la pléiotropie (modèle de Pavlicev & Wagner [2012]).

### 12.3.4 Perspectives sur l'origine des molaires tribosphéniques

Les molaires de souris, comme toutes les molaires d'ailleurs, ont perdu leur capacité de remplacement. Bien que les prémolaires, qui sont des dents molariformes aient conservées ce mécanisme de remplacement mais il est réduit à la production d'une seule dent de remplacement. Ce qui pose la question du lien entre les dents de remplacement, qui rappelons le se forment côté lingual, et le fait que le molaires sont des dents qui forment une deuxième rangée de cuspides linguales [Peterkova *et al.* , 2002]. Les dents molariformes intégreraient-elles une dent de remplacement pour former leur deuxième rangée de cuspide? Cette hypothèse avait déjà été avancé au début du $XX^{ème}$ siècle (notamment par L. Bolk, voir Butler [1956] et Peterkova *et al.* [2002]).

Comme abordé dans la partie introductive de ce manuscrit (*cf.* section 3.5.1, *p.* 3.5.1), les dents de remplacement se forment à partir de la lame dentaire de succession (figure 12.3A.) lingualement à la dent qui est en train de se développer. Comme mentionné plus haut (*cf.* section 12.2.3.1, *p.* 191), nos travaux ont permis de mettre en évidence l'existence d'une forte identité insoupçonnée des côtés buccal et lingual des germes de molaires au cours du développement (*cf.* chapitre 10). De plus, les résultats de nos collaborateurs montrent que les boucles épithéliales buccale et linguale ont en fait une dynamique de croissance qui est différente (figure 12.2 et section 12.2.3.1, *p.* 191), ce qui suggère que ces deux boucles sont en fait deux structures distinctes, parce qu'elles ont des modes de croissance différentes. Ainsi, au stade ED14.5-ED15, juste pendant et après la transition coiffe, le nœud d'émail primaire (PEK) se met en place au milieu de germe et l'épithélium croît asymétriquement (dans l'axe bucco-lingual) autour du mésenchyme en l'enserrant (*cf.* chapitre 10), ces "bras" d'épithélium sont appelés boucles cervicales. Cet épithélium est considéré comme étant l'épithélium formant la dent (*cf.* figure 12.3.3.2, représenté en violet).

Une possibilité intéressante serait que la boucle épithéliale buccale soit effectivement l'épithélium formant la dent, alors que a boucle linguale correspondrait à la lame dentaire de succession (figure 12.2.3.1B). La véritable boucle linguale de ce premier germe ne se formerait donc pas, hormis dans certains contextes, comme le mutant *caspase-3* où on voit clairement une petite boucle supplémentaire entre la boucle buccale et la boucle linguale [Matalova *et al.* , 2006]. Puis, plus tard au cours du développement de la molaire, cette boucle linguale perdrait ce potentiel et deviendrait une "vraie" boucle cervicale, elle prendrait une identité d'épithélium formant la dent. À partir de cette hypothèse, nous pouvons proposer que l'acquisition de cette identité d'épithélium formant la dent se déclare plus tardivement dans la boucle épithéliale linguale de

la molaire supérieur de souris. Ceci expliquerait alors pourquoi cette dent est capable de mettre en place une troisième rangée de cuspide. Ainsi, l'idée que la boucle linguale corresponde à la lame dentaire de succession juxtaposée à une véritable boucle linguale dont le développement est réprimé est aussi en accord avec morphologie transitoire adoptée par la molaire supérieure de souris au cours de la transition coiffe à cloche [Peterkova *et al.* , 2002] (*cf.* schéma bilan figure 7 du chapitre 10). Cette hypothèse permettrait d'expliquer la si forte différence transcriptomique entre le côté lingual et le côté buccal que l'on a observé au stade ED15 (*cf.* chapitre 10). Si cette hypothèse est vérifiée alors on s'attend à avoir beaucoup moins de différences entre les côtés buccal et lingual à des stades où le côté lingual a perdu son identité de lame dentaire de succession au profit d'une identité d'épithélium formant dent.

Pour tester cette hypothèse, on pourrait réaliser une analyse transcriptomique comparée entre nos données bucco-linguale de souris et des données transcriptomiques obtenues chez les reptiles. Plus précisément, il s'agirait des transcriptomes de la lame dentaire de succession ainsi que des transcriptomes des boucles cervicales buccale et linguale de la dent en développement chez le reptile. Si notre hypothèse est avérée, alors on s'attend à ce que la boucles épithéliale linguale de souris au stade ED15 partage plus de caractéristique avec la lame dentaire de succession de reptile qu'avec les boucles buccale et linguale de la dent en formation.

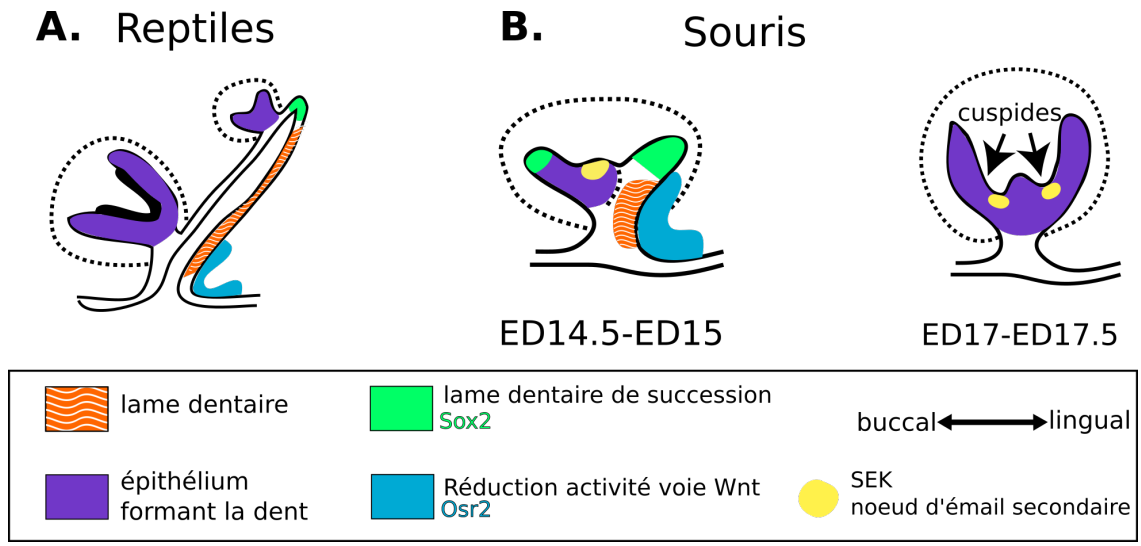



FIGURE 12.3 – **Identité des boucles épithéliales de la molaire de souris au cours du développement.**
On peut supposer que l'inhibition de la voie Wnt et BMP mise en évidence sur le côté lingual de la molaire de souris [Jia *et al.* , 2016] pourrait limiter à la fois la formation de la boucle linguale du premier germe et la formation du germe de succession. On voit donc dans cette inhibition linguale un mécanisme évolutif assez simple pour expliquer l'évolution des molaires et conjointement la réduction du potentiel de remplacement.

### 12.3.5 Perspectives sur l'origine du plan murin : comment la souris met en place deux cuspides linguales supplémentaires ?

#### 12.3.5.1 Les spécificités de la molaire supérieure de souris

Nos travaux soulignent qu'une des spécificités de la première molaire supérieure de souris est d'être enrichie en mésenchyme, alors que cet excès de mésenchyme n'est pas retrouvé dans la molaire supérieure de souris (voir chapitre 10). Ces résultats sont très intéressants car il a été montré que le nombre de cuspides mises en place sur un germe de molaire est dépendant de la quantité de cellules mésenchymateuses [Hu *et al.* , 2006]. Par conséquent, l'excès de mésenchyme de la première molaire supérieure de souris pourrait être responsable de la mise en place des deux cuspides supplémentaires de la molaire supérieure. De la sorte, il serait intéressant de réaliser des expériences expériences de recombinaisons hétérotypiques entre l'épithélium et le mésenchyme de germes de molaires de souris et de hamster. Est-ce que la recombinaison d'un mésenchyme de molaire supérieure de souris avec un épithélium de molaire de hamster permet le développement d'une molaire avec des cuspides supplémentaires ? Réciproquement, est-ce que l'association d'un mésenchyme de hamster avec un épithélium de molaire supérieure de souris donnera une molaire avec un nombre ancestral de six cuspides ? Ces expériences pourraient également nous permettre de tester si le modèle de réaction-diffusion qui contrôle la taille et l'espacement des cuspides dans le germe molaire [Cai *et al.* , 2007; Cho *et al.* , 2011; Salazar-Ciudad & Jernvall, 2002] est aussi un facteur qui entre en jeu dans la mise en place des deux cuspides supplémentaires dans la molaire supérieure de souris, alors que la molaire supérieure de hamster conserve un nombre ancestral de cuspides.

Nous avons aussi noté que, malgré la conservation des différences des programmes de développement entre la molaire supérieure et la molaire inférieure, ces différences sont plus prononcées chez la souris tout le long de la cuspidogenèse, ce qui est cohérent avec la plus grande différence morphologique entre les molaires de souris. Néanmoins, nous n'avons pas encore pu isoler cet excès de différences inf/sup de la plus grande proportion de mésenchyme dans la molaire supérieure de souris. C'est pourquoi la déconvolution complète nous permettrait de voir ce qui reste de la différence supérieure-inférieure une fois que l'effet de «proportion de mésenchyme» est éliminé. De plus, l'excès de différences entre molaire inférieure et supérieure à mi-cuspidogenèse qui est conservé chez les deux espèces, arrive plus tôt chez la souris ce qui suggère une hétérochronie entre les programmes de développement des molaires de souris et de hamster. Une analyse de de Gene Ontology nous a appris que les gènes qui sont responsables de cet excès de différence inf/sup pourraient être responsable dans la molaire supérieure d'une plus forte morphogenèse, avec moins de cellules mitotiques et plus de cellules en migration à la fois chez la souris (*cf.* annexe A et chez le hamster (*cf.* chapitre 9). Ces résultats suggèrent que la molaire supérieure de souris pourrait anticiper une étape de morphogenèse essentielle au développement de la molaire supérieure (car commune entre souris et hamster), ce qui lui permettrait de "gagner du temps" à la fin de la morphogenèse de la couronne pour pouvoir

mettre en place ses deux cuspides supplémentaires.

De façon à pouvoir tester les hypothèses mentionnées dans cette section, nous pourrions utiliser le modèle *in silico* de développement des dents proposé de Salazar-Ciudad & Jernvall [2002] et en modifier les paramètres, puis nous relieront la modification des paramètres à des processus biologiques susceptibles de se dérouler au cours du développement de la molaire.

Ainsi, nous pourrions tester si une plus grande quantité de cellules mésenchymateuses est suffisant lors de la phase d'initiation du modèle, pour permettre la mise en place de cuspides supplémentaires sur le germe. De plus, nous pourrions tester si le modèle de réaction-diffusion est différent dans la molaire supérieure de souris, ce qui permettrait aux cuspides mises en place d'avoir une zone d'inhibition réduite. Nous pourrons aussi coupler ces deux paramètres : (i) réduction de la zone d'inhibition des nœuds d'émail et (2) une augmentation des cellules mésenchymateuses, qui pourrait signifier que le germe est légèrement plus gros ou a plus de matière pour croître d'avantage, alors il y aurait assez d'espace sur le germe pour mettre en place des cuspides supplémentaires, afin d'estimer si l'interaction entre ces paramètres est nécessaire pour permettre le développement de cuspides supplémentaires. Enfin, comme nous avons mis en évidence une exagération hétérochronique des processus morphogénétiques dans la molaire supérieure de souris, il nous faudra aussi rechercher et modifier les paramètres qui influent sur la dynamique de la morphogenèse du germe.

#### 12.3.5.2 La différence bucco-linguale est plus forte dans la molaire supérieure de souris

Il y a trois hypothèses pour expliquer la mise en place de cuspides supplémentaires par la molaire supérieure de souris. (1) L'addition terminale, dans ce cas on s'attend à ce que les programmes de développement des molaires supérieures et inférieures soient assez similaires et diffèrent vers la fin de la morphogenèse de la couronne lorsque la molaire supérieure met en place ses dernières cuspides, qui sont les cuspides linguales supplémentaires. (2) Une hétérométrie, sous cette hypothèse on s'attend à ce que l'asymétrie bucco-linguale soit exagérée dans la molaire supérieure. (3) Une hétérochronie, sous cette hypothèse on s'attend à ce que l'asymétrie bucco-linguale perdure plus longtemps dans la molaire supérieure. Une quatrième hypothèse pourrait être un mélange entre hétérométrie et hétérochronie, où les différences bucco-linguales seraient plus fortes et dureraient plus longtemps

Nos données ont montré que les différences bucco-linguales étaient en effet plus grandes dans la molaire supérieure (*cf.* chapitre 10). Cet excès de différences bucco-lingual (BL) dans la molaire supérieure qui arrive assez précocement dans la morphogenèse de la couronne (au stade ED15, où seulement une cuspide est mise en place) élimine donc l'hypothèse de l'addition terminale. Cet excès de différences BL dans la molaire supérieure est donc en accord avec notre hypothèse d'hétérométrie ; néanmoins, nous n'avons pas pu conclure sur l'action d'une éventuelle hétérochronie car nous ne disposons de données BL que pour un unique stade de développement. Pour lever cette zone d'ombre, il faudrait générer un autre jeu de données BL à un stade plus

tardif.

Enfin, il est intéressant de noter que le modèle *in silico* de développement de dents [Salazar-Ciudad & Jernvall, 2002] a besoin de paramètres contrôlant des biais de croissance dans l'axe bucco-lingual, afin d'être capable de modéliser correctement le développement d'une dent. Ceci est une indication de plus sur le rôle important joué par cette asymétrie bucco-linguale pendant le développement des dents. Cependant, tel qu'il est implémenté aujourd'hui, ce modèle n'est pas capable de reproduire précisément l'ordre d'apparition des cuspides. En effet, dans e modèle les deux premières cuspides apparaissent en même temps alors que pendant le développement elles apparaissent séquentiellement, la première se met en place du côté buccal et la deuxième se met en place côté lingual et cette formation à l'air d'être conserver entres les molaires (*cf.* chapitre 9). Cette divergence entre le modèle et la réalité souligne qu'une partie de la dynamique de l'asymétrie bucco-linguale n'est pas prise en compte par le modèle.

Un développement futur de ce travail sera d'explorer cette modélisation et d'améliorer le modèle afin d'être en mesure de former une molaire supérieure de souris avec ses deux cuspides linguales supplémentaires. Ensuite, nous aimerions explorer l'espace des paramètres pour déterminer si l'acquisition de cuspides supplémentaires côté lingual pourrait être atteint avec différents paramètres. Si tel est le cas, cela suggérerait qu'il existe plusieurs voies de développement différentes pour obtenir la même morphologie finale.

#### 12.3.5.3 Combinaison des analyses bucco-linguale chez la souris et inférieure-supérieure chez la souris et le hamster

Les différences inf/sup que nous avons identifié entre les molaires de souris et de hamster sont deux de types : (i) d'abord il y a des différences intrinsèques entre les programmes de développement des molaires inférieures et des molaires supérieures, puis (ii) il y a les différences qui permettent à la molaire supérieure de souris de mettre en place deux cuspides linguales supplémentaires. Au contraire, le hamster a le même nombre de cuspides sur ses molaires inférieures et supérieures, c'est pourquoi nous avons utilisé le hamster comme une espèce contrôle où le même nombre de cuspide est formé sur la molaire inférieure et la molaire supérieure. Néanmoins, les molaires de hamster et de souris possèdent un certain nombre caractéristiques spécifiques, telles que la présence de zone exempte d'émail dans les molaires de souris qui ne sont pas présentes chez le hamster. Nous avons donc besoin d'éliminer les gènes responsables de ces caractéristiques, par exemple nous pouvons le faire en identifiant les gènes exprimés de manière différentielle entre les molaires inférieures de souris et de hamster puis enlever ces gènes de ceux identifiés comme étant différentiellement exprimés entre molaires supérieures de souris et hamster. Un critère supplémentaire pour ajouter un gène à la liste des gènes candidats participant à l'acquisition de cuspides supplémentaires linguales sur la molaire supérieure est évidemment que ce gène soit exprimé avec un biais bucco-lingual. De plus, de bons gènes candidats auraient également un biais asymétrique bucco-lingual précoce.

Pour le moment, nous avons concentré nos analyses sur l'évolution du programme de développement entre les molaires de souris et le hamster. Désormais, nous souhaitons nous concentrer sur comment la molaire supérieure de souris se différencie de la molaire supérieure de hamster et est capable deux former deux cuspides linguales supplémentaires. Pour cela, nous utiliserons encore une fois les données transcriptomiques dans leur ensemble pour identifier les signatures transcriptomiques spécifiques ou exagérées dans la molaire supérieure de souris, de plus cet approche nous permettra d'identifier le gènes et les processus cellulaire impliqué dans cette formation. Pour cela, nous tirerons profit de l'ensembles de jeux données, ceux de germe entier de premières molaires inférieures et supérieures de souris et de hamster au cours de la cuspidogenèse, notre jeu de données bucco-lingual et notre jeu de données de tissus purs composé d'épithélium et mésenchyme de molaire supérieure et inférieure à deux stades de développement .

#### 12.3.5.4 *Bmper*, un gène candidat du développement des cuspides linguales supplémentaires chez la souris

Au cours de nos travaux, nous avons identifié un gènes candidat, il s'agit de *Bmper* (*cf.* chapitre 11). Ce gènes est différentiellement exprimé entre la les molaires supérieures et inférieures de souris mais pas de hamster. De plus, nous avons révélé que l'expression de ce gène diminue brusquement lorsque le germe de molaire n'est plus asymétrique le long de l'axe bucco-lingual (*cf.* chapitre 11, figure 11.2, *p. 176*). Enfin, nos données bucco-linguales (*cf.* chapitre 10) nous ont appris que *Bmper* est exprimé avec un biais buccal dans les molaires supérieures ei inférieures de souris et nous avons confirmé ces résultats par de hybridations *in situ* (*cf.* chapitre 10, figure 11.3, *p. 178*).

Pour parfaire sa liste de bon candidat, les données de la littératures nous ont appris que *Bmper* est connu pour interagir avec *Bmp4* et réguler et ainsi réguler l'action des BMPs, dans un autre système biologique que la dent [Moser *et al.* , 2003; Kelley *et al.* , 2009]. Comme présenté dans la partie introductive de ces travaux (*cf.* introduction *p. 46*), *Bmp4* est connu pour son importance au cours du développement de la molaire [Vainio *et al.* , 1993; Jernvall *et al.* , 1998; Jernvall & Thesleff, 2000]. De même, il est intéressant de mentionner que *Bmp4* est exprimé dans le mésenchyme avec un bias buccal au cours du développement de la molaire [Chen *et al.* , 1996; Åberg *et al.* , 1997; Cho *et al.* , 2007; Jia *et al.* , 2012]. Des résultats préliminaires sur l'étude de germes de molaires de souris homozygotes $bmper^{-/-}$ au stade ED18.5, quand toutes les cuspides ont été mises en place, suggèrent que *Bmper* affecte la formation des cuspides supplémentaires de la molaire supérieure de souris. Cependant, ces résultats ont été obtenu par l'analyses de coupes histologiques, dont l'analyse n'est pas aisée. C'est pourquoi nous souhaiterions utiliser la microphotographie à rayons-X de manière à pouvoir imager des dents entiers, sans avoir à passer par un stade de coupe. Cette technique d'imagerie nous permettrait alors de reconstruire en 3-dimensions le germe de molaire d'une souris mutante $bmper^{-/-}$ et d'être plus précis sur l'effet de *Bmper* au cours du développement de la molaire. Cependant, ce projet n'a pas pu aboutir dans les limites de mes travaux de thèse car nous avons rencontré des problèmes pour avoir

accès à de nouveaux échantillons, ralentissant ainsi cette analyse.

Pour conclure sur cette approche gène centrée, l'analyse de *Bmper* nous a fournis des lignes directrices à suivre afin d'identifier de nouveaux gènes candidats à l'acquisition de cuspides linguales supplémentaires sur la molaire supérieure de souris. Ainsi, au cours de nos prochaines analyses, nous prévoyons de rechercher systématiquement des gènes différentiellement exprimés entre la molaire supérieure et la molaire inférieure, avec un biais bucco-lingual et dont les niveaux d'expression montrent un point d'inflexion correspondant à l'arrêt de la croissance bucco-linguale.

### 12.3.6 Comment se traduit la convergence évolutive dans les transcriptomes ?

La souris et la souris épineuse partagent la même morphologie dentaire : une première molaire inférieure avec un nombre ancestral de cuspide et une première molaire supérieure avec deux cuspides supplémentaires linguales (figure 12.4). Ces deux espèces de rongeurs ont acquis cette morphologie de manière convergente au cours de l'évolution. À terme nous disposerons de données pour germes entier de molaires inférieures et supérieures pour quatre espèces de rongeurs : le hamster et la souris, pour lesquels nous possédons déjà les données et dont l'étude représente une partie des travaux présentés dans ce manuscrit, la souris épineuse et la gerbille, dont la totalité des données a été collectée et est prête à être séquencée. Le hamster et la gerbille seront utilisées comme des espèces contrôles car elles possèdent le même nombre de cuspides sur leurs molaires inférieures et supérieures, ce qui représente l'état ancestral.

De la sorte, l'analyse du jeu de données de transcriptomes de premières molaires inférieures et supérieures au cours de la cuspidogenèse pour nos quatre espèces de rongeurs, dont deux sont considérés comme des contrôles avec le même nombre de cuspides sur la molaire inférieure et supérieures, va nous permettre d'interroger la convergence au niveau transcriptomique. Pour comparer les niveaux d'expression entre espèces nous utiliserons Amalgam, la méthode qui a été développée dans le laboratoire (*cf.* section 12.2.1). Nous utiliserons ensuite le cadre méthodologique que nous avons développé au cours de ces travaux, qui passent par des analyses multivariées, l'analyse des signatures transcriptomiques, l'estimation des proportions tissulaires relatives, l'analyse de gènes différentiellement exprimés ainsi que la comparaison des profils d'expression, tout ceci afin d'identifier les similitudes et les différences ainsi que les spécificités entre l'ensemble de ces programmes développementaux.

D'après nos analyses comparatives entre les molaires de souris et de hamster (chapitre 9), on s'attend à ce qu'il y ait une signature de dérive des programmes de développement entre ces quatre espèces, c'est-à-dire que les variations majoritaires transcriptomes seraient dues d'abord à un effet espèce. Néanmoins, la souris et la souris épineuse partage une morphologie identique de leurs molaires supérieures, est-ce que l'on détecte cette acquisition convergente de morphologie au niveau de leur transcriptome, en d'autres termes est-ce que les programmes de

développement ont subi des une évolution similaire pour aboutir à une morphologie similaire? Ensuite, nous pourrions identifier et quantifier ces similarités, il y a trois niveaux de similarité :(1) similarité génétique pure, ce sont les mêmes gènes qui ont été recruté, (2) similarité génétique large, ce sont les même voies de signalisations qui entre en jeu, mais ce sont des gènes différents de ces voies qui entrent en jeu ou encore (3) similarité au niveau des processus/des modules, dans ce cas c'est une fonction ou un processus cellulaire qui permet d'aboutir au même phénotype. Dans ce dernier cas de figure, on peut par exemple émettre l'hypothèse que le germe de molaire supérieur de la souris épineuse a, comme le germe de molaire supérieur de la souris, plus de mésenchyme, dont l'augmentation du nombre de cellules composant le germe est connu pour augmenter le nombre de cuspides [Hu *et al.* , 2006].

Pour le moment, j'ai réalisé l'analyse de données transcriptomiques préliminaires de gerbille et de souris épineuse, qui suggère que *Bmper* ne soit pas impliqué dans la formation des cuspides supplémentaires chez la souris épineuse, car il n'est pas différentiellement exprimé entre les molaires supérieures et inférieures de souris épineuses. De plus, ces données suggèrent que la molaire supérieure de souris épineuse ne soit pas enrichie en mésenchyme, comme c'est le cas pour la souris. Néanmoins, nous avons peu mettre en évidence d'autres gènes convergents. Cependant, cette analyse est encore trop préliminaire (un seul stade de développement comparé et pas avec les séquences obtenues par Amalgam).

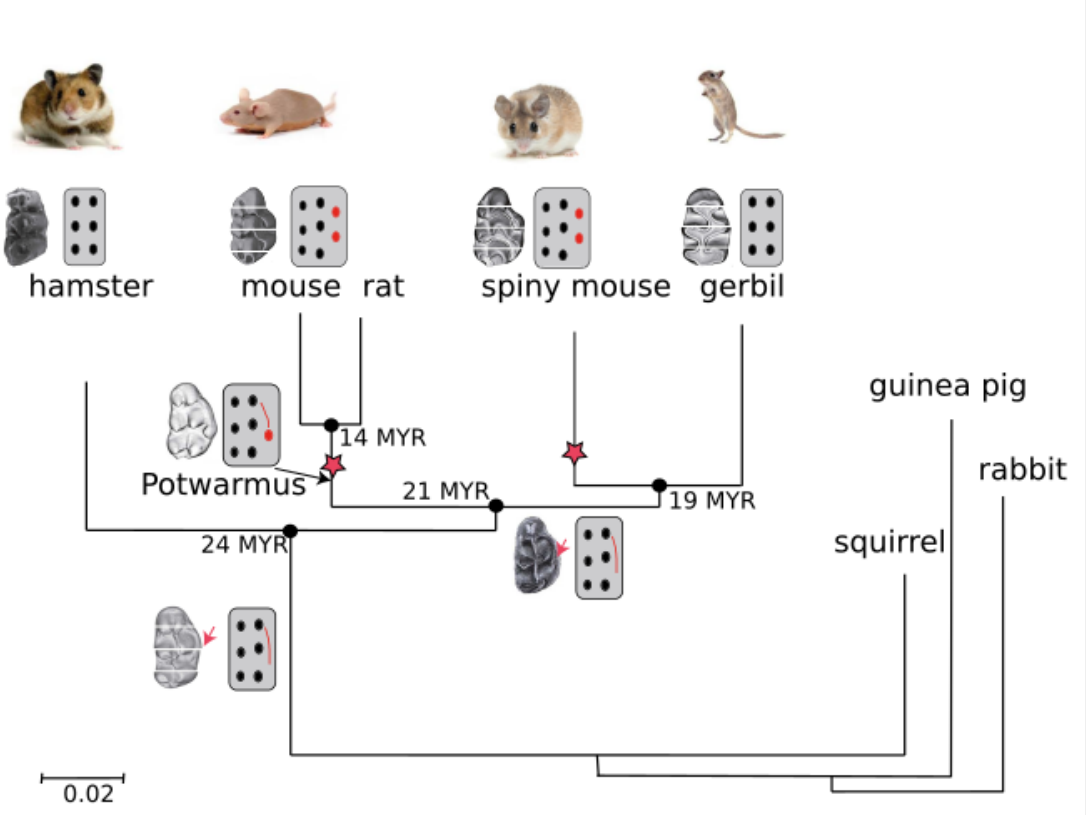



FIGURE 12.4 – **Convergence de l'acquisition des cuspides linguales supplémentaires chez la souris et la souris épineuse.** Arbre représentant le liens phylogénétiques entre le hamster, la souris, la gerbille et la souris épineuse, retraçant l'histoire évolutive de l'apparition des cuspides supplémentaires de molaires supérieures de la souris et de la souris épineuse. MYR = millions d'années.

# Appendix Part V

# A Transcriptomic signature of serial organs : the mouse first upper and lower molar

# Transcriptomic signatures shaped by cell proportions shed light on comparative developmental biology

December 19, 2016

Article accepted in Genome Biology

Sophie Pantalacci[1]*; Laurent Guéguen[2] ; Coraline Petit[1] ; Anne Lambert[1] ; Renata Peterkova[3]; Marie Sémon[1]*.

Affiliations: [1] UnivLyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, Laboratoire de Biologie et Modélisation de la Cellule, 15 parvis Descartes, F-69007, Lyon, France
[2] Laboratoire de Biométrie et Biologie Évolutive (LBBE), Université de Lyon, Université Lyon 1, CNRS, Villeurbanne, France
[3] Department of Teratology, Institute of Experimental Medicine, Academy of Sciences AS CR, Videnska 1083, 142 20 Prague, Czech Republic

* Corresponding authors: Sophie Pantalacci and Marie Sémon
sophie.pantalacci@ens-lyon.fr ; marie.semon@ens-lyon.fr

**Abstract**

**Background**
Comparative transcriptomics can answer many questions in developmental and evolutionary developmental biology. Most transcriptomic studies comparing developing embryos or organs start by showing global patterns of variation in transcriptomes that differ between species or organs through developmental time. However, little is known about the kinds of expression differences that shape these patterns.

**Results**
We compared transcriptomes during the development of two morphologically distinct serial organs, the upper and lower first molars of the mouse. We found that these two types of teeth largely share the same gene expression dynamics but that three major transcriptomic signatures distinguish them, all of which are shaped by differences in the relative abundance of different cell types. First, lower/upper molar differences are maintained throughout morphogenesis, and stem from differences in the relative abundance of mesenchyme and from constant differences in gene expression within tissues. Second, there are clear time--shift differences in the transcriptomes of the two molars related to cusp tissue abundance. Third, the transcriptomes differ most during early--mid crown morphogenesis, corresponding to transiently exaggerated morphogenetic processes in the upper molar involving fewer mitotic cells but more migrating cells. From these findings, we formulate hypotheses about the mechanisms enabling the two molars to reach different phenotypes. We also successfully applied our approach to fore and hindlimb development.

**Conclusions**
Gene expression in a complex tissue reflects not only transcriptional regulation but also abundance of different cell types. This knowledge provides valuable insights into the cellular processes underpinning differences in organ development. Our approach should be applicable to most comparative developmental contexts.

**KEYWORDS** Comparative transcriptomics, developmental biology, transcriptomic signature, temporal dynamics of gene expression, heterochrony, serial homology, tooth

# BACKGROUND

In multicellular organisms, organ development is a dynamic process manifested by changes in gene expression together with changes in tissue organisation. Both interact to finally shape the adult phenotype. Therefore, studying the dynamics of gene expression during development helps to decipher developmental processes and their evolution across species [1, 2]. Historically, this was long done by looking at the expression of individual genes across conditions (time and/or mutants, organs, species). The raise of transcriptomics, triggered a growing interest for identifying groups of genes with concerted expression differences across conditions. Clustering analysis of microarray or RNAseq datasets are grounding numerous developmental studies, and have provided helpful insights in comparative developmental biology (e.g. [3]). More recently, data mining with purposely-developed tools allowed extracting information carried collectively by hundreds of genes, sometimes called "transcriptomic signatures", in other fields such as toxicology and cancerology). These more global comparisons of transcriptomes have proven extremely useful to decipher cell lineages (e.g. [4, 5]), date embryonic samples (e.g. [6]) or extract organ or cell-specific genes during organogenesis and adulthood (e.g. [7]) (see also this review: [8]). In evo-devo, they were used as a measure of similarity, to assess the conservation of development depending on the stage and the phylogenetic scale (e.g. [9] ,hourglass pattern or inverted hourglass patterns in transcriptomes [10-13],, developmental milestones in nematode species, [13],). or question homology relationships [14, 15] or the origin of new cell types [16]. The reciprocal question, how transcriptomic differences inform us of the developmental basis of morphological disparities between species or homologous organs has been overlooked. Do transcriptomic signatures distinguish the development of homologous organs with different morphologies? Does the study of these signatures bring out an integrated view on differences in their developmental processes?

These questions can be more easily tackled, in a single species, in the context of serial homology. Serial organs are organs of the same "type", but with different phenotypes, iterated at different positions of the same individual (e.g. leaves in plants or body appendages such as limbs in animals). Even when these organs are unambiguously of the same type, their shape can be quite different depending on their position (e.g. fore/ hind limb). Because they develop using the same genome, it is obvious that differential expression is the only factor promoting phenotypic differences, by building on context-dependent differences related to the position of the organ in the body. The comparison of gene expression in serial organs therefore helps to understand which differences in developmental processes result in organs that are similar, yet different.

In animals, the most commonly studied serial organs are body appendages, typically arthropod appendages and vertebrate limbs [17-20]. Molars, the model system used in this study, are another example, whether considered in the same or in opposite jaws [21, 22]7̃d. Research on serial appendages has put a strong focus on the role of what we called herein "identity genes". These are genes with position-dependent expression, found specifically or in a specific combination in one type of appendage ([18-20]. The focus on these genes is consistent with the spectacular "homeotic transformations" of one appendage into another, that arise when manipulating their function. These identity genes are most often transcription factors with homeodomains expressed specifically at organ initiation (e.g. homeotic genes for insect appendages, and homeodomain transcription factors Tbx4 and Pitx1 for vertebrate limbs [19, 23]., Dlx5,6 for jaws, and by extension, molars [22, 24]). They introduce specific regulations in an otherwise "default" developmental program made of the cascade regulation of thousands of common genes. These differential regulations are thought to be responsible for shape differences between serial appendages, yet we know little about them. Indeed these models focus on the initiation of development, somewhat discounting the rest of the process. In contrast, from the roots of evolutionary developmental biology, studies were interested in comparing the developmental sequence producing two serial organs (within and between species). They have shown numerous heterochronies, that is, temporal shifts in and/or between developmental processes (followed with marker genes or histological staining), trying to relate them to the final phenotypic outcome [25, 26].

A relatively poorly explored yet crucial question is to what extent, and how, gene expression differs during the development of serial appendages and how this reflects differences in developmental programs underpinning ultimate phenotypic differences. Previous studies comparing gene expression in serial organs lacked a detailed temporal resolution - when not focusing on a single stage (e.g. limbs [27], different type of teeth [28]). Moreover, they did not look for transcriptomic signatures, because they interpreted differential expression analysis at the level of the gene or group of genes. Here we go beyond these limits by studying the dynamics of gene expression during the development of two phenotypically well-differentiated serial appendages, upper and lower molars in mouse. Below, we provide general background on molar development and more specific background on developmental differences between lower and upper molar.

Using this model, we benefit from the strong body of work on molar evolution and development [29-31]. Although mouse lower and upper molars are both immediately recognizable as bona fide molars, exhibiting a mineralized crown with little hills named "cusps", their shape is significantly different (Fig 1). It is especially noticeable that the upper molar exhibits a supplementary row of cusps on its lingual side (red arrow on Fig 1), on top of other differences (inclination, shape and arrangement of cusps, presence of accessory cusps). Present-day mouse lower and upper molar developmental programs have been shaped by at least 150 million years of evolution: molars evolved in early mammals from less complex teeth, and since then mammalian lower and upper molar most of-

ten have had different morphologies [21]. Molars develop as jaw appendages and lower/upper molar specification is intimately linked to lower and upper jaw regionalization from the first branchial arch of the embryo (e.g. dependent on Dlx genes [22, 24, 31-33] ). Of course, there are common principles for the development of lower and upper molars. The earliest steps of molar development (see figure 1 for more details) involve molar specific genes (e.g. as compared to incisor), like *Barx1.* These steps give rise to a molar tooth germ arranged in two compartments, epithelium and neural-crest derived ectomesenchyme (herein called mesenchyme for simplicity). Crown morphogenesis rests on shaping the epithelium-mesenchyme interface and differentiating enamel and dentin-secreting cells. These cells secrete a mineralizing matrix, which fixes the shape adopted by the soft tissue into a crown. This process depends heavily on self-organizing principles, involving epithelium- mesenchyme interactions through diffusible molecules and tissue mechanical properties [34,35]. Key to the process are epithelial signaling centers called Enamel Knots, which organize crown morphogenesis in time and space [36, 37]. The Primary Enamel Knot (PEK) directs the folding of the epithelium around the condensing mesenchyme, defining the molar germ. Later, the Secondary Enamel Knots (SEK) direct further folding of the epithelium, defining the cusps. These SEK are patterned sequentially (from ED16.0) while molar germ grows [36, 37]. They also act as epicenters for the progression of proliferation arrest and differentiation. Beyond these common principles, lower and upper molar developments of course show developmental and genetic differences. Our knowledge of these differences is however patchy, as a strong focus has been put on lower molar development ([29, 30], see also the database for expression in tooth: `http://bite-it.helsinki.fi/` and transcriptomic data generated with microarray technology ([38, 39]). One such difference is the degree of participation of a vestigial bud to the anterior part of the first molar germ. Its relevance to final phenotypic differences between lower and upper molar is unclear [40]. A second difference is the persistence of lingual extension in the upper molar germ at ED16.5 [41], which paves the way for the formation of the third cusp row starting at ED17.5. Although it has not be put in these terms, this could be interpreted as a heterochrony, whereby a process shared by the two molars persists longer in the upper germ and is responsible for the supplementary cusp row. Developmental genetics revealed a few genes specifically expressed and necessary in one type of molar, and these are all lower-molar specific genes (Dlx5 and Dlx6 [22, 24], Nkx2.3 [28, 42], mesenchymal expression of Pitx1, [43]). On top of that, a number of genes expressed in both types of molars show differential requirements (Activin betaA [44], Bmp4 [45], Runx2 [46], Follistatin [47], Dlx1-2 [48, 49]). In some extreme cases, tooth development is arrested early in the mutant condition for one type of molar, but proceeds normally for the other type (e.g. Activin betaA, Bmp4, see references above). This indicates that the gene networks, although shared, can adopt very different states during lower and upper molar development. In line with this idea, a microarray study revealed numerous differentially expressed genes, including genes of key developmental signaling pathways and homeobox transcription factors [28]. However, this study was based on a single, and quite early, time point of the molar morphogenesis (ED14.5).

In this context, we took an original bioinformatic approach to compare the transcriptomic dynamics of the lower and upper molar germs, during the crown morphogenesis period (Fig 1), that is the period when lower and upper molar form, sequentially, their 6 and 8 cusps, respectively. In order to match gene expression dynamics with cusp formation, we not only sampled closely spaced RNA-seq transcriptomes (8 stages) but we also followed cusp formation using a cusp marker whose expression was evidenced in situ in the tooth germ. Our in-depth study of transcriptomic signatures provides evidence of upper/lower specific differences in widely shared developmental mechanisms, and this, at three different levels. i) First, the two developing teeth have a clear lower/upper transcriptomic identity carried by a large number of genes and maintained throughout molar development. This signature most notably stemmed from upper/lower biases in mesenchymal gene expression and relative abundance of the tooth mesenchymal component. This was consistent with experimental observations showing that the nature of mesenchyme and its abundance determine tooth final morphology ii) Second, we evidenced a transcriptomic signature of heterochrony between the two teeth and related it to differences in the proportion of the territory occupied by a particular tissue (later forming molar crown cusps). Given our knowledge of molar development, this suggested specific differences in the activation-inhibition mechanisms ruling cusp formation in the two teeth. iii) Third, we evidenced a transient exaggeration of expression profiles in upper molar for genes involved in mitosis, adhesion and migration. This counters the assumption that differences should accumulate during development and suggested transiently exaggerated morphogenetic processes. Taken together, this integrated transcriptomic analysis enabled us to formulate specific hypotheses of how the upper molar crown develops additional crown cusps. More generally, we note that all three transcriptomic signatures were shaped by differences in relative abundance of cell types within samples (i.e. proportion of mesenchymal/epithelial tissue, proportion of cusp tissue, proportion of mitotic or migrating cells), rather than to direct differences in gene regulation.

# RESULTS

In this study, we targeted the period of lower and upper molar development during which most of the crown morphogenesis, and notably cusp patterning, happen (Fig 1). We sampled lower and upper first molar tooth germs, in eight individuals, corresponding to eight consecutive stages each separated by 12 hours.

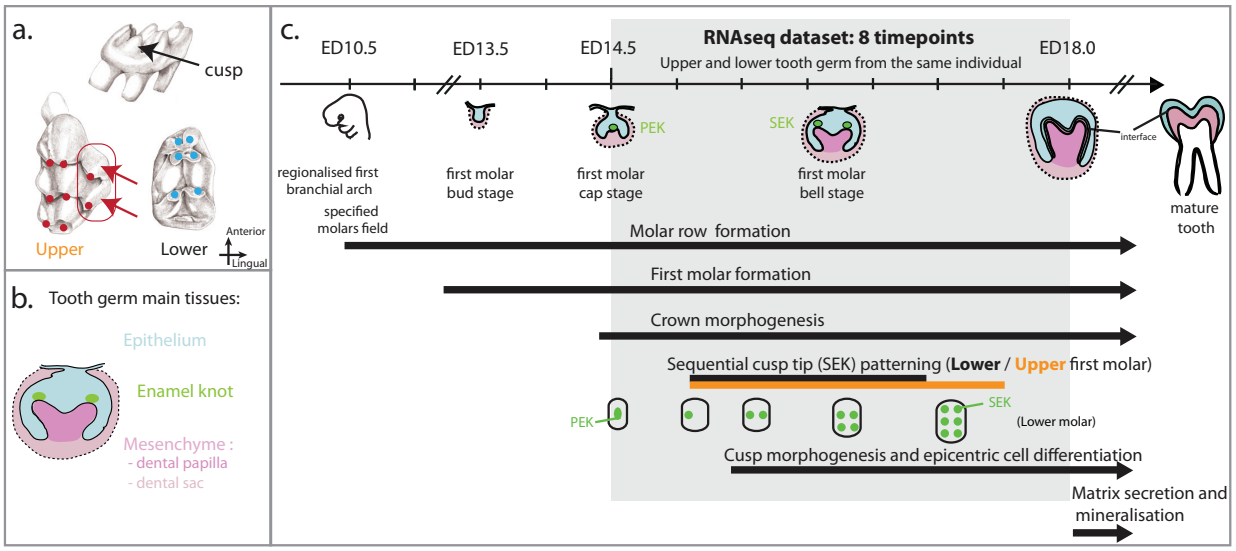



Figure 1 – **Mouse lower and upper first molars and their development.**
**(a)** Top : Lateral view of the mouse upper first molar showing its cusps. bottom: Crown views of mouse lower and upper first molars. Drawings after [101]. **(b)** Simplified view of the tooth germ, composed of three main tissues: mesenchyme, epithelium and (epithelial) enamel knot, which serves as an epicenter for epithelium differentiation. (c) Scheme summarizing the main steps of molar development. At ED10.5, the lower and upper molars presumptive fields are specified in regionalized first branchial arch. At this point, the development of the molar rows is starting. By ED13.5, the first molars reach the bud stage: the oral epithelium has invaginated and tops a condensing mesenchyme of neural crest origin. Epithelium-mesenchyme interactions then lead to the formation of a signaling center in the invaginated oral epithelium, called the primary enamel knot (PEK), which origin can be complex [102, 103]. The PEK directs the growth of dental epithelium around the further condensing and rapidly growing dental mesenchyme (bud-cap transition), resulting in the cap stage. This marks the beginning of the morphogenesis of the molar crown. Next, secondary enamel knots (SEK) are patterned sequentially. They seat at the tip of the future cusps and drive their morphogenesis, resulting in the bell stage. SEK patterning in lower molars is schematized with a domino, and the period of SEK patterning in lower (black bar) versus upper molars (orange bar) is shown. SEK also act as epicenters for the progression of proliferation arrest and differentiation. Epithelial and mesenchymal cells situated at the inner interface between these two compartments respectively differentiate into ameloblasts and odontoblasts, and at the very end of the fetal life, those cells will start producing enamel and dentin respectively. Note that the successive steps start sequentially but largely overlap because the downgrowth of the epithelium continues while cusps are patterned, and the differentiation of the first cusps starts before the last cusps are formed.

## Whole organ developmental transcriptomes carry an embedded signal of developmental timing.

As a starting point, we used multivariate analysis to extract the main axes of variation in gene expression within the time series for each type of tooth. We combined the lower molar samples, the upper molar samples, or all samples together. In each case, we found that the first axis represented about half of the total variation and ordered the samples according to developmental time (Fig 2a, Sup Fig 1). This time ordering was extremely robust and resisted sub setting data, either by randomly sampling genes, by taking only the bottom (or top) 20% expressed sets of genes, or by retaining only tissue-specific genes (see later in Sup Fig 6). It is not caused by the fact that the expression of many genes is switched on and off between consecutive stages. We found a set of 641 genes showing such drastic pattern (i.e. very weakly expressed at one stage and expressed at least twice more in the consecutive stage). The removal of all genes does not impact the strength of PCA1 (61%), and the ordering of the samples.

## Most of the transcriptome series is similar between upper and lower molars.

We next wanted to compare gene expression during lower and upper molar development. In a PCA combining lower and upper samples, the first axis of variation ordered samples with time (Fig 2a). So, the main axis of variation was common to upper and lower molars, suggesting that the development path is largely common to these two types of tooth. This developmental time axis represented about 52% of variance, that is a bit less than 61-64% of variance (which was observed in lower or upper-specific PCAs, see Sup Fig 1). This is however high, suggesting that the vast majority of time variation in gene expression is shared by the lower and the upper developing molars. This strong

similarity was also shown by the analysis of the main clusters of developmental expression time series. We extracted ten main developmental profiles in upper and in lower tooth (see Methods). These profiles are shown in Fig 2b, together with the number of genes that were assigned to each of them. Linearly increasing/decreasing profiles made up one third of the genes transcribed in each organ (34%, clusters 4 and 5 in lower tooth or clusters 7 and 5 in upper tooth).

Some of the clusters (1, 2, 4, 5 for lower tooth and 5, 7 for upper tooth) were strongly enriched in genes involved in development, morphogenesis, signaling and adhesion, cell proliferation and differentiation (Sup Fig 2) including genes that are annotated with especially relevant developmental processes, such as odontogenesis, skeletal system or appendage development. Clusters 9 and 10 of lower tooth (which have resembling profiles) and cluster 1 of upper tooth were enriched in genes involved in mitosis (e.g. GO mitotic nuclear division; note that this enrichment is much stronger and the profile much sharper for the upper cluster). The profile of this upper cluster 1 resembled upper cluster 4, which was enriched in genes involved in DNA metabolism and DNA replication, and may thus be the counterpart for the G1/S phase of the cell cycle. Taken together these clusters suggested that lower and upper tooth germs are depleted in proliferating cells at ED 15.5, an effect that was sharper in the upper molar. Other clusters show no general pattern of functional enrichment, but nevertheless revealed small groups of genes involved in specific processes (*e.g.* N-glycosylation for lower and upper cluster 3; cell migration and integrin signaling pathway for upper cluster 6; synaptic transmission for upper cluster 9, lamellipodium organization for upper cluster 10; inflammatory response for upper cluster 2). Finally, some clusters were only poorly enriched for specific functions (*e.g.* clusters 6, 8 of lower tooth).

We then used these clusters as a way to quantify the similarity between development in upper and in lower tooth. From the ten main representative profiles extracted from lower molar data, we retained only the genes which were robustly associated to one of these representative profile (see methods for the test employed), that is, 8,187 genes. Among these genes, we counted the number of genes which were associated together in the same profile in upper and in lower tooth. Such genes are expected to have more or less the same temporal expression profile in both organs. This was the case much more often than expected (Chi-squared test, P-value <10-16), and represented altogether 56% of the genes. The same result was obtained with the reciprocal analysis (Fig 2c-right), mapping expression from upper tooth to the lower tooth.

## Intrinsic difference in transcriptome between upper and lower tooth.

### An intrinsic difference in transcriptome between upper and lower tooth is carried by a large number of genes.

Although most of the gene expression was similar between upper and lower tooth, the differences were far from negligible. This was true in term of time profiles, as exemplified by the frequency of genes that do not map to the same cluster in upper and lower tooth (44%, Fig 2c) and by the importance of the second axis of the principal component analysis. This second axis of variation segregated samples between upper and lower molar, whatever the stage considered (Fig 2a) and represented 13.2% of the total variation.

We first wondered whether this pattern could be caused by specific or strongly biased genes that would mark a clear lower versus upper molar identity, as identified for early jaws [50]. A single gene, *Nkx2-3*, was specifically expressed in the lower molar, and we found no gene specific for the upper molar, although the top upper gene, Pou3f3, was about 10 times more expressed in the upper molar. If fact, when taking all samples from each kind of tooth in the large-scale dataset as replicates (so in total, 8 replicates for upper, and 8 for lower molar), we found 1,347 genes (out of 14,808) differentially expressed ("upper/lower DE genes", adjusted P-value $< 0.1$), out of which only 83 show more than a twofold excess difference. This included genes known for their role in molar or jaw specification (like *Dlx5* and *6*, *Pou3f3* and 2 associated ncRNAs, *2900092D14Rik*, *2610017I09Rik* [50], *Pitx1* [43]). We concluded that there was relatively few genes that were consistently biased with LR>2 throughout stages. On a developmental point of view however, these consistently biased genes were possibly sufficient to provide and sustain different orientations for upper and lower molar development. Are those consistently biased genes sufficient to explain the genomic signature that separates upper and lower tooth? To answer this question, we looked how far this genomic signature would resist the removal of differentially expressed genes. Removing the 83 above mentioned genes had a marginal effect on the second axis of PCA (11.4% of variation explained instead of 12.9% with the 83 genes). In fact, the second axis of the PCA still separated upper and lower samples and represented a significant amount of the total variation, even when all differentially expressed genes were removed: After removing 1347 DE genes that were found when the 8 stages are taken as replicates, the axis that splits upper and lower tooth represented 9.3% of the total variation. Upper/lower DE genes can also be estimated taking time into account (DESeq2, adjusted P-value<0.1), which is less stringent and resulted in 3155 DE genes : After removing these genes, the axis that splits upper and lower tooth represented 7.8% of the total variation. We concluded that the upper/lower transcriptomic signature was not only carried by sets of genes that are moderately to strongly biased throughout the developmental period, but

also by more subtle gene expression differences in a very large number of genes.

Figure 2 – **Lower and upper first molar germs share temporal dynamics of gene.**
**(a)** Map of 16 transcriptomes from upper (black) and lower (grey) molar germs, at 8 stages of development, on the two first principal components of a PCA analysis. PCA1 orders samples with time, hence represents shared temporal variation. PCA2 separates upper and lower molar germs. The percentage of variation explained by each axis is indicated. Lower and upper first molar pure-mesenchyme transcriptomes (pink) and epithelial cell transcriptomes (blue) projected on this map are segregating near upper and lower samples respectively. **(b)** Ten main temporal profiles of expression determined independantly for the lower (left) and upper (right) tooth. 10 clusters were obtained by K-mean clustering (see methods), and associated number of genes are indicated.Cluster numbering and color are arbitrary and do not indicate a correspondence between lower and upper clusters. Each profile represents the median of relative expression level (normalized with expression level at ED 14.5) for all genes associated to a particular cluster. **(c)** Similarity of expression dynamics between upper and lower tooth. Left : The height of each bar corresponds to the number of genes which lower timeprofile is robustly assigned to one of the 10 main "lower" clusters (Pearson R>0.7). In black, the number of genes whose timeprofile in upper molar associates best (and with R>0.7) to the same lower cluster: such genes have similar timeprofile during lower and upper molar development. In dark grey, the number of genes for which expression profile in upper data associates best (and with R>0.7) to another lower cluster. In light grey, the number of genes, for which the upper expression profile is not well associated to any lower profile (R<0.7). Right: the experiment was done in the opposite direction, starting with genes which upper timeprofile is robustly associated to a given upper cluster and challenging their lower timeprofile.

**Upper/lower transcriptomic differences are partly driven by differences in proportions between tooth epithelial and mesenchymal compartments.**
Because experimental work has suggested that the mesenchyme carries molar identity at the examined stages [51], we wondered whether the above lower-upper DE genes would rather be mesenchyme-specific genes. We took benefit of a previous study that had sampled epithelium and mesenchyme transcriptomes separately at an earlier stage [39] to attribute a tissue specificity score, by simply calculating a mesenchyme/epithelium ratio. Then, we extracted stage-by-stage differentially expressed genes, by considering 7 pairs of consecutive developmental stages as replicates (see methods). We obtained 1,146 genes that differed between upper and lower tooth for only a portion of the whole time-serie (1 or 2 such "pairs of consecutive stages"), 1,646 genes that differed at in all comparisons (7 such "pairs of consecutive stages"), and 10,933 genes that never differ significantly, in any pair comparison. We found that genes that differ between upper and lower molar were on average significantly biased towards mesenchyme specificity (Wilcoxon test, P-value < 10e- 16, Fig 3a in white). When we further separated these sets in two categories, depending whether they are enriched in lower or in upper molar (genes assigned to "lower" in grey, or "upper" in black, in Fig 3a), we were however surprised that the tissue specificity of lower and upper genes was on average significantly different: upper genes are clearly more mesenchyme-specific than lower genes (Fig 3a, Wilcoxon test P-value < 10e-16 in all comparisons of mesenchyme ratio for the 3 gene sets). This was even more pronounced for genes that were biased at all stages, but interestingly, this was also true, although to a minor extent, for genes that were not significantly differentially expressed (Fig 3a, P-value < 10-16). Such bias impacting a large part of the transcriptome including non DE genes could be explained if the proportions of epithelium and mesenchyme differed between lower and upper molar, the lower molar being enriched in epithelium and the upper molar enriched in mesenchyme. We performed deconvolutions (see methods) using three sets of marker genes (markers from bite-it [52], 11 genes for enamel knot –EK– , 12 for epithelium, and 14 for mesenchyme) to estimate the relative proportions of mesenchyme, epithelium and enamel knot in the molar germs. We show that the estimated proportions of mesenchyme were systematically higher in upper teeth, especially in earlier stages (Fig 3b, the difference was significant at each developmental stage, obtained by paired Wilcoxon tests on bootstrapped values, P-values < 10-16). Similar results were obtained from another set of markers defined using microarray data (Sup Fig 3). Such a difference had not been pointed before, thus we performed a direct quantification of epithelium and mesenchyme compartments on histological sections taken throughout development and we found that the upper molar was indeed enriched in mesenchyme as compared with the lower molar throughout the time window of interest (Fig 3c, Wilcoxon test, P-value = 2*10-5). These results thus highlighted that the upper molar germ was enriched in mesenchyme relative to the lower molar germ throughout the period of cusp formation.
Next we reasoned that such a pervasive effect on the transcriptome, seen even for non-DE genes, could typically shape the second axis of PCA (PCA2). This prompted us to examine the mesenchyme specificity of the genes contributing to this axis. One main advantage of PCA is that it is possible to extract coordinates for each gene. Genes with large coordinates contribute the most to the axis. We found a clear correlation between the coordinates in PCA2 and a mesenchyme/epithelium ratio extracted from published tissue-specific expression obtained at earlier stages using microarray data (Fig 3d, [39, 51], Pearson R = 0.280, P-value <10-16). Thus genes co-localizing with upper molar tend to be mesenchyme- specific, while genes co-

localizing with lower molar tend to be epithelium-specific, as seen above. Finally, it was also possible to map external datasets on a PCA. In such mapping, pure tooth mesenchyme RNA-seq data (taken from upper and lower tooth from stage ED13.5 [45]) co-localized with early upper tooth samples when they were projected on the PCA plot, while epithelial cell lines (not even from tooth nor mouse : human mammary epithelial cells, [53]) located on the side of lower tooth samples. All these data illustrated that PCA2 discriminates genes with epithelial and mesenchyme biases.

We then wished to evaluate how far differences in epithelium-mesenchyme ratio take part in shaping expression differences between lower and upper molar by two complementary approaches. First, we removed the genes whose expression is strongly biased in mesenchyme or in epithelium pure tissue transcriptomes (From [39]) We retained 9,263 genes whose epithelium/mesenchyme ratio is less biased than the 3rd quartile. Using this set of genes to draw a PCA still separated upper and lower tooth on the second axis, but the importance of this axis decreased (this second axis represented 10.8which was slightly less than 13% on the total set of genes). Second, we simulated a reasonable increase of mesenchyme proportion in a lower ED 14.5 sample (See methods), and mapped this artificial sample on the PCA. This artificial sample was shifted along the second axis, however it was not sufficient to morph a lower tooth into an upper tooth: it covers only 23 à 35% of the upper/lower distance on axis 2 between lower and upper tooth germ at this stage. We concluded that pervasive differences in gene expression due to tissue proportion were clearly taking part in shaping axis PCA2, although they were not sufficient to explain it.

Figure 3 – **The upper molar germ is enriched in mesenchyme.**
**(a)** Boxplot showing the distribution of an index of mesenchyme-specificity, for the totality of genes (white) and for genes more expressed in upper (black) or in lower (grey) molar germ. This index was measured as the ratio of expression levels between mesenchyme and epithelium (median of log ratio(mesenchyme/epithelium) at 12.0, 13.0 and 13.5 dpc, estimated using microarray data [39]). non-DE: genes non-significantly biased between upper and lower toot. All : genes significantly biased for at least one stage. DE 1 or 2 stages: genes significantly biased for one or 2 developmental stages. DE all stages : genes significantly biased for all developmental stages. **(b)** Mesenchyme proportions estimated from deconvolutions for our 8 developmental stages. Markers were obtained from bite-it database (11 for EK, 12 for epithelium and 14 for mesenchyme). Confidence intervals were extracted by resampling (500 bootstraps). RNA-seq samples of mesenchyme isolated from lower and upper molar germs [45] were used as positive controls. **(c)** Boxplot showing the difference between mesenchyme proportion between 24 pairs of upper and lower molar germs, taken between ED 14.5 and ED 17.5. Mesenchyme proportion is expressed as a percentage of mesenchyme volume relative to total germ volume, as measured on complete series of histological sections. (d) Map showing the location of the genes on the two first principal components of a PCA analysis obtained from upper (black) and lower (grey) molar germs, at 8 stages of development. The interpretation of PCA axes (respectively developmental timing for PCA1, and upper/lower for PCA2) is reminded with arrows. Genes are colored according to their tissue-specificity in a tissue specific dataset ([39]; expressed mesenchyme>epithelium in pink; expressed epithelium>mesenchyme in blue).

**Upper/lower transcriptomic differences are also driven by differences within compartments.**

We then reasoned that pervasive differences in gene expression not due to tissue proportion, but constantly opposing lower and upper molars, should also be responsible for shaping this axis 2. To get rid of tissue proportion effects, we again made use of a previously published RNAseq dataset of pure mesenchyme. This was obtained from laser dissected tissues, in a lower molar germ that is one day younger than the youngest sample of our dataset (ED13.5, that is, before cap stage, [45]). As seen in the previous section, the upper and lower mesenchyme samples mapped correctly on the second axis of the PCA (Fig 2a). They were more closely mapped than samples extracted from whole tooth germ, and their distance only corresponded to 22 to 41% of the upper/lower distance obtained on whole tooth germ samples (a more realistic estimate, if we consider only early stages of development, would be closer to 25% of the upper/lower distance). This confirmed that the transcriptomic differences shaping PCA2 are not only a matter of epithelium versus mesenchyme proportion, but also (at least) of mesenchyme quality. Moreover, this suggested that these differences are already established before cap stage, as the mapped transcriptomes were younger (13.5 days). We have shown earlier that there was a set of constantly biased genes (1347 genes with upper/lower bias), most of them being only slightly to moderately biased. The data above suggested that, even if part of the upper/lower difference is driven by epithelium- mesenchyme proportions, another part is related to a different mesenchyme quality in upper compared to lower molar. To confirm this, we performed two additional experiments. When performing a new PCA after removing those 1347 genes from our dataset, and mapping again the pure mesenchyme samples, the distance drops to 8% on average (median: 8.3%, from 5.2 to 10.1%) of the upper/lower distance obtained on whole tooth germ samples. Reciprocally, performing the same experiment on a dataset consisting only of these 1347 genes, the distance between pure mesenchyme sam-

ples increases to 46% (median : 46%, from 32.4% to 59.3%) of the upper/lower distance obtained on whole tooth germ samples. The sensitivity of these mapping experiments to this set of 1347 genes was a good indication that this set is partly shaped by constant upper-lower differences (at least) in the mesenchyme, already seen at ED13.5. Said differently, part of the lower-upper variation of gene expression making up this second axis was also found in these mesenchymal datasets. It could represent one fourth to one third of the transcriptome difference between upper and lower whole tooth germs and would therefore be attributable to differences in the level of expression, irrespective of differences in tissue compartment proportions.
In summary, our data suggest that lower and upper molar transcriptomes exhibit clear signatures of identity: "lowerness" vs "upperness". This signature is partly made by differences in tissues proportions, the upper molar germ being enriched in mesenchyme as compared with the lower molar germ, but also by differences in expression within tissues (at least mesenchyme), which are established as early as before the cap stage (ED13.5).

## Early peak of the difference in transcriptome between upper and lower tooth.

We noted on the PCA plot (Fig 2a) that the distance between lower and upper samples on axis 2 is maximal at ED15.0-ED16.0 (40-60% more than the average distance), which is also true when differentially expressed genes are removed. This suggests that some of the gene expression differences driving axis 2 are exacerbated during this period. A simple plot of the number of differentially expressed genes along development (Fig 4a) also shows such transitory excess. “Upper genes” (genes with a significant positive upper > lower difference) are overall more numerous than “lower genes”, but the underlying distributions are different for genes differentially expressed at a single time point, and for other genes. Finally, for genes differentially expressed at all time points, “lower genes” are very rare compared to “upper genes”.
We also designed a method, based on hidden Markov Models, to track differences in time profile, in a way which explicitly models the temporal series (see methods). This models, for each gene, the variation of the magnitude of upper/lower ratios along time. More specifically, we considered two states, one with “small” and one with “large” upper/lower ratio. After modeling, each gene corresponds to a succession of 8 states, corresponding to the variation of upper/lower ratios during the time course. This kind of modeling has a property of smoothing the time profiles (because the state probabilities in consecutive stages are not independent), rendering the results robust to artifactual changes of expression in a particular time point. In Fig 4b, we counted the transitions for each pair of consecutive stages in our time series, that is, how many genes switch from one state of upper/lower ratio to another. Although the whole modeling is symmetrical along time, the resulting pattern is clearly asymmetrical (Chi-squared test, P-value <10-16). Most transitions “small”-”large” were found between the first two time points, and on the contrary most transitions “large”-”small” were found between stages ED16 and ED16.5. This again emphasizes that upper/lower differences were time dependent, with a transient excess of upper/lower difference.
To understand what drove this transient excess, we retrieved 739 genes that were differentially expressed between lower and upper molars at ED15.5-16.0 but not at other time points (adjusted P-value $< 0.1$ when both timepoints are considered replicates, and significant in no more than one other time point). These genes were enriched for genes involved in mitotic cell cycle process (regulation but also accomplishment, e.g. formation of spindle...), but also in cell adhesion and cell migration (notably *integrin* binding), signaling (notably Wnt signaling pathway, with 3 ligands and 2 antagonists), ossification related process and programmed cell death. We classified these genes as lower or upper genes (Fig 4c) and compared their relative enrichment for GO processes and functions (Sup Fig 2). Lower genes (and most notably the 200 genes transiently down regulated in upper tooth, Fig 4c) were strongly enriched in genes involved in cell cycle and mitosis (e.g. “cell cycle process” 5.10-57; spindle organization 1.10-18). This was reminiscent of the enrichment seen in upper cluster 1 (see above Fig 2b), and again pointed an exaggerated decrease in proportion of mitotic cells in the upper molar germ at ED15.0-16.0. Conversely, upper genes were enriched in genes involved in cell migration/locomotion.

## Origins of the transcriptomic delay between the two molars.

### The transcriptomic delay of the upper molar is evidenced with several methods.

We noticed that the first axis of the PCA also tends to segregate lower and upper samples, although in a less intuitive manner than the second axis, since upper samples mapped systematically ahead from their lower counterpart (Fig 5a). Because the upper and lower tooth at each developmental stage were taken from the same embryo, this cannot be due to sampling effect, and rather suggests that upper tooth germs show a delay in shared developmental variation of gene expression. For instance at ED 15 we estimated the offset between the two teeth worth 3.5-7.5 hours of development (from simple ratio of the distances on the first axis, with consecutive development stages in upper or lower tooth being about estimated at 12 hours). We later refer to this delay as an “heterochrony” between upper and lower tooth development. Then we wished to estimate this heterochrony using a completely independent method, based on expression profiles. It is not possible, for genes with a linear profile of expression, to disentangle shifts in expression levels (heterometry) and shifts in expression timing (heterochrony) on an individual basis. Nevertheless, on a statistical basis, if the upper germ is delayed, then linear genes increasing with time should tend to be more strongly expressed in lower

germ whereas those decreasing should tend to be more strongly expressed in upper germ. This was indeed the case when taking genes that vary significantly during development, in a linear manner (comparison of 3,281 genes whose expression level is increasing linearly, with 3,181 genes which are decreasing linearly : Wilcoxon test, P-value < 10-16; DE-genes with a significant time-related bias, adjusted P-value < 0.05, DESeq2).
To go further, we decided to focus on genes for which the expression profile is a bell curve, with a single extremum of the same kind (either a maximum or a minimum) (Fig 5b). We found 1,029 such genes. We then time-shifted the expression profile for the upper tooth, so that to match the lower profile as well as possible (using a simple translation, see methods). The distribution of the translation factors was significantly skewed towards positive values, in accordance with the hypothesis that the upper tooth is overall delayed by comparison to lower tooth (Fig 5c). We then sought to quantify the relative importance of heterochronies in the upper/lower difference. For that purpose, we quantified the upper/lower distance before and after time-translation. The ratio indicates that on average, 32% of the difference between upper and lower (bell-type) profiles is attributable to simple heterochronic shift (median = 29%, Fig 5d ). We concluded that a substantial part of gene expression differences (one fourth to one third) between lower and upper tooth germ can be interpreted as delayed changes in gene expression in the upper molar.
The period of interest is marked by crown morphogenesis, including cusp sequential addition and the initiation of their differentiation. Heterochrony in this process could thus logically result in transcriptome heterochrony, and especially along PCA1, since as seen earlier, enamel knot marker genes are among strong contributors to this axis. To try to understand this heterochrony in expression, we first examined the link between incremental cusp patterning and PCA1. We documented the temporal dynamics of cusp addition in the two molars, using the cusp tip marker *Fgf4* (*in situ* hybridization data; see Fig 6a for lower and Sup Fig 5 for lower/upper data). The dynamics is clearly different between upper and lower molars (Fig 5e). However, from these data, one does not intuitively expect a systematic delay of the upper molar, like the one seen in expression data, which starts from early stages and continues throughout the development.

### The transcriptomic delay between the two molars relates to the proportion of "cusp tissue".

To better understand this pattern of upper/lower heterochrony, we needed first to better understand the link between PCA1 and the number of cusps for one tooth germ, for example the lower one. The number of cusps that are patterned at each developmental stage was correlated with the coordinate on PCA1 (Fig 6d top: P-value associated to a Poisson model in which PCA1 explains cusp number = 0.003, McFadden's pseudo R-squared = 0.29). This is a good correlation of course, but the number of cusps does not increment at each developmental stage (eg. between 14.5 and 15.0 or between 15.5 and 16.0), and yet the transcriptome captures time differences between those stages. This is not surprising as correlating with absolute cusp number is over simplistic for two reasons: first, a transcriptome made from a whole tooth germ can only record a proportion of tissue engaged in cusp formation relative to the rest of the tooth germ, and not absolute number of cusps. Second, cusps and in particular cusp epithelium, are known to mature progressively from tip (the so-called enamel knots, see Fig 1) into the valleys (starting with cell cycle arrest and much later with ameloblast differentiation), and this process starts soon after the enamel knots are patterned [54]. We thus anticipate that the patterning of enamel knots should be very rapidly followed by conversion of naïve tissue to "cusp tissue" from the tip into the valleys, and this would also change cusp proportions. Based on these assumptions, we built a simple model for quantifying "cusp tissue" patterning and expansion in time, fueled with our *Fgf4 in situ* hybridization results on time of cusp tip (secondary enamel knot) patterning. For simplicity, we chose to model cusp proportions with a simple grid model. This is still oversimplified as there is no pre-pattern in tooth, and cusp patterning occurs concomitantly with some tooth growth. For simplicity also, once patterned, the territory of each cusp expands at the same speed (Fig 6c). Despite this simplicity of the model, its application to lower tooth fits better the advancement of development monitored by PCA1 than the simple number of cusps, albeit in a non-linear manner, which was expected given that we estimate proportions (Fig 6d bottom; P-value associated to a binomial regression in which PCA1 explains cusp proportion = 10-12, McFadden's pseudo R-squared = 0.42). This shows that the overall proportion of the territory occupied by cusp tissue, very simply explains the major component of the transcriptomes.
We then sought to understand whether this main model could explain the offset observed on the PCA1 between upper and lower tooth. We applied a similar modeling to the upper tooth, following the sequence of cusp addition that ranges from 1 to 8 cusps, because the upper tooth adds 2 more cusps at the end of the developmental sequence (ED17.5-ED18.0, Fig 6e and Sup Fig 5a; P-value associated to a binomial regression = 10-24, McFadden's pseudo R squared = 0.71). We fueled it with *Fgf4* in situ hybridization data obtained for upper molar (see methods), since the dynamics of cusp patterning is specific to each molar (see above Fig 5e). As such, the model intrinsically implies that in the early steps, the cusp tissue will occupy a smaller proportion of the total tooth germ in upper as compared to lower tooth and also that cusp expansion rate is similar between lower and upper molars. This may not be realistic, and therefore we examined three alternative models (Sup Fig 5b-c-d). A first model, in which the two supplementary cusps are omitted (implying that the formation of extra cusps does not translate in a change in cusp proportions), a second model, in which upper tooth is progressively acquiring the extra space cor-

responding to the extra cusps, and a third model, in which the rate of cusp expansion is different in upper as compared to lower tooth. With respect to the overall fit to PCA, none of these alternative models performs better than the main model, and they are all worse considering specifically upper/lower heterochrony (see details in the legend of the Sup Fig 5).

The offset between upper and lower tooth at the same stage on the PCA (which we called earlier heterochrony) is well correlated with the offset of cusp proportions computed from the model (R2 = 41%, Fig 6e, model 2). This correlation becomes nearly perfect (R2 = 95%, P- value = 10-4) when the last stage (ED 18) is omitted. This suggests that our model, in which cusp expansion is infinite, is not adequate at this late stage, possibly because cusp expansion is in fact finite. We modified our model in accordance with this hypothesis, by reducing the time for complete expansion attributed to each cusp. We reduced this time, from 8 steps (corresponding to the entire time series, therefore to the non-finite expansion model mentioned above) down to 2 steps. The pattern of heterochrony observed with the PCA is best recapitulated by a finite model in 5 steps, ie which allows a duration of 2 days and a half for each cusp to expand to its final territory (Sup Fig 5e).

A simple model of "cusp tissue" proportions in upper and lower molar germs was thus sufficient to recapitulate the main axis of variation observed at the transcriptome level and the heterochrony between the two molars. However, at this stage we did not make explicit what could be this "cusp tissue" whose expansion impacts the transcriptomes. In theory, both the epithelial and mesenchymal component may be concerned but at least this "cusp tissue" should include the secondary enamel knot (SEK), and the neighboring epithelial tissue, which gets patterned soon. This process is exemplified by Shh expression, which is stronger in the SEK but spread in a much larger territory of dental epithelium around ([55] and `http://bite-it.helsinki.fi/` [52]). In the bite-it database, we screened for other genes which expression territory is first confined to the primary enamel knot (cap stage) and later (bell stage) expanded to a larger epithelial territory around the SEK. We thus defined 6 markers of what we refer below as "epithelial cusp tissue" ( *Shh, Bmp7, Edar, Fgf9, Cdkn1a, Wnt10b*). With this set, we performed deconvolutions to monitor the progression of epithelial cusp tissue proportion in lower and upper molar germs. As expected, the proportion is increasing with developmental stage in upper and lower tooth, and epithelial cusp tissue proportions themselves are also very well correlated with PCA1 (P-value = 10-8; R2 = 0.91; Fig 6f). epithelial cusp tissue proportion are comparatively smaller in upper than in lower molar germs (Fig 6g). The upper/lower offset in the advancement of the maturation estimated by our model was nearly perfectly correlated with epithelial cusp tissue proportions estimated by deconvolutions (P-value = 10-12; R2 = 0.97). The offset was small at the beginning of the developmental series, then tends to increase with developmental time, and is reduced again at the end. This further suggested that the upper germ transcriptome may look delayed because less "cusp tissue" has formed, and therefore genes expressed in this "cusp tissue" were less represented in the whole tooth germ.

Figure 6 – **Cusp proportions estimated with a simple model recapitulate time ordering on first PCA axis and upper-lower heterochrony.**
**(a)** Cusp patterning in the lower molar, from 1 to 6 main cusps, which corresponds to its adult morphology. The timing and sequence of cusp addition reflects observations based on the analysis of in situ hybridization data using the cusp tip marker Fgf4 (see methods and Fig 5e). **(b)** Model of cusp proportion for the lower molar without maturation. The contribution of each cusp to a final proportion of "cusp tissue" (represented by the last stage) is modeled with a simple grid. In this model, the total contribution is achieved as soon as the cusp has been patterned. Time for cusp patterning is derived from bench data in (a). **(c)** Model of cusp proportion for the lower molar with maturation. In this model, the contribution of each cusp to a final proportion of "cusp tissue" increases following patterning. "Cusp tissue" expands at the same speed in all cusps. The average shade of grey of grid unit, at each developmental stage, gives a visual impression on the average degree of the expansion of the cusp tissue territory, from unpatterned cusp (white) to just-patterned cusp (light grey) and fully expanded cusp (black). We computed cusp expansion in each tooth based on this model, 0 corresponding to no cusp patterned (all white), and 1 to a complete expansion (all black). **(d)**-Top: Relationship between the number of cusps, as shown in (a), and the coordinate on the first PCA axis (developmental stages are indicated). Dashed lines: glm model in which PCA1 explains cusp number (corresponding P-value = 0.003). Bottom: Relationship between cusp proportions, estimated using the model 2 (shown in c), and the coordinate on the first PCA axis (developmental stages are indicated). Dashed lines: glm model in which PCA1 explains cusp proportion (corresponding P-value = 1e-12). **(e)** Lower / upper difference as estimated in terms of cusp number (as in a.), cusp proportion estimated with model 1 in b, cusp proportion estimated with model 2 in c and coordinates on PCA axis 1. Model c recapitulates best the heterochrony measured on PCA1. **(f)** Relationship between the proportion of "cusp tissue," estimated using the maturation model in (c), and the proportion of "epithelial cusp tissue" estimated by deconvolutions. 6 markers from bite-it ( Bmp7, Edar, Fgf9, Cdkn1a, Wnt10b, Shh) were chosen, because their expression territory is large and therefore may well reflect the advancement of cusp expansion. (g) Proportions of "epithelial cusp tissue" estimated by deconvolution as in (f) for upper (black) and lower (grey) samples, at each developmental stage. RNA-seq samples of mesenchyme isolated from lower and upper molar germs [45] serve as negative controls ("mes").

217

**Mosaic heterochrony in different tissue compartments.**

Previous section of the paper suggested that most of the information concerning development time in the transcriptome was carried by "epithelial cusp tissue" proportion. Upper and lower heterochrony seemed also easily explained by differences in the advancement of "cusp tissue" proportion between upper and lower tooth. However, bell- shaped genes also show heterochronies and also order samples on PCA, albeit their bell profile is not directly explainable by a steady progress in cusp formation. In consequence we can imagine that time information may also be carried by other tissues, e.g. mesenchyme and non-cusp epithelium. To get hints into this question, we selected roughly the same number of marker genes for three tissues: enamel knot (421 specific genes), epithelium (336 specific genes) and mesenchyme (566 specific genes). We then first asked, whether each set of marker is carrying time-related information, that is, whether the 3 separate PCAs made using each tissue-specific set of markers order samples on their first axis (PCA1). The answer is yes for each "tissue-specific PCA", although the percentage of variation associated to each of these first axes is varying. As expected, PCA1 is strongest for EK-specific PCA (made with EK markers, 68.4% of variation), but mesenchyme-specific PCA (57% of variation) and epithelium-specific PCA (50.6% of variation) also carry a strong time-related signal (Sup Fig 6). The three tissue-specific PCAs also carry a signal of heterochrony. EK- specific PCA carries a heterochrony (up to 15% of the total distance ED 14.5-ED 18) which is best correlated with the heterochrony measured on total PCA (63.0%, P-value = 0.02), and reassuringly, with differences in EK proportions measured with deconvolutions (63.0%, P- value = 0.02). Epithelium-specific PCA carries very little heterochrony, but very interestingly, mesenchyme-specific PCA carries a very strong heterochrony (up to 40-50% of the temporal signal of mesenchyme). Contrarily to the pattern of heterochrony carried by EK, the pattern of heterochrony carried by mesenchyme is quite stable throughout the time period and starts in early development. Using a trick of normalization (see methods) we sought to disentangle the influences of tissue proportions and changes in cell types in time-related effects. As expected, time- related effects in EK seem to be carried both by tissue proportions and by cell-specific effects (PCA1 : diminishes from 68.4% in non-normalized PCA to 59.5% in normalized PCA). In mesenchyme, the magnitude of time-related effects is a bit less impacted by normalization (PCA1 : decreases from 62% to 57% for mesenchyme), and in epithelium there is hardly an effect (from 50.6% to 49.9%). This meant that temporal effects for in mesenchyme are not much carried by tissue proportion, but rather, mainly, by tissue maturation (see also deconvolutions of mesenchyme proportion in Fig 3b which are quite stable with time). From the same tissue-specific PCAs we observed that the patterns of upper/lower heterochrony in EK and mesenchyme stand against normalization, which means that heterochrony is also carried by tissue maturation (Sup Fig 6).
We conclude that the transcriptomic delay of the upper molar seen from the earliest stage on is the result of at least two phenomenon: i) a smaller proportion of "epithelial cusp tissue", ii) a slower maturation of EK and mesenchyme tissues. An interesting possibility would be that this delay in mesenchyme maturation is linked with the patterning of 2 extra cusps in the upper molar at ED17.5-18.0 (Fig 5e). This apparent delayed maturation might reflect a developmental mechanism whereby progression of differentiation is delayed, thus lengthening the phase when the germ grows and new cusps are patterned.

# DISCUSSION

This study asks the question which transcriptomic differences distinguish the development of two serial organs with different morphologies: the lower and upper first molar of mouse. We identified a shared developmental program despite heterochronies (temporal shifts) in this program. These are common findings in comparative developmental biology, but usually made from a limited number of genes. Here we show this is mirrored in the whole transcriptome dynamics. Furthermore, we revealed several collective gene expression differences, which we call transcriptomic signatures. First, we evidenced a transcriptomic signature for lower-upper molar heterochrony that we could relate to differences in proportions of the tissue later forming the crown cusps. Second, we found a transient signature due to exaggerated expression profiles in the upper molar, early during crown molar morphogenesis. Related to this, transcriptomic differences did not accumulate as in a funnel model, but peaked early on during molar crown morphogenesis. Third, we evidenced a clear lower/upper transcriptomic identity carried by a very large number of genes, for which individual variation was generally small but consistent over developmental stages. This signature stemmed from upper/lower biases in tissue composition of the tooth germs, although expression within individual tissues (at least mesenchyme) also participated. Below we discuss how these transcriptomic signatures highlight variations in a core developmental program that are possibly relevant for final phenotypic differences and evidence properties of whole organ transcriptomes applicable beyond our specific biological context.

**Transcriptomes reveal a core developmental program modified by cell proportions, heterochronies and inverted hourglass variation.**

It is commonly thought that serial organs develop using similar gene networks (*e.g.* limbs, [18, 20]). We already

knew that the two molars express more or less the same genes [28]. By showing that more than half of those genes have very similar temporal dynamics, we demonstrate that the two molars use similar gene networks, with similar temporal deployment. This identifies a molar core developmental program. Of note, some authors propose this is a criterion to establish serial homology [56]. This would replace the strict criterion of common ancestry, which implies that, in a distant ancestor, serial organs were identical and sharing identical gene networks. This is generally difficult to establish with certainty [57] - exactly the case for molars [21]. According to the extended definition of serial homology, our finding of a core molar developmental program establishes that lower and upper molars are serial homologs. In contrast to these shared aspects, the final differences in upper/lower morphology imply that the two molar germs exhibit extensive differences during morphogenesis. We evidence a clear genomic signature of lower/upper identity, consistent throughout very different periods of molar development (bud to end bell stage). We dissected this genomic signature and showed that this identity is influenced both by the proportion of the different tissues composing the tooth germ (epithelium/mesenchyme) and the tooth specific nature of tissues (at least for mesenchyme, since this was not specifically examined for epithelium). A recent study evidenced that organ-specific transcriptomic signatures in recently specified organs can be accounted for by a limited number of master transcriptional regulators, specifically expressed in the organ and activating a vast amount of target genes [7]. A similar effect may be at work here, and could rely on the few specific or markedly biased transcription factors already known or identified here (*e.g. Nkx2.3*, reported previously in molars [42], *Pou3f3* and two genetically linked ncRNA (*2610017I09Rik* and *2900092D14Rik*, reported previously only in early jaws [50], or the identity genes Dlx5 and Dlx6 that are no longer specific for lower molar but expressed 2 to 3 time less in upper molar). This could be further investigated in the future. It remains that most of the transcriptomic identity is made by differences in mesenchyme proportion within the germ (Fig. 7a). We revealed this by transcriptomes analyses, and then confirmed it at the bench: we quantified the respective volumes of the mesenchyme/epithelial compartments in 3D reconstructed tooth germs (dissected with the same procedure as the RNAseq samples, sup Fig 4), and the respective areas of the two compartments delineated on histological sections (which rules out the possibility that it could be a dissection artifact). The upper molar showed a larger proportion of mesenchyme from ED14.5, a very early step for crown morphogenesis. This finding is interesting in light of a tooth engineering study [58]. When an isolated ED14.5 epithelium was allowed to develop on a pellet of dissociated mesenchymal cells, tooth and cusp formation were obtained, and the number of cusps formed increased with the number of mesenchymal cells. A provocative scenario could be that the excess of mesenchymal cells present since ED14.5 in the upper molar is sufficient to raise the levels of cusp-inducing molecules secreted by the mesenchyme (such that Activin, [55]) and favors the formation of additional cusps. In this case, tissue architecture, without any need for differential gene regulation within cells during this period, would be one determinant of upper-lower differences. It is ironic that cell proportions play a major role in shaping the genomic signature of lowerness/upperness, such that our study initially focused on genes reminds us not to set cells aside: a view centered on cascades of gene expression is insufficient, because both molecular and cellular differences feed off one another over the course of development.

Another obvious factor well known in comparative developmental biology is difference in the relative timing of developmental processes (heterochronies). Heterochronies have been repeatedly pointed out in studies comparing serial organ development (and more generally in comparative developmental biology) and to date have somehow impeded the comparison of transcriptomes between serial organs [59]. Our data are consistent with this as heterochronies explain a substantial part of expression profile differences (25%). We moreover evidenced a transcriptomic signature of heterochrony: the upper molar is systematically "younger". To our knowledge, no study has ever reported a delay in the upper molar development during the period we examined here. Nevertheless, it is consistent with a delay reported for stages older than our oldest timepoint, in the doctoral work of Moullec [60]. He showed that odontoblast differentiation and enamel secretion, two processes that start at end of the fetal period, become visible one day later in upper molar compared to lower molar. Moreover, it would not have been expected from standard comparative developmental biology data (see our Fgf4 data, figure 5). The transcriptomic delay is seen from ED14.5, before cusp addition starts. Importantly, it is not due to a constant offset in the timing of development of the two molars, since cap transition, which is seen at late ED14.0, is synchronous in both molars (data not shown). It is thus far from trivial. Nevertheless, the heterochronic shift measured at each timepoint was recapitulated in a simple model predicting proportions of cells engaged in cusp formation from timing of cusp formation estimated at the bench with the Fgf4 marker. According to this model, from ED14.5, the proportion of cells engaged in cusp formation is smaller in the upper molar than it is in the lower molar. This finding suggests a mechanistic difference in cusp formation between the two molars, possibly in relation with the formation of two additional cusps in the upper molar. Cusps are known to inhibit the formation of new cusps in their vicinity [34]). From early on in crown morphogenesis, a decreased cusp proportion will reduce the inhibition on the unpatterned part of the tooth germ. This mechanism may promote the upper molar making ultimately more cusps. This could be achieved by change(s) in the growth rate and/or activation-inhibition mechanisms that together rule cusp addition [34, 61]. Heterochronies are further discussed in the last section.

We expected the temporal structure of expression differences to culminate in the latest stage. This intuitive idea

is integrated in two models of interspecific variation in embryogenesis: the funnel-like model where differences in gene expression increase linearly, and the hourglass model where differences are minimal at mid-embryogenesis but increase afterwards (*e.g.* [11, 12, 62-65]). In our case, we had additional reasons to expect such a pattern. The supplementary cusp row of the upper molar starts forming at a time when all main lower cusps are patterned (ED17.5, Fig 5d). It follows lingual extension of the upper germ beyond ED16.5, at a time when this process has stopped in the lower germ [41]. This prolonged duration in the upper molar should logically emphasize differences with the lower molar late in development. However, all transcriptomic data analysis (PCA, DE genes, Markov modeling) pointed to an early peak of differences (ED15.0-16.0), in conflict with the funnel-like model. The transcriptomic variation follows what we call an inverted hourglass model. This observation suggests that differences in developmental processes can be much sharper in early stages than commonly thought and can peak before partially resuming.
Why then such an early peak of differences? This ED15.0-16.0 period may be critical for the phenotypic differences between the two molars. The sharper gene expression profiles seen for upper molar at ED15.0-16.0 indicated that it experiences exaggerated processes (see Fig. 2). Interestingly, genes differentially expressed at this period included genes involved in cell adhesion and cell migration, suggesting that in the upper molar, a higher proportion of cells are engaged in these processes. This is consistent with intense morphogenetic changes in the epithelium, but also in mesenchymal cells that are proliferating, migrating and segregating to form the dental papilla and the dental follicle [66, 67]. Moreover, transcriptomes also showed that molar germs in this period are slightly depleted in mitotic cells, and this effect is marked in the upper molar. This depletion may also reflect the intense morphogenetic changes at this stage, as a dichotomy between cell division and cell shape change has been observed (in cancer cells [68], during embryogenesis in various species [69]). Independent of the transcriptomic analysis, our bench data are consistent with differences from this early period, as we see delayed patterning of the second cusp in the upper molar (Fig. 5d). As the two molar germs enter morphogenesis ("cap stage") synchronously, just before ED14.5 (Fig 1), we have to assume that the upper molar takes more time before the second cusp is patterned. Interestingly, this cusp is patterned following the rapid lingual extension of the molar germ, and the persistence of this process in the upper molar enables the formation of a third cusp row [41]. It is possible that morphogenetic processes linked with the lingual development of the molar germ are exaggerated in this ED15.0-16.0 period, and prepare for the later formation of the supplementary cusp row. Such a mechanism for supplementary cusp formation is not mutually exclusive with the difference in activation-inhibition mechanisms proposed above. Both type of evidence concern an early period of crown morphogenesis, while the formation of the third cusp row occurs at the very end of morphogenesis: we thus suggest a scenario where early variation in a core molar program is responsible for ultimate phenotypic differences.

**A general approach of transcriptomes for evolutionary developmental biology and developmental biology.**

Beside our biological question, our results evidence properties of whole organ transcriptomes with strong implications for developmental and evolutionary developmental biology. Transcriptome comparisons are commonly used in both fields, whether it be between mutant and wild type, between species at specific timepoints (including adulthood) or between stages in a specific time course.

First, we show that transcriptomes allow determining the stage of an organ, even at a fine scale. This self-ordering of developmental time series in a multivariate analysis and its convenience for blindly staging samples has been demonstrated in other datasets (e.g. [6, 70]). Our dataset represents a much more restricted portion of the development, because it is sampled at the level of the organ, and not at the level of the whole organism, during a tight window of development. The self-ordering of developmental samples appears because a sufficient number of genes show variations that are subtle, yet consistent between consecutive stages. Second, we show that cells proportions shape clear signatures in the transcriptome of a whole organ (figure 7a,b). This is true comparing different organs (mesenchymal versus epithelial cell proportions) or different developmental stages of the same organ (proportions of mitotic cells, or cusp forming tissue). Importantly, this finding was not limited to molar germ development: by analyzing the most comprehensive dataset of fore and hindlimb development [27], we recovered a strong time signal on PCA1 that correlated well with proportions of chondrocytes estimated by deconvoluting 50 key marker genes (Figure 7c,d and Supplementary text with figure). The signature can be a matter of a proportion of cell types per se (like we see in the case of epithelial versus mesenchyme, cusp tissue or chondrocyte proportion), or a matter of a proportion of cells, engaged in different parts of the cell cycle, regardless of cell type. The effect of cell cycle activity on the heterogeneity of gene expression is reminiscent of the major impact seen in single cell RNAseq data, where it is often treated as a confounding factor [71-73]. In batches of yeast cells, the proportion of cells that stand at different time of the cell cycle is also largely impacting the transcriptomes [74]). To our knowledge, this is the first time such an effect was interpreted in a comparison of complex organs, even if it is easily conceivable that gene expression of a whole organ reflects its cellular composition.

In summary, changes in gene expression taken at the transcriptome level reflect 1) the restriction and expansion of cells populations, each expressing a given repertoire of genes 2) the appearance of new “cell types” expressing

radically different repertoires of genes, and 3) above that, the variations in proportions of cells found in different phases of the cell cycle. This view should be applied to many other types of transcriptome comparisons common in developmental biology and evo-devo. In developmental biology, it is common to compare, at a given timepoint, the same organ in a wild type situation versus a mutant situation. Our findings suggest that the transcriptomic changes provide a monitoring tool of the expansion of some cellular populations to the expense of others (e.g. in mutants with a patterning defect). It is now common to compare the transcriptomes of different species (developing or adult organs) [12, 75-77]. Here, transcriptomic differences may reveal the physiological adaptations of an organ resulting from the relative expansion of a given tissue or cell type within the organ. Combining whole organ transcriptomes with tissue or cell-specific transcriptomes is a powerful approach to get an integrated description of developmental gene expression.

Finally, transcriptomes reveal heterochronies, a key source of differences between species [78]. A few studies have reported transcriptomic heterochronies [79-83]. On the other hand, many "classical" heterochronies have been reported in serial organs, most notably between mammalian limbs, based on morphological criteria or marker genes [25, 26]. To our knowledge, our study is the first report of a transcriptome heterochrony at such a fine scale (a few hours for molars) and between developing serial organs. Importantly, our approach applied to another dataset also recovered the expected heterochronic signal: the forelimb was slightly advanced compared to hindlimb, as expected (Figure 7c, d and Supplementary text). We showed that the proportion of differentiating "cusp-tissue" for molars or differentiating chondrocytes for limbs (figure 7d) increases with time. The relatively smaller proportion of cusp tissue in the upper molar, or chondrocytes in hindlimbs, makes them look younger. In molars, transcriptomic heterochrony associated with cusp epithelium is a matter of cell proportion (figure 7b). In addition, we also see that mesenchyme displays its own pattern of heterochrony: the whole-tooth heterochrony is most likely the result of different patterns. As pointed out by SJ Gould [78] and others [84], heterochronies are most often patterns, and only rarely a direct mechanism for phenotypic difference. Yet identifying heterochronies may help to formulate hypotheses, as seen above. The detection and elucidation of transcriptomic heterochronies will be a very helpful tool to gain insights into the developmental basis of species differences. We anticipate that in any system, where the proportion of a given cell type changes over time, a difference in that proportion between organs will also be perceptible as a transcriptomic heterochrony (independently of technology, as we employed here both RNAseq (molars) and microarrays (limb) technologies).

# CONCLUSION

Our study extends the use of transcriptomes as molecular phenotypes by enhancing our comprehension of the transcriptomic patterns that are reported. We evidenced three transcriptomic signatures, all stemmed from differences in cell proportions between organs and stages. From these, we could formulate specific hypotheses how two serial organs develop in two different morphologies. We predict that transcriptomes taken as "molecular microscopes" will reinvigorate comparative embryology [1]. More generally, we believe transcriptomic signatures in complex organs will help deciphering biological variation in a variety of contexts, from developmental biology to evolutionary biology or medical science (see also, [1, 70]).

# METHODS

## Mice breeding and embryo sampling.

*CD1 (ICR) adult mice were purchased from Charles River (France). Females were mated overnight (night at 20:00 pm) and the noon after morning detection of a vaginal plug was indicated as embryonic day (ED) 0.5. Other breeding pairs were kept in a light-dark reversed cycle (night at 12:00 pm), so that the next day at 16:00 pm was considered as ED1.0. Pregnant females were killed by cervical dislocation and embryos were harvested on cooled Hank's or DMEM advanced medium, weighted as described in [85] and immediately decapitated. This study was performed in strict accordance with the European guidelines 2010/63/UE and was approved by the Animal Experimentation Ethics Committee CECCAPP (Lyon, France; reference ENS_2012_045).*

## Epithelium dissociations and in situ hybridizations.

*Complete or hemi mandibles and maxillae were dissected in Hank's medium and treated with Dispase (Roche) 100% at 37°C for 1h30 to 2h20 depending on embryonic stage. Epithelium was carefully removed and fixed overnight in PFA 4% at 4°C. DIG RNA antisense Fgf4 probe were prepared from plasmids described elsewhere [86]. In situ hybridization was done according to a standard protocol. Photographs were taken on a Leica M205FA stereomicroscope with a Leica DFC450 digital camera (Wetzlar, Germany) or on a Zeiss LUMAR stereomicroscope with a CCD CoolSNAP camera (PLATIM, Lyon).*

## RNA-seq sample preparation

*A total of 16 samples were prepared for the RNA-seq analysis representing 8 stages (ED14.5, 15.0, 15.5, 16.0, 16.5, 17.0, 17.5, 18.0), coming from 8 individuals. Each sample contained 2 tooth germs, the left and right first molars (M1) of the same male individual, and for a given stage, the upper and lower samples came from the same individual. The heads of harvested embryo were kept for a minimal amount of time in cooled PBS (small scale) or advanced DMEM medium (large scale). The M1 lower and upper germs were dissected under a stereomicroscope and stored in 200µl of RNA later (SIGMA). Similarly dissected tooth germs from the same litter and same weight were fixed overnight in PFA 4% for immunolocalization and 3D reconstruction (see later). Another embryo of the same litter and same weight was processed as indicated above for Fgf4 in situ hybridization. Total RNA was prepared using the RNeasy micro kit from QIA-*

221

*GEN following lysis with a Precellys homogenizer. RNA integrity was controlled on a Bioanalyzer (Agilent Technologies, a RIN of 10 was reached for all samples used in this study). PolyA+ libraries of the large-scale dataset were prepared with the Truseq V2 kit (Illumina), starting with 150ng total mRNA and reducing the amplification step to only 12 cycles and sequenced on an Illumina Hi-seq2000sequencer at the GENOSCOPE (Evry, France).*

## Data analysis

*R scripts corresponding to the main methods and processed data are available on GitHub (`https://github.com/msemon/ToothTranscriptomeAnalyses`) [87] and in zenodo (doi:10.5281/zenodo.197077). Source code is free software governed byCeCILL License.*

## Expression levels estimation using RNA-seq and differential expression analysis

*Sequences were obtained for each samples (Illumina sequencing, 100 bp paired-end sequences). These reads were mapped to the mouse genome (mm10, october 2012), using Bowtie2 (version 2.0.2; [88]) and Tophat2 (version 2.0.6; [89]) with standard settings., we obtained a median of 37.8 million reads that mapped uniquely to the genome (minimum = 32.9 to maximum = 44.3 M reads). The number of reads per gene was then counted using HTSeq (version 0.5.3p9; option : "intersection-nonempty"; 21938 genes with an UCSC annotation; [90]). Raw data is available in the gene expression omnibus repository. (GSE76316). Published RNAseq were uploaded (accession number GSE39918 and GSE36863) and treated as our own data for read mapping, counting and normalization. Statistical analysis was performed using R (version 3.1.2, 2014-10-31; [91]), with ggplot2 for graphics (ggplot2_1.0.0; [92]). Two main analysis used in the text include differential gene expression analysis (DE) and multivariate analysis. Differential expression analysis was performed using DEseq package (DESeq2, version 1.6.3, [93]), with settings indicated in the main text. Multivariate analyses (mainly principal component analyses) were performed using ade4 package (ade4_1.6-2; [94]).*

### Gene Ontology analysis

*Gene ontology (GO) analysis were performed and visualized with GORILLA [95] and GOStats (version 2.34.0, with Mmusculus.UCSC.mm10.ensGene_3.1.2 and GO.db_3.1.2;[96]), using the full list of genes expressed in the corresponding dataset as a background. R custom scripts were used for graphical representation of the GO-terms enrichments.*

### Clustering of upper and lower time-series

*For each gene and at each stage of development, we took the level of expression divided by the initial level of expression (i.e. level at stage ED 14.5). Following [97], we then ranked the genes according to the maximum obtained value and removed the first quartile (relatively flat genes), obtaining a set of 17,506 remaining genes. To classify the genes we first generated theoretical profiles representing all possible combinations of transitions between stages (expression level being either increasing, decreasing or flat between the 8 consecutive stages), making up a total of 2,187 profiles. We then correlated each real expression profile to each theoretical profile using Spearman correlations. We clustered the obtained correlation matrix using K-means (choosing 10 clusters, and checking that the obtained profiles are repeatable over several iterations of the method). The median of the expression profiles of these 10 clusters are drawn on Fig 2b.*

### Simulations of realistic tooth transcriptomes with different tissue proportions

*To simulate realistic tooth transcriptomes while controlling the relative proportions of mesenchyme (M), enamel knot (K) and non-enamel knot (E) epithelium, we used expression levels in isolated tissue obtained from microarray data, taking the median of ED12.5, ED13.0, ED13.5 ([39], `http://compbio.med.harvard.edu/ToothCODE/`). We consider that for each gene G, expression level is proportional to a mixture of its microarray levels (denoted respectively mMG, mKG and mEG) in the three tissues. If we set the hypothesis that lower tooth is a mixture of pE epithelium tissue and (1-pE) mesenchyme, and that epithelium tissue is a mixture of pP non-enamel knot epithelium and (1-pP) enamel knot, then gene G specific expression proportion factor equals:*

$$kG = LG/((1-pE)*mMG+pE*(pP*mEG+(1-nP)*mKG)) \quad (1)$$

*where LG is this gene expression level is the RNAseq library of the lower tooth.*
*From this, we can simulate artificial RNA libraries with several tissue proportions, such as pE' epithelium and pP' non-enamel knot in epithelium. For each gene G, artificial expression level is:*

$$AG = kG*((1_pE')*mMG+pE'*(pP'*mEG+(1-pP')*mKG)) \quad (2)$$

*First, to measure the impact of epithelium proportion on the position of the sample on the first axis of the PCA, we sampled artificial teeth corresponding to the expression levels ED14.5, with a range of plausible initial conditions (pE). We made several simulations, with plausible initial K compositions (pP varying from 5% to 20% of the total epithelium) and with a range of coherent epithelium proportions (pE' representing 45 to 60% of total tooth). From these simulations, we then built artificial samples in which the proportion of epithelium is increased, and projected the obtained artificial samples on the PCA. On average, 15.7% of the PCA1 distance between ED 14.5 and ED 15.0 is covered when pE' increases by 30% (maximum = 28%, obtained when enamel knot represents 20*
*Second, to measure the impact of non-enamel knot proportion (pP) on the position of the sample on the second axis of the PCA, we sampled an artificial teeth corresponding to a ED14.5 lower molar transcriptome and, from this, made artificial samples in which the proportion pP' of epithelium decreases. Starting from 55% non-enamel knot epithelium, and decreasing to 40 or 50% covers 23 to 35% of the real distance between upper and lower tooth on PCA2 at ED14.5.*

### Estimating tissue proportions from RNAseq data : deconvolutions

*Performing gene expression deconvolution, is estimating cell type proportions (and/or cell- specific gene expression signatures) from global expression data in heterogeneous samples. Here, we used the R package CellMix (CellMix_1.6.2, [98]), to estimate the proportions of the three tissue compartments in our tooth germ transcriptomes. CellMix implements in particular the method Digital Sorting Algorithm ( DSA) proposed by Zhong et al. [99] which performs complete gene expression deconvolution using a set of marker genes only. We needed therefore to define 3 sets of markers, as specific as possible for each tissue compartment. We first used markers obtained from bite-it database (used in Fig 3b): 11 for EK, 12 for epithelium and 14 for mesenchyme. In the case of the study of maturation, we further restricted EK markers to 6 genes with a large pattern of expression (used in Fig 6g). The*

*results were confirmed with markers obtained from microarray data ([39]; used in Sup Fig 3): 18 EK (defined such as expression level in ek > 1.2x expression level in mesenchyme and level in ek > 1.2x level in epithelium), 9 for epithelium (such as epithelium > 1.5x mesenchyme and epithelium > 1.2x ek) and 27 mesenchyme (such as mesenchyme > 1.5x epithelium and mesenchyme > 1.2x ek). The thresholds used (1.5x and 1.2x) were set up to guarantee enough markers relatively equilibrated among the 3 tissues, and yet that are specific enough for the tissue under study. Of note, results were qualitatively unchanged when the thresholds are modified (within a range 1.2-1.5 for both). To get an idea of the sensitivity of the results to the choice of the set of markers, for each analysis, we computed 500 bootstraped values by randomly resampling half of the total number of markers (keeping only samples such as there are more than 5 markers per compartment) and recomputing deconvolutions on these 500 random subsets. The confidence intervals of these resamplings are indicated in the barplots (Fig 3b, Fig 6g and Sup Fig 3).*

### Morphological determination of the volume of tissue components

*The mouse embryos and fetuses were harvested during ED14.5–17.5, weighted and fixed in Bouin-Hollande fluid. At each stage, 1-2 specimens representative specimens exhibiting median body weight were selected. Their heads were routinely embedded with paraffin and series of 10µm thick frontal histological sections were prepared and stained by hematoxylin- eosin. On each section, the surfaces of dental epithelium and dental mesenchyme were measured for the fours molars. From these data, total volumes of the two compartments were calculated.*

## Measuring heterochrony on genes with a bell-shaped profile.

*To estimate the gene specific heterochrony offset, we computed the cubic splines of both expression curves (upper and lower) : up(t) and lo(t). We model the offset hbetween up and lo as, for all times t:*

$$up(t) = lo(t+h) + \epsilon \; where \epsilon \; \sim \; N(0,1) \tag{3}$$

*To compute the maximum likelihood estimation of h, we considered 100 equidistant time points in the whole interval, $t_1,\ldots,t_{100}$; for any tested value of h, we kept all time points $t'_i$ such that $t'_{i+h}$ is also in the interval, and computed the average squared difference:*

$$\frac{1}{number_of_t'_i} \sum (up(t'_i) - lo(t'_{i+h}))^2 \tag{4}$$

*We define the heterochrony for a gene as the value of h which minimizes this measure.*

### Method based on hidden Markov Models to model upper/lower differences

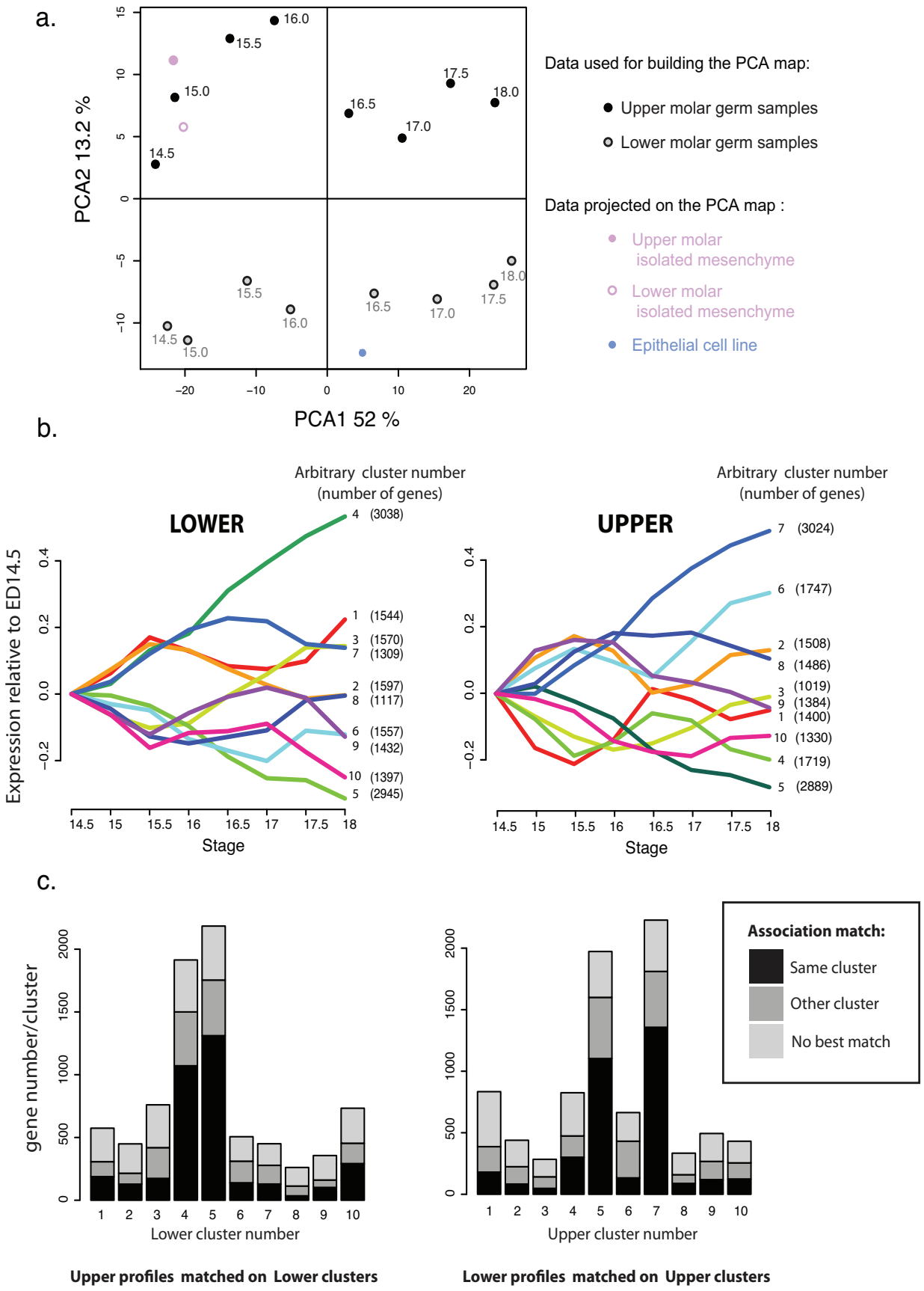
*We designed a method, based on hidden Markov Models, to track differences in time profile, in a way which explicitly models the temporal series. First, we looked for a description in 2 classes of the overall distribution of the absolute value of upper/lower ratios, taking into account the temporal constraint To do this, we modeled this distribution as a mixture of 2 normal distributions, organized in an hidden Markov model. This makes 7 free parameters to estimate, using 8 time points from 2477 genes. These 2 distributions (hidden states) correspond to medium (ME) and high (HI) upper/lower difference in expression levels. Considering all time series independently with the classical forward algorithm, it is possible to compute the likelihood of the whole data, and we optimized this likelihood to fit at best the transition probabilities between the states of the HMM the whole data (2477 genes with a differential expression between upper and lower), using R library RHmm version 2.0.3 [87]. Finally, from this optimized HMM, we computed the a posteriori probabilities of transition between states for each gene, in each interval of time points in each time series. In an attempt to better fit the distribution of the ratios, we performed similar analyses, using an HMM with two hidden states, where each hidden state is now modeled by a mixture of gaussian distributions. In Fig 4b, for each type of transition and each interval of time points, we counted how many genes followed this transition with a probability above 0.8. We also obtained similar results using an HMM with three hidden states.*

### Models of cusp maturation

*Within a mouse litter, embryos of the same weight are rigorously at the same developmental stage. Therefore to estimate the number of cusp patterned for a given RNA-seq sample, we used an embryo harvested from the same litter and having the same weight. Following epithelium dissociation, we performed Fgf4 in situ hybridization to reveal the number and position of Fgf4 expressing enamel knots. We built a simple model for quantifying cusp patterning and maturation in time, fueled with the data on cusp number and position mentioned above. For simplicity, we chose to model cusp proportions with a simple grid model of 6 (lower molar) or 8 (upper molar) equally contributing units, corresponding to the cusps. Once patterned, the territory of each cusp expands at the same speed within a unit. (Fig 6c). The expansion of the territory is simply coded by increments of one unit (from 0: no cusp patterned to 8: fully expanded cusp) per half day. The proportion of cusp tissue per tooth germ is then computed as the sum of the values assigned to each cusp, divided by the total volume for a fully patterned tooth (total number of cusps x maximal value = 8). In this model, the cusps are still being incremented at the last time point. Alternative models were build by incrementing the cusp proportions over y steps only after the cusp patterning (the model with 5 steps, corresponding to 2.5 days, has the best fit). The fit of these models were computed using generalized linear models (glm in R), which is modeling the proportion of cusp tissue over the total molar tooth germ.*

## Estimating heterochrony in the different tissue compartments

*Using microarray data [39], we selected 3 sets of markers corresponding to each tissue compartment, that contain a similar number of genes. We obtained 336 markers relatively specific to epithelial cells, such as the ratio Epithelium/Mesenchyme is among the top 5% and the ratio Epithelium/EK is in the top 20%. Similarly, we selected 421 EK markers (EK/Epithelium in the top 5% and EK/Mesenchyme in the top 5%) and 566 mesenchyme markers (Mesenchyme/Epithelium in the top 5% and Mesenchyme/EK in top 5%). By drawing PCAs based on these subsets of genes that are relatively specific from each tissue compartment, we obtained a map of the samples seen by this tissue. This permits to evaluate whether each tissue is carrying the time component and the heterochrony seen in the whole dataset. Furthermore, as explained briefly in Sup Fig 6a, the normalization of expression data only on these genes removes part of the influence of tissue composition.*




224 Figure2 - **Lower and upper first molar germs share temporal dynamics of gene.**

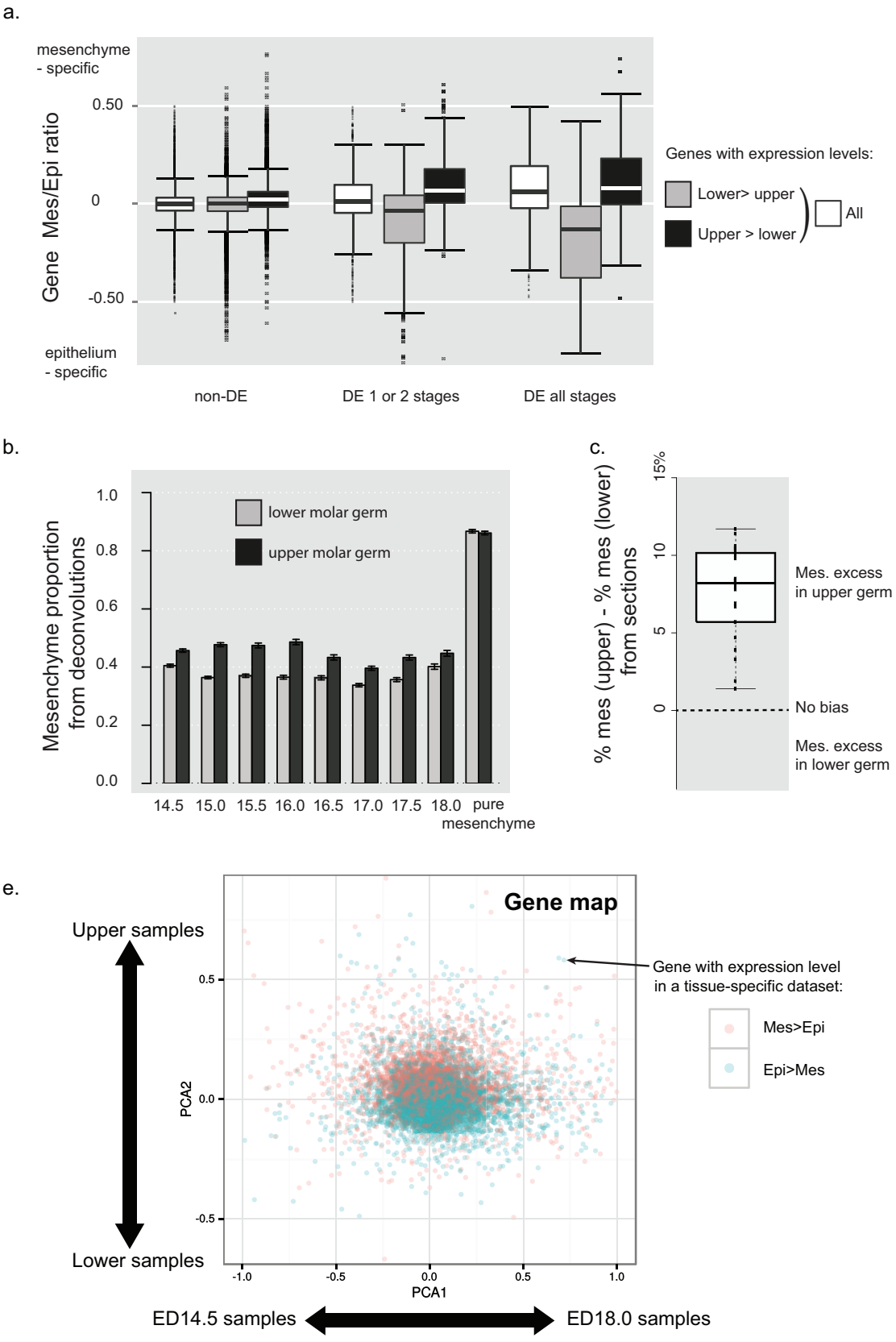



Figure3 - **The upper molar germ is enriched in mesenchyme.**

225

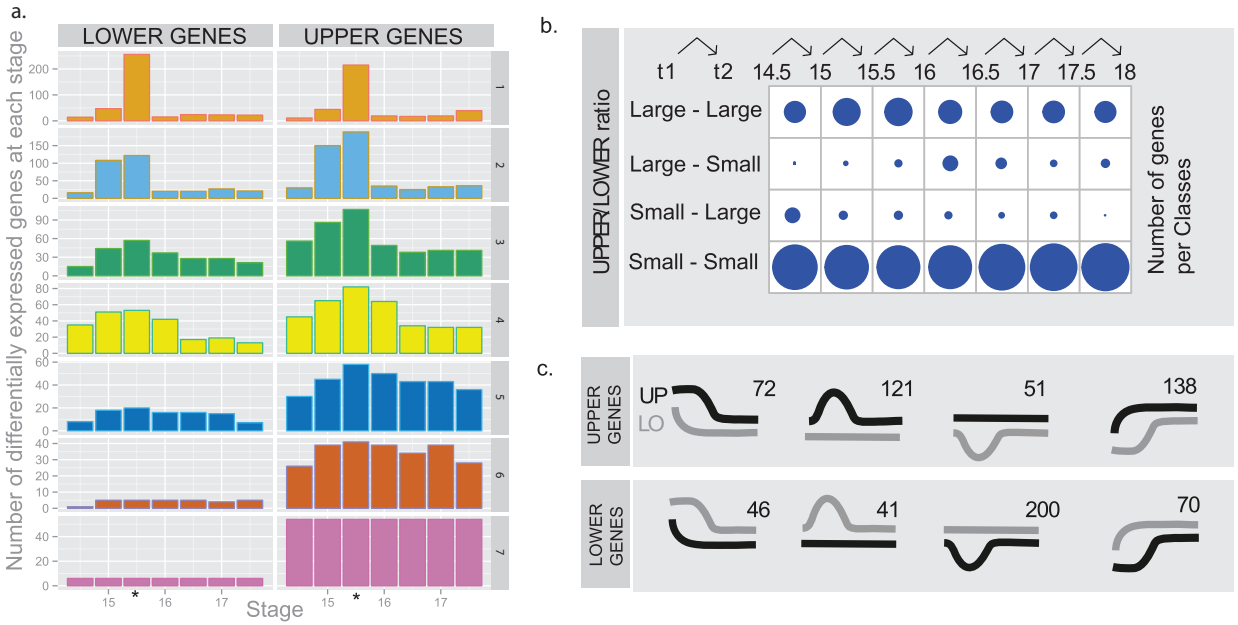



Figure 4 – **Time-dependent differences in upper/lower gene expression.**
**(a)** Number of differentially expressed genes for 7 consecutive periods. We defined these 7 periods, by considering pairs of consecutive developmental stages as replicates (i.e. ED14.5-15.0, ED15.0-15.5, *etc*). We extracted differentially expressed genes at each of these 7 periods, and split genes in two categories: those with a higher expression level in lower molar and those with a higher expression level in upper molar. We also split genes that were found to be differentially expressed at 1, 2, ..., up to all (7) periods. **(b)** Distribution of temporal shifts of the upper/lower expression ratios, extracted by a method based on hidden Markov model. The size of the spots are proportional to the number of genes for which upper/lower expression ratio follows a particular kind of profile. For instance, small-large at 14.5-15 indicates that the upper/lower expression ratio was small at ED 14.5, and larger at stage ED 15. **(c)** Classes and number of genes assigned to temporal profiles of the genes that are differentially expressed at ED 15.5-16.0 but not earlier or later during the developmental series.

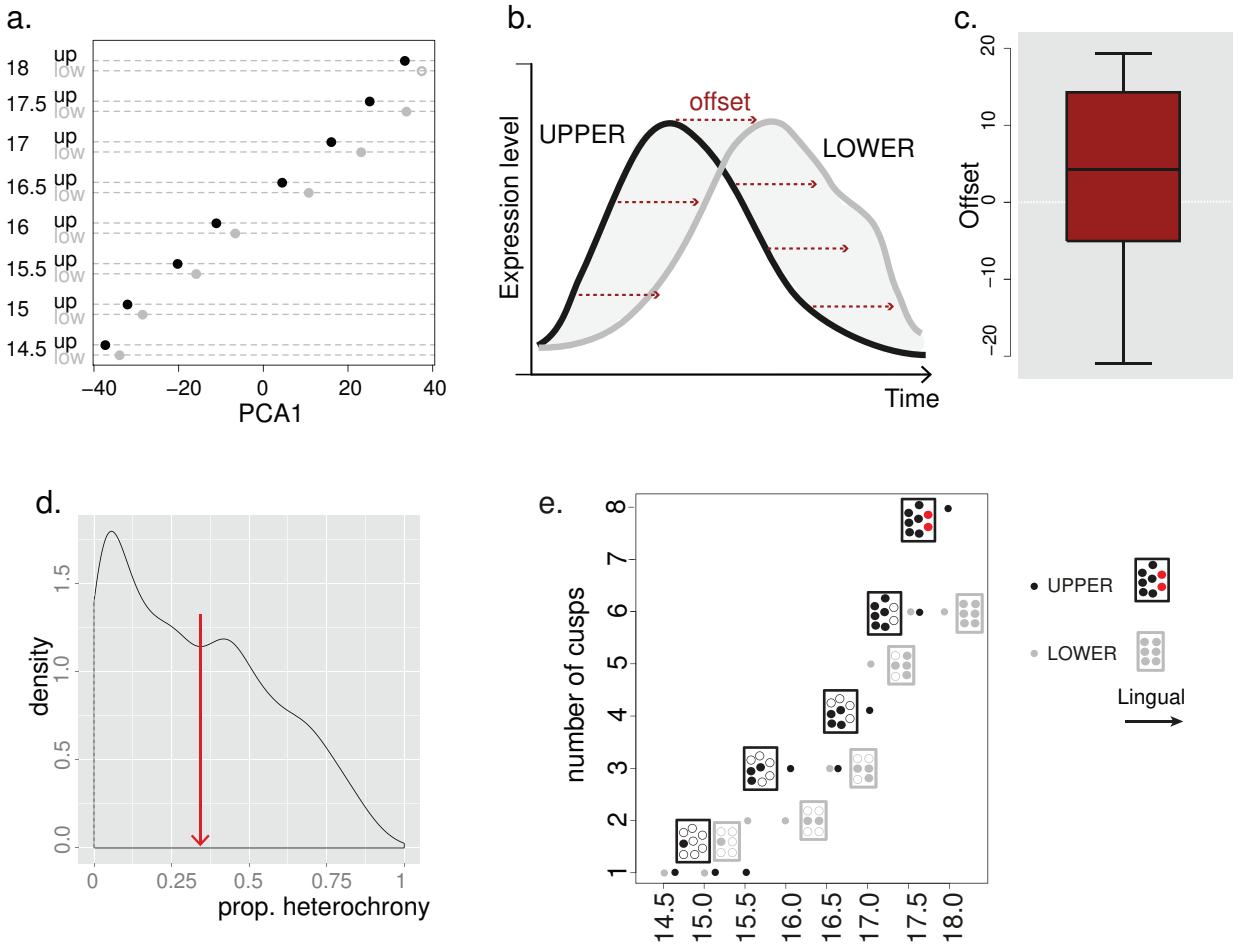



Figure 5 – **The heterochrony seen at the transcriptome level is not directly consistent with the number of patterned cusps .**
**(a)** Coordinates of the samples on PCA1 showing a systematic offset between upper (black) and lower (grey) samples, the latter being more "mature" than the former. **(b)** Heterochrony measured on bell-shaped genes: Toy example showing how, as a proxy for heterochrony, we are using the simple time shift from upper to lower. **(c)** Distribution of this time shift, which is significantly positive, indicating that the upper molar germ is most often delayed in comparison with the lower germ. **(d)** Proportion of the upper/lower distance, which can be explained by a simple time-translation. The average is indicated by a red arrow. **(e)** Total number of patterned cusps at each developmental stage for upper (black) and lower (grey) molar of the RNAseq timeseries. At each stage, embryos of the same litter and same weight as those chosen for the RNAseq timeseries in panel (a) were used for in situ hybridizations with a *Fgf4* probe marking the secondary enamel knots.

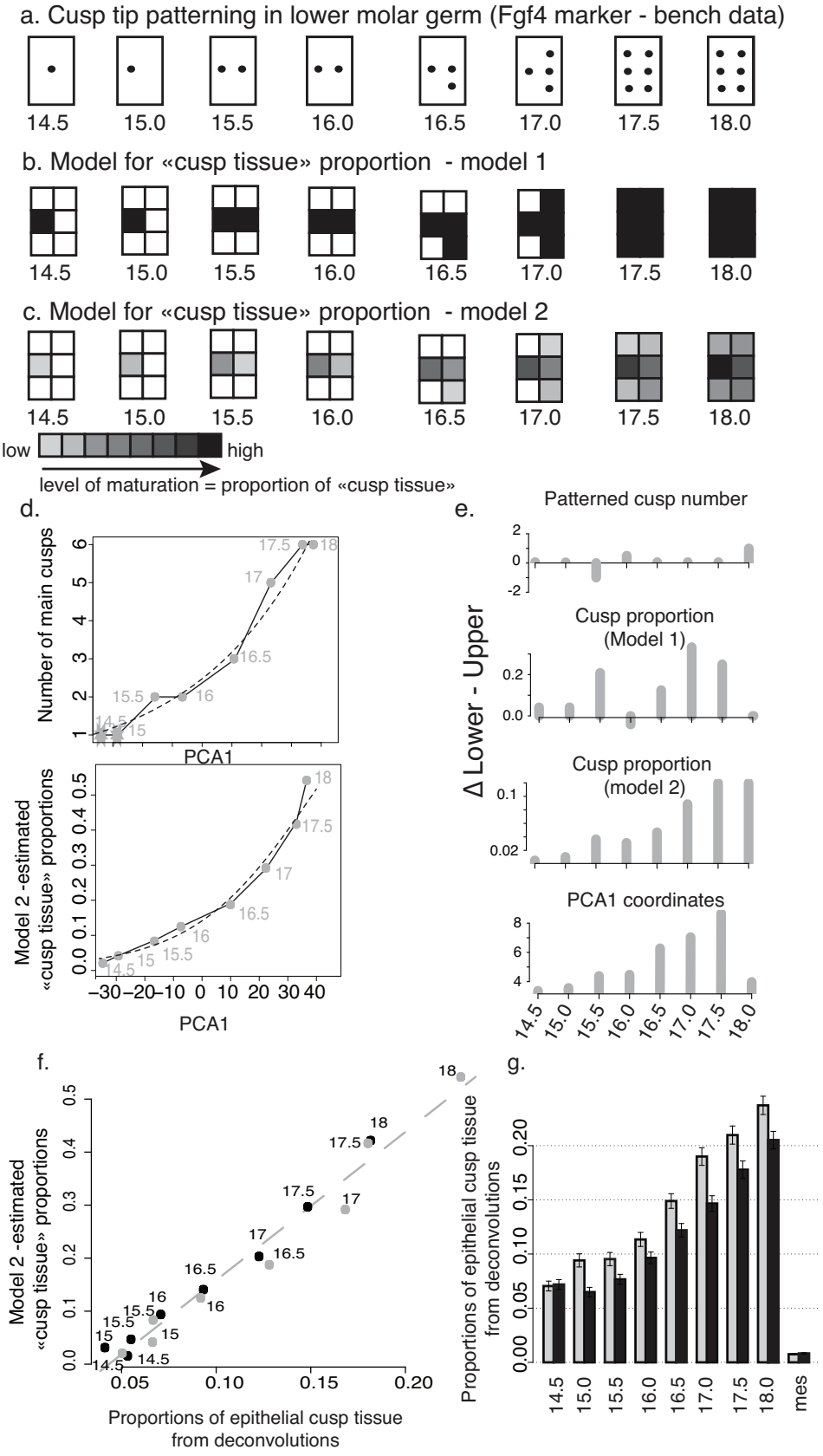



Figure6 - **Cusp proportions estimated with a simple model recapitulate time ordering on first PCA axis and upper-lower heterochrony.**

228

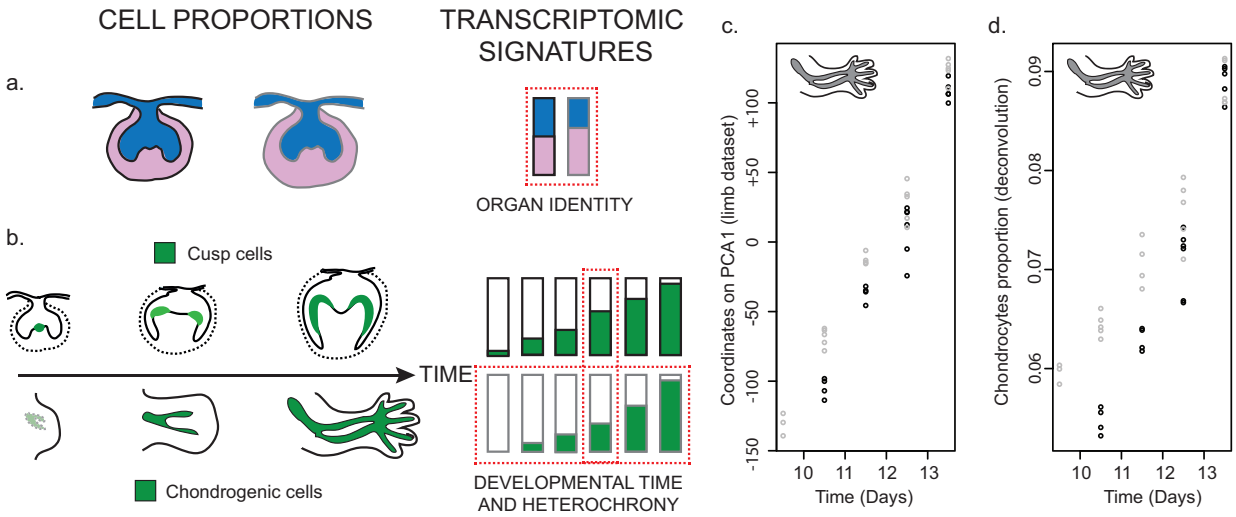



Figure 7 – **Transcriptomic signatures shaped by cell proportion reveal differences in organ development.** **(a)** A difference in mesenchyme (pink) versus epithelium (blue) proportion in the tooth germ takes part in a transcriptomic signature of organ identity. **(b)** Increasing proportion of differentiating cells over time takes part in a transcriptomic signature of developmental time and heterochrony. We evidenced this both in molars (cusp cell differentiation) and limb (chondrogenic cells, see c and d). **(c and d)** We reanalyzed a published transcriptome dataset on fore/hind limb development obtained with microarrays by [27]. On the first axis of PCA, we recovered a strong time-related signal (7c) and the expected heterochrony between fore and hindlimb (forelimb samples are systematically ahead on PCA1, 7c). Both the time signal and the heterochrony correlated well with the proportion of chondrogenic cell types estimated by deconvolution with a set of marker genes extracted from [104] (7d, compare 7c and 7d).

# DECLARATION

### Ethic approval

Mouse experiments for this study were performed in strict accordance with the European guidelines 2010/63/UE and were approved by the Animal Experimentation Ethics Committee CECCAPP (Lyon, France; reference ENS_2012_045).

### Consent for publication

Not applicable

### Competing interests

The authors declare no competing interests.

### Availability of data and materials

The dataset supporting the conclusions of this article is available in the gene expression omnibus (GEO) repository, GSE76316 [100]; R scripts corresponding to the main methods and processed data are available on GitHub [87] and in zenodo (doi:10.5281/zenodo.197077). Source code is free software governed byCeCILL License.

### Funding

SP and MS benefited from a grant from the Agence Nationale pour la Recherche (`http://www.agence-nationale-recherche.fr/`): ANR 2011 JSV6 00501 "CONVERGDENT" and a grant from the GENOSCOPE - Centre National de Séquençage. RP benefited from a grant from the Grant Agency of the Czech Republic, No. 14-37386G.
Salaries were funded by the Centre National pour La Recherche Scientifique, the Ecole Normale Supérieure de Lyon and the Université de Lyon, Université Lyon I.

### Author's contribution

MS and SP conceived, analyzed and interpreted the experiments. MS, SP, LG, CP, AL, RP made the experiments. MS and SP wrote the manuscript. MS and SP contributed equally to this study.

### Acknowledgements

We acknowledge the contribution of Alain Rubod for 3D reconstructions, Alice Joffard for heterochronies, the Plateau de Biologie Expérimentale de la Souris (PBES) and the Plateau Technique Imagerie/Microscopie (PLATIM) of SFR Biosciences Gerland-Lyon Sud (UMS344/US8). We are grateful to Irma Thesleff for the *fgf4* probe plasmid. We are much grateful to Pauline Sémon for her old style drawings of molars, to Vincent Laudet for support, to Vanessa Ribes for helpful reading of the manuscript, to Joanne Burden for english proofreading and more, and to Ulysse for its coming back.

230

# REFERENCES

1. Pantalacci S, Semon M: **Transcriptomics of developing embryos and organs: A raising tool for evo--devo.** J Exp Zool B Mol Dev Evol 2015, 324:363--371
2. Roux J, Rosikiewicz M, Robinson--Rechavi M: **What to compare and how: Comparative transcriptomics for Evo--Devo.** J Exp Zool B Mol Dev Evol 2015, 324:372--382.
3. Papatsenko D, Levine M, Goltsev Y: **Clusters of temporal discordances reveal distinct embryonic patterning mechanisms in Drosophila and anopheles.** PLoS Biol 2011, 9:e1000584.
4. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR: **Reconstructing lineage hierarchies of the distal lung epithelium using single--cell RNA--seq.** Nature 2014, 509:371--375
5. Scialdone A, Tanaka Y, Jawaid W, Moignard V, Wilson NK, Macaulay IC, Marioni JC, Gottgens B: **Resolving early mesoderm diversification through single--cell expression profiling.** Nature 2016, 535:289--293.
6. Anavy L, Levin M, Khair S, Nakanishi N, Fernandez--Valverde SL, Degnan BM, Yanai I: **BLIND ordering of large--scale transcriptomic developmental timecourses.** Development 2014, 141:1161--1166.
7. Gerrard DT, Berry AA, Jennings RE, Piper Hanley K, Bobola N, Hanley NA: **An integrative transcriptomic atlas of organogenesis in human embryos.** Elife 2016, 5.
8. Francesconi M, Lehner B: **Reconstructing and analysing cellular states, space and time from gene expression profiles of many cells and single cells.** Mol Biosyst 2015, 11:2690--2698.
9. Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH: **Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid.** Proc Natl Acad Sci U S A 2014, 111:E4736--4742.
10. Tian X, Strassmann JE, Queller DC: **Dictyostelium development shows a novel pattern of evolutionary conservation.** Mol Biol Evol 2013, 30:977--984.
11. Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, et al: **The mid--developmental transition and the evolution of animal body plans.** Nature 2016.
12. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P: **Gene expression divergence recapitulates the developmental hourglass model.** Nature 2010, 468:811--814.
13. Levin M, Hashimshony T, Wagner F, Yanai I: **Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo.** Dev Cell 2012, 22:1101--1108.
14. Wang Z, Young RL, Xue H, Wagner GP: **Transcriptomic analysis of avian digits reveals conserved and derived digit identities in birds.** Nature 2011, 477:583--586.

15. Tschopp P, Sherratt E, Sanger TJ, Groner AC, Aspiras AC, Hu JK, Pourquie O, Gros J, Tabin CJ: **A relative shift in cloacal location repositions external genitalia in amniote evolution.** Nature 2014, 516:391--394.

16. Kin K, Nnamani MC, Lynch VJ, Michaelides E, Wagner GP: **Cell-type phylogenetics and the origin of endometrial stromal cells.** Cell Rep 2015, 10:1398--1409.

17. Angelini DR, Kaufman TC: **Insect appendages and comparative ontogenetics.** Dev Biol 2005, 286:57--77.

18. Ruvinsky I, Gibson--Brown JJ**: Genetic and developmental bases of serial homology in vertebrate limb evolution.** Development 2000, 127:5233--5244.

19. Duboc V, Logan MP: **Regulation of limb bud initiation and limb-type morphology.** Dev Dyn 2011, 240:1017--1027.

20. Weatherbee SD, Carroll SB: **Selector genes and limb identity in arthropods and vertebrates.** Cell 1999, 97:283--286.

21. Van Valen L: **Serial homology: the crests and cusps of mammalian teeth.** Acta Palaeontologica Polonica 1994, 38:145--158

22. Weiss K, Stock D, Zhao Z, Buchanan A, Ruddle F, Shashikant C: **Perspectives on genetic aspects of dental patterning.** Eur J Oral Sci 1998, 106 Suppl 1:55--63.

23. Mann RS, Carroll SB: **Molecular mechanisms of selector gene function and evolution.** Current Opinion in Genetics & Development 2002, 12:592--600.

24. Cobourne MT, Sharpe PT: **Tooth and jaw: molecular mechanisms of patterning in the first branchial arch.** Arch Oral Biol 2003, 48:1--14.

25. Bininda--Emonds OR, Jeffery JE, Sanchez--Villagra MR, Hanken J, Colbert M, Pieau C, Selwood L, Ten Cate C, Raynaud A, Osabutey CK, Richardson MK: **Forelimb--hindlimb developmental timing changes across tetrapod phylogeny.** BMC Evol Biol 2007, 7:182.

26. Richardson MK, Gobes SM, van Leeuwen AC, Polman JA, Pieau C, Sanchez--Villagra MR: **Heterochrony in limb evolution: developmental mechanisms and natural selection.** J Exp Zool B Mol Dev Evol 2009, 312:639--664.

27. Taher L, Collette NM, Murugesh D, Maxwell E, Ovcharenko I, Loots GG: **Global gene expression analysis of murine limb development.** PLoS One 2011, 6:e28358.

28. Laugel--Haushalter V, Paschaki M, Thibault--Carpentier C, Dembele D, Dolle P, Bloch-- Zupan A: **Molars and incisors: show your microarray IDs.** BMC Res Notes 2013, 6:113.

29. Jernvall J, Thesleff I: **Tooth shape formation and tooth renewal: evolving with the same signals.** Development 2012, 139:3487--3497.

30. Tucker A, Sharpe P: **The cutting--edge of mammalian development; how the embryo makes teeth.** Nat Rev Genet 2004, 5:499--508.

31. Weiss KM, Stock DW, Zhao Z: **Dynamic interactions and the evolutionary genetics of dental patterning.** Crit Rev Oral Biol Med 1998, 9:369--398.

32. Ferguson CA, Tucker AS, Sharpe PT: **Temporospatial cell interactions regulating mandibular and maxillary arch patterning.** Development 2000, 127:403--412.

33. Zhao Z, Stock D, Buchanan A, Weiss K: **Expression of Dlx genes during the development of the murine dentition.** Dev Genes Evol 2000, 210:270--275.

34. Salazar--Ciudad I: **Tooth morphogenesis in vivo, in vitro, and in silico.** Curr Top Dev Biol 2008, 81:341--371.

35. Salazar--Ciudad I: **Tooth patterning and evolution.** Current Opinion in Genetics & Development 2012, 22:585--592.

36. Jernvall J, Thesleff I: **Reiterative signaling and patterning during mammalian tooth morphogenesis.** Mech Dev 2000, 92:19--29.

37. Thesleff I, Keranen S, Jernvall J: **Enamel knots as signaling centers linking tooth morphogenesis and odontoblast differentiation.** Adv Dent Res 2001, 15:14--18.

38. Landin MA, Nygard S, Shabestari MG, Babaie E, Reseland JE, Osmundsen H: **Mapping the global mRNA transcriptome during development of the murine first molar.** Front Genet 2015, 6:47.

39. O'Connell DJ, Ho JW, Mammoto T, Turbe--Doan A, O'Connell JT, Haseley PS, Koo S, Kamiya N, Ingber DE, Park PJ, Maas RL: **A Wnt--bmp feedback circuit controls intertissue signaling dynamics in tooth organogenesis.** Sci Signal 2012, 5:ra4.

40. Peterkova R, Hovorakova M, Peterka M, Lesot H: **Three--dimensional analysis of the early development of the dentition.** Aust Dent J 2014, 59 Suppl 1:55--80.

41. Cam Y, Fausser JL, Vonesch JL, Peterkova R, Peterka M, Halaskova M, Lesot H: **Asymmetrical morphogenesis and medio--lateral positioning of molars during mouse development.** Eur J Oral Sci 2002, 110:35--43.

42. Biben C, Wang CC, Harvey RP: **NK--2 class homeobox genes and pharyngeal/oral patterning: *Nkx2--3* is required for salivary gland and tooth morphogenesis.** Int J Dev Biol 2002, 46:415--422.

43. Mitsiadis TA, Drouin J: **Deletion of the Pitx1 genomic locus affects mandibular tooth morphogenesis and expression of the *Barx1* and *Tbx1* genes.** Dev Biol 2008, 313:887--896.

44. Ferguson CA, Tucker AS, Christensen L, Lau AL, Matzuk MM, Sharpe PT: **Activin is an essential early mesenchymal signal in tooth development that is required for patterning of the murine dentition.** Genes Dev 1998, 12:2636--2649.

45. Jia S, Zhou J, Gao Y, Baek JA, Martin JF, Lan Y, Jiang R: **Roles of *Bmp4* during tooth morphogenesis and sequential tooth formation.** Development 2013, 140:423--432.

231

46. Aberg T, Wang XP, Kim JH, Yamashiro T, Bei M, Rice R, Ryoo HM, Thesleff I: ***Runx2*** **mediates FGF signaling from epithelium to mesenchyme during tooth morphogenesis.** Dev Biol 2004, 270:76--93.
47. Wang XP, Suomalainen M, Jorgez CJ, Matzuk MM, Wankell M, Werner S, Thesleff I: **Modulation of activin/bone morphogenetic protein signaling by follistatin is required for the morphogenesis of mouse molar teeth.** Dev Dyn 2004, 231:98--108.
48. Qiu M, Bulfone A, Ghattas I, Meneses JJ, Christensen L, Sharpe PT, Presley R, Pedersen RA, Rubenstein JL: **Role of the Dlx homeobox genes in proximodistal patterning of the branchial arches: mutations of Dlx--1, Dlx--2, and Dlx--1 and --2 alter morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches.** Dev Biol 1997, 185:165--184.
49. Thomas BL, Tucker AS, Qui M, Ferguson CA, Hardcastle Z, Rubenstein JL, Sharpe PT: **Role of Dlx--1 and Dlx--2 genes in patterning of the murine dentition.** Development 1997, 124:4811--4818.
50. Jeong J, Li X, McEvilly RJ, Rosenfeld MG, Lufkin T, Rubenstein JL: **Dlx genes pattern mammalian jaw primordium by regulating both lower jaw--specific and upper jaw-- specific genetic programs.** Development 2008, 135:2905--2916.
51. Schmitt R, Lesot H, Vonesch JL, Ruch JV: **Mouse odontogenesis in vitro: the cap--stage mesenchyme controls individual molar crown morphogenesis.** Int J Dev Biol 1999, 43:255--260.
52. Tooth and Craniofacial Development Group of the Developmental Biology Programme IoB, University of Helsinki: Bite--it database. Tooth and Craniofacial Development Group of the Developmental Biology Programme, Institute of Biotechnology, University of Helsinki.
53. Kaur H, Mao S, Li Q, Sameni M, Krawetz SA, Sloane BF, Mattingly RR: **RNA--Seq of human breast ductal carcinoma in situ models reveals aldehyde dehydrogenase isoform 5A1 as a novel potential target.** PLoS One 2012, 7:e50249.
54. Lisi S, Peterkova R, Peterka M, Vonesch JL, Ruch JV, Lesot H: **Tooth morphogenesis and pattern of odontoblast differentiation.** Connect Tissue Res 2003, 44 Suppl 1:167--170.
55. Harjunmaa E, Kallonen A, Voutilainen M, Hamalainen K, Mikkola ML, Jernvall J: **On the difficulty of increasing dental complexity.** Nature 2012, 483:324--327.
56. Rutishauser R, Moline P: **Evo--devo and the search for homology ("sameness") in biological systems.** Theory Biosci 2005, 124:213--241.
57. Minelli A: **The origin and evolution of appendages.** Int J Dev Biol 2003, 47:573--581.
58. Hu B, Nadiri A, Kuchler--Bopp S, Perrin--Schmitt F, Peters H, Lesot H: Tissue engineering of tooth crown, root, and periodontium. Tissue Eng 2006, 12:2069--2075.
59. Ouimette JF, Jolin ML, L'Honore A, Gifuni A, Drouin J: Divergent transcriptional activities determine limb identity. Nat Commun 2010, 1:35.
60. Moullec N: [Contribution to the morphological study of mouse odontogenesis. Causality of tooth differentiation (author's transl)]. Arch Anat Histol Embryol 1978, 61:151--175.
61. Salazar--Ciudad I, Jernvall J: **A computational model of teeth and the developmental origins of morphological variation.** Nature 2010, 464:583--586.
62. Irie N, Kuratani S: Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat Commun 2011, 2:248.
63. Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson--Rechavi M: **The hourglass and the early conservation models-–co--existing patterns of developmental constraints in vertebrates.** PLoS Genet 2013, 9:e1003476.
64. Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I: **A transcriptomic hourglass in plant embryogenesis.** Nature 2012, 490:98--101.
65. Wang Z, Pascual--Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, Li C, White S, Xiong Z, Fang D, et al: **The draft genomes of soft--shell turtle and green sea turtle yield insights into the development and evolution of the turtle--specific body plan.** Nat Genet 2013, 45:701--706.
66. Lesot H, Hovorakova M, Peterka M, Peterkova R: **Three--dimensional analysis of molar development in the mouse from the cap to bell stage.** Aust Dent J 2014, 59 Suppl 1:81--100.
67. Rothova M, Peterkova R, Tucker AS: **Fate map of the dental mesenchyme: dynamic development of the dental papilla and follicle.** Dev Biol 2012, 366:244--254.
68. Giese A, Bjerkvig R, Berens ME, Westphal M: **Cost of migration: invasion of malignant gliomas and implications for treatment.** J Clin Oncol 2003, 21:1624--1636.
69. Duncan T, Su TT: **Embryogenesis: coordinating cell division with gastrulation.** Curr Biol 2004, 14:R305--307.
70. Francesconi M, Lehner B: **The effects of genetic variation on gene expression dynamics during development.** Nature 2014, 505:208--211.
71. Tsang JC, Yu Y, Burke S, Buettner F, Wang C, Kolodziejczyk AA, Teichmann SA, Lu L, Liu P: Single--cell transcriptomic reconstruction reveals cell cycle and multi--lineage differentiation defects in Bcl11a--deficient hematopoietic stem cells. Genome Biol 2015, 16:178.
72. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, Schneider RK, Wagers AJ, Ebert BL, Regev A: **Single--cell RNA--seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells.** Genome Res 2015, 25:1860--1872.
73. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O: **Computational analysis of cell--to--cell heterogeneity in single-- cell RNA--sequencing data reveals hidden subpopulations of cells.** Nat Biotechnol 2015, 33:155--160.

74. O'Duibhir E, Lijnzaad P, Benschop JJ, Lenstra TL, van Leenen D, Groot Koerkamp MJ, Margaritis T, Brok MO, Kemmeren P, Holstege FC: **Cell cycle population effects in perturbation studies.** Mol Syst Biol 2014, 10:732.

75. Israel JW, Martik ML, Byrne M, Raff EC, Raff RA, McClay DR, Wray GA: **Comparative Developmental Transcriptomics Reveals Rewiring of a Highly Conserved Gene Regulatory Network during a Major Life History Switch in the Sea Urchin Genus Heliocidaris.** PLoS Biol 2016, 14:e1002391.

76. Fushan AA, Turanov AA, Lee SG, Kim EB, Lobanov AV, Yim SH, Buffenstein R, Lee SR, Chang KT, Rhee H, et al: **Gene expression defines natural changes in mammalian lifespan.** Aging Cell 2015, 14:352--365.

77. Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu--Petri A, Kircher M, et al: **The evolution of gene expression levels in mammalian organs.** Nature 2011, 478:343--348.

78. Gould S: **Ontogeny and Phylogeny.** Belknap Press of Harvard University Press; 1977.

79. Liu X, Somel M, Tang L, Yan Z, Jiang X, Guo S, Yuan Y, He L, Oleksiak A, Zhang Y, et al: **Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques.** Genome Res 2012, 22:611--622.

80. Somel M, Franz H, Yan Z, Lorenc A, Guo S, Giger T, Kelso J, Nickel B, Dannemann M, Bahn S, et al: **Transcriptional neoteny in the human brain.** Proc Natl Acad Sci U S A 2009, 106:5743--5748.

81. Somel M, Rohlfs R, Liu X: **Transcriptomic insights into human brain evolution: acceleration, neutrality, heterochrony.** Current Opinion in Genetics & Development 2014, 29:110--119.

82. Yanai I, Peshkin L, Jorgensen P, Kirschner MW: **Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility.** Dev Cell 2011, 20:483--496.

83. Yuan Y, Chen YP, Ni S, Xu AG, Tang L, Vingron M, Somel M, Khaitovich P: **Development and application of a modified dynamic time warping algorithm (DTW--S) to analyses of primate brain expression time series.** BMC Bioinformatics 2011, 12:347.

84. Raff RA WG: **Heterochrony: developmental mechanisms and evolutionary results.** J Evol Biol 1989, 2:409–434.

85. Peterka M, Lesot H, Peterkova R: **Body weight in mouse embryos specifies staging of tooth development.** Connect Tissue Res 2002, 43:186--190.

86. Jernvall J, Aberg T, Kettunen P, Keranen S, Thesleff I: **The life history of an embryonic signaling center: BMP--4 induces p21 and is associated with apoptosis in the mouse tooth enamel knot.** Development 1998, 125:161--169.

87. `https://github.com/msemon/ToothTranscriptomeAnalyses`

88. Langmead B, Salzberg SL: **Fast gapped--read alignment with Bowtie 2.** Nat Methods 2012, 9:357--359.

89. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: **TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.** Genome Biol 2013, 14:R36.

90. Anders S, Pyl PT, Huber W: **HTSeq-—a Python framework to work with high-- throughput sequencing data.** Bioinformatics 2015, 31:166--169.

91. R: A language and environment for statistical computing. [`https://http://www.R-\ discretionary{-}{}{}-project.org/`.]

92. Wickham H: **ggplot2: Elegant Graphics for Data Analysis.** Springer--Verlag New York; 2009.

93. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA--seq data with DESeq2**. Genome Biol 2014, 15:550.

94. Dray SaD, A.B.: **The ade4 package: implementing the duality diagram for ecologists.** Journal of Statistical Software 2007, 22:1--20.

95. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z: **GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists.** BMC Bioinformatics 2009, 10:48.

96. Falcon S, Gentleman R: **Using GOstats to test gene lists for GO term association.** Bioinformatics 2007, 23:257--258.

97. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I: **Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer**. Nature 2015, 519:219--222.

98. Gaujoux R, Seoighe C: **CellMix: a comprehensive toolbox for gene expression deconvolution.** Bioinformatics 2013, 29:2211--2212.

99. Zhong Y, Wan YW, Pang K, Chow LM, Liu Z: **Digital sorting of complex tissues for cell type--specific gene expression profiles.** BMC Bioinformatics 2013, 14:89.

100. `http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE76316`

101. Gruneberg H: **Genes and genotypes affecting the teeth of the mouse.** J Embryol Exp Morphol 1965, 14:137--159.

102. Lochovska K, Peterkova R, Pavlikova Z, Hovorakova M: **Sprouty gene dosage influences temporal---spatial dynamics of primary enamel knot formation.** BMC Dev Biol 2015, 15:21.

103. Prochazka J, Pantalacci S, Churava S, Rothova M, Lambert A, Lesot H, Klein O, Peterka M, Laudet V, Peterkova R: **Patterning by heritage in mouse molar row development.** Proc Natl Acad Sci U S A 2010, 107:15497--15502.

104. Cameron TL, Belluoccio D, Farlie PG, Brachvogel B, Bateman JF: **Global comparative transcriptome analysis of cartilage formation in vivo.** BMC Dev Biol 2009, 9:20.

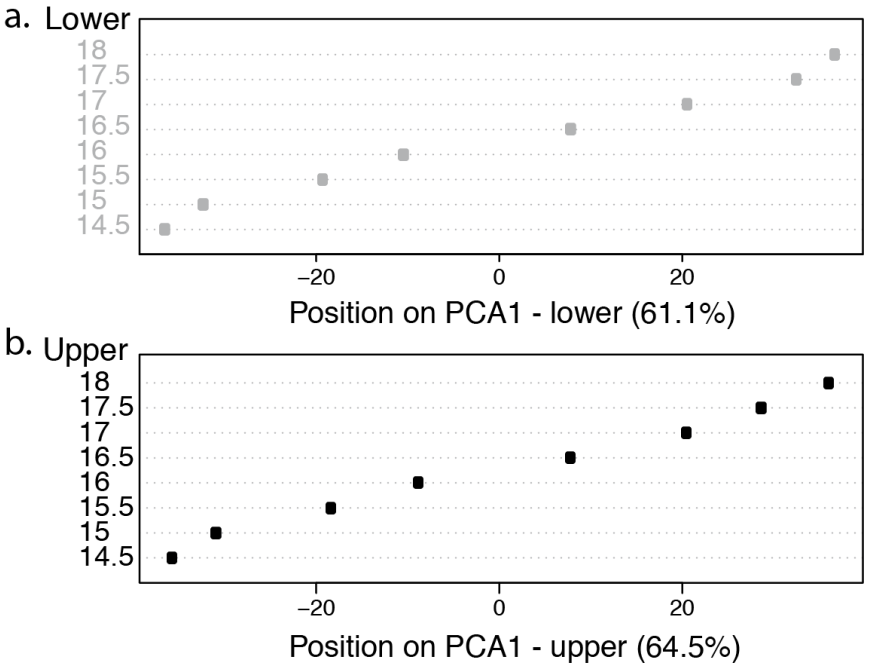
105. Ahnfelt-Ronne et al. J Hist Cytochem 2007 55(9):925-30

106. Yokomizo and Dzierzak Development 2010, 137 :3651-61

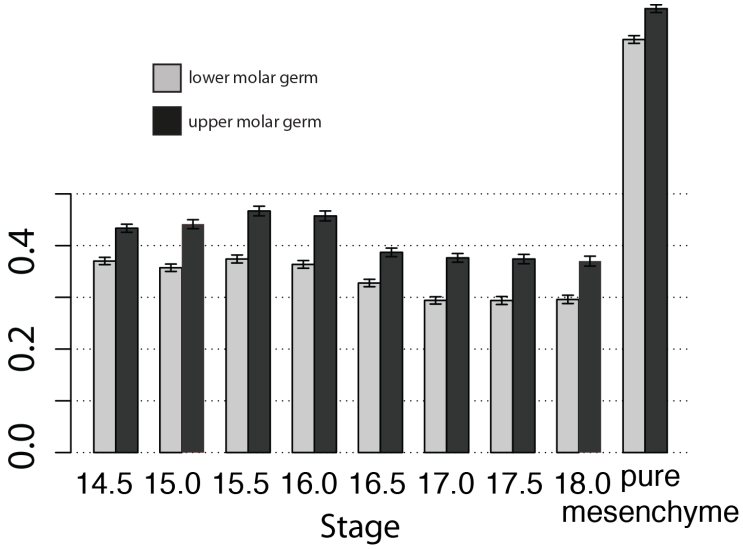
233

# SUPPLEMENTARY FIGURES

## Supplementary figures



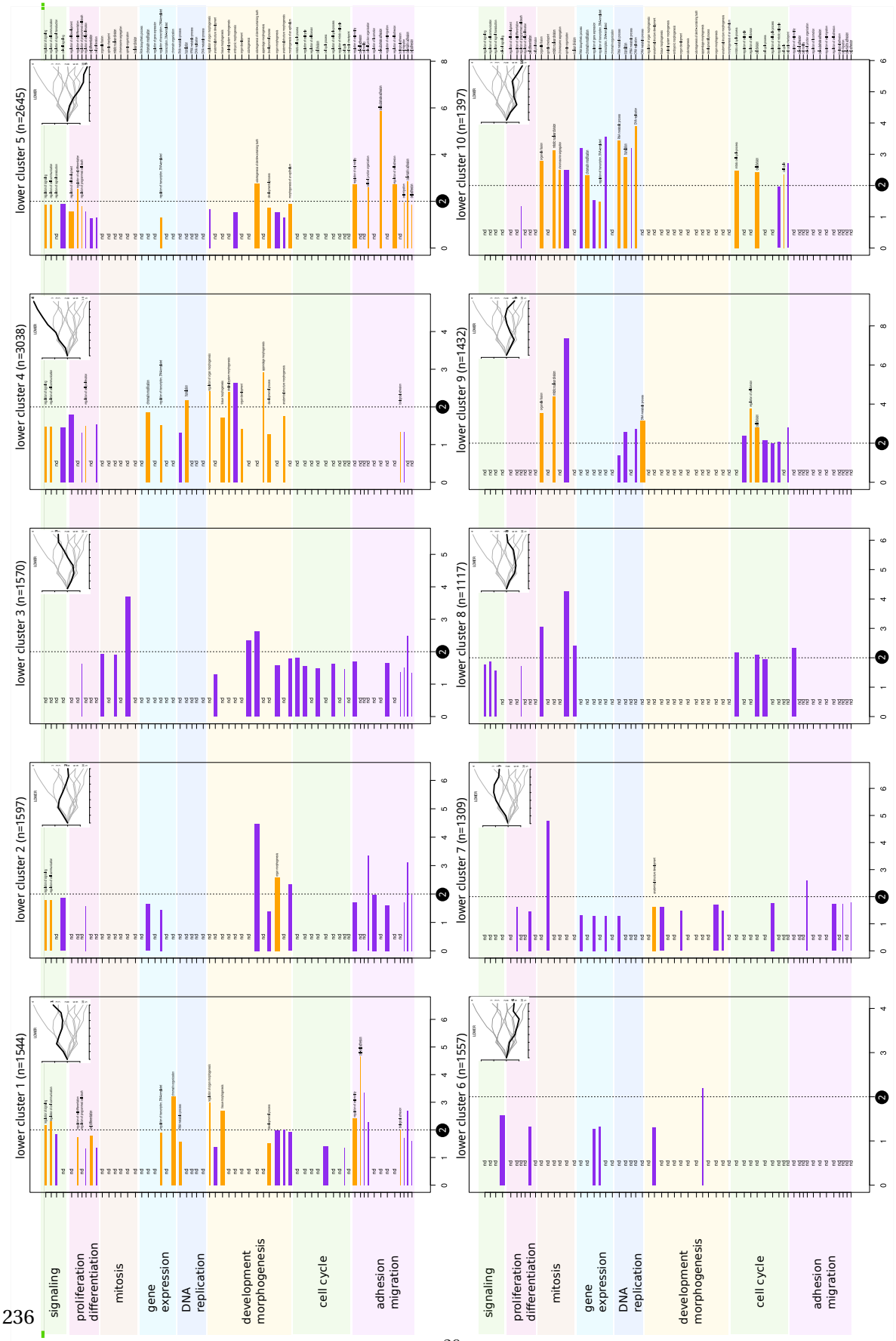

**SUP Fig 1 - First axis of the lower and upper-specific PCAs orders samples with time.** **(a)** The location of lower molar germs on the first axis of a PCA drawn specifically with the 8 lower samples. **(b)** The location of upper molar germs on the first axis of a PCA drawn specifically with the 8 upper samples.




**SUP Fig 3 - Mesenchyme proportion estimated using deconvolutions with markers based on microarray data.** Markers were obtained from microarray data ([39]; 18 for EK, 9 for epithelium and 27 for mesenchyme). Confidence intervals were extracted by resampling (500 bootstraps). RNA-seq samples for tooth mesenchyme [45] were used as positive controls.

**SUP Fig 2 - Gene ontology analysis of the genes robustly assigned to one of the 10 clusters obtained for lower and upper molar.** The columns represent the ten main temporal profiles of expression in the lower (**a.** top) and upper (**b.** bottom) tooth, as described Figure 3. Cluster numbering is random (same as in Figure 3), and do not coincide between upper and lower tooth. The number of genes is indicated for each cluster, and the corresponding profile is in represented in bold. Lists of genes assigned to a each cluster were submitted to GO analysis using GOstats. Enrichment for a list of relevant GO terms (grouped by 8 broad functional category, see colors on the left side) are represented. The width of each bar is proportional to the size of the GO term (in number of genes). The length of each bar represents the enrichment (measured by an odd ratio), orange bars corresponding to GO Terms which are significantly enriched in the analysis (adjusted pvalue <0.1), and purple bars corresponding to GO Terms which are not significantly enriched. n.d means that no genes assigned to this GO Term belong to the cluster.

**SUP Fig 4 - Mesenchyme proportion are always larger in upper molar as measured from 3D reconstructions of dissected tooth germs.** **(a)** A series of confocal sections of a dissected ED15.0 lower molar germ stained with pan- cytokeratin antibody. **(b)** Epithelial component and total whole germ 3D reconstructions from the sections shown in (a). **(c)** Table showing measures of mesenchyme proportion from 5 pairs of 3D- reconstructed lower-upper molar germs. Supplementary material for this figure: Tooth germs dissected from littermate embryos of RNA-seq samples (same weight) were fixed overnight in PFA4% and dehydrated through a methanol series. *In toto* immunolocalisation protocol was adapted from [105] . Following incubation in H202 5%, DMSO 10% in methanol for 4 hours, they were rehydrated, blocked and incubated successively with a pan- cytokeratin antibody (overnight,1/200, Novus Biologicals) and a Dylight 549 conjugated Donkey Anti-rabbit antibody (overnight 1/200, Jackson immunoresearch) and finally with Hoechst (overnight , 50$mu$g/ml). Following dehydration, they were clarified and mounted in BABB as described in [106]. They were imaged with a Zeiss LSM710 confocal microscope at the PLATIM (Lyon, France). The outline of the epithelium labeled by the pan-cytokeratin antibody and the outline of the tooth germ labeled with hoechst were delineated manually in the AMIRA software. Total epithelium and total tooth germ were reconstructed in 3D and their volume measured in the AMIRA software as well.

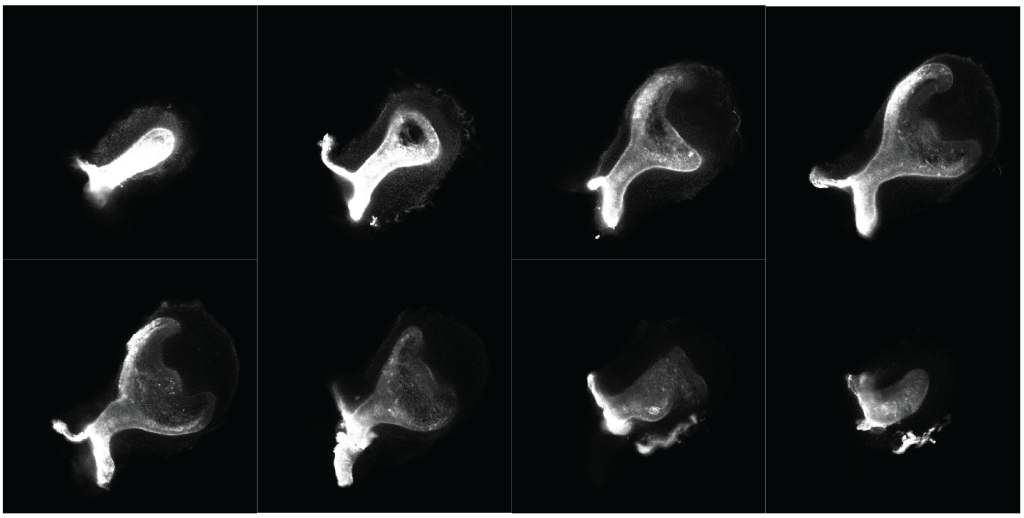




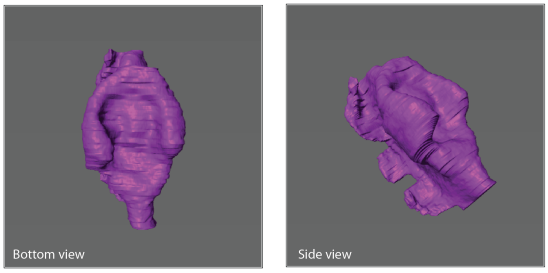



**SUP Fig 2 - Gene ontology analysis of the genes robustly assigned to one of the 10 clusters obtained for lower and upper molar.**

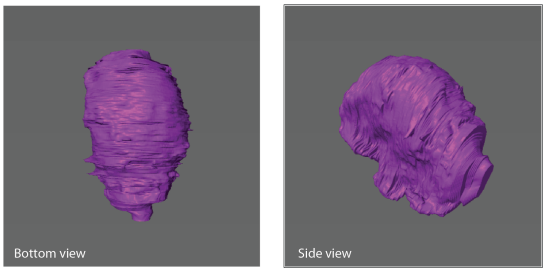
a. A series of confocal sections of a dissected ED15.0 lower molar germ stained with pan-cytokeratin antibody



b.Epithelial component and total whole germ 3D reconstructions from the sections shown in a.




3D-reconstructed epithelial component

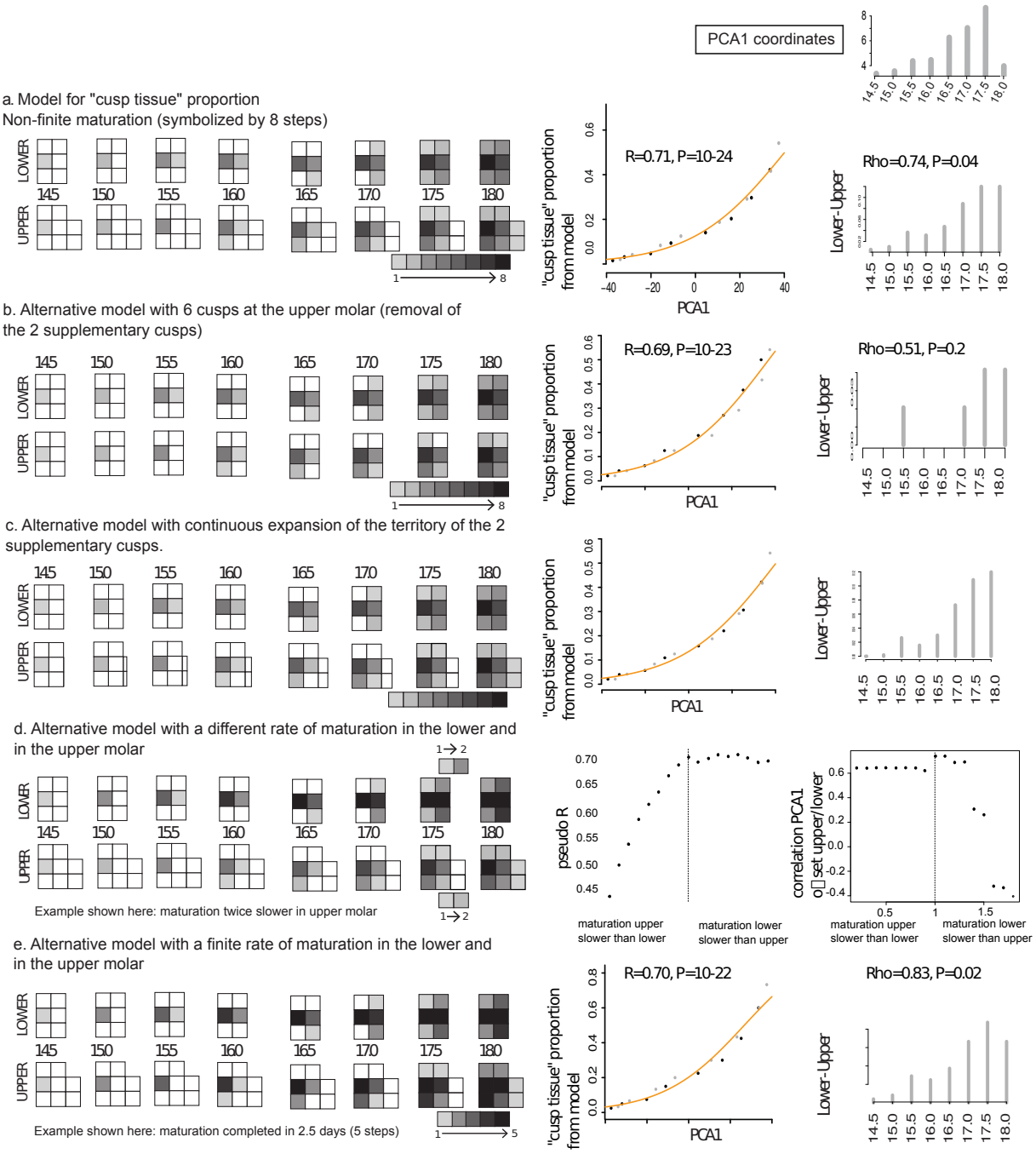



3D-reconstructed total molar germ

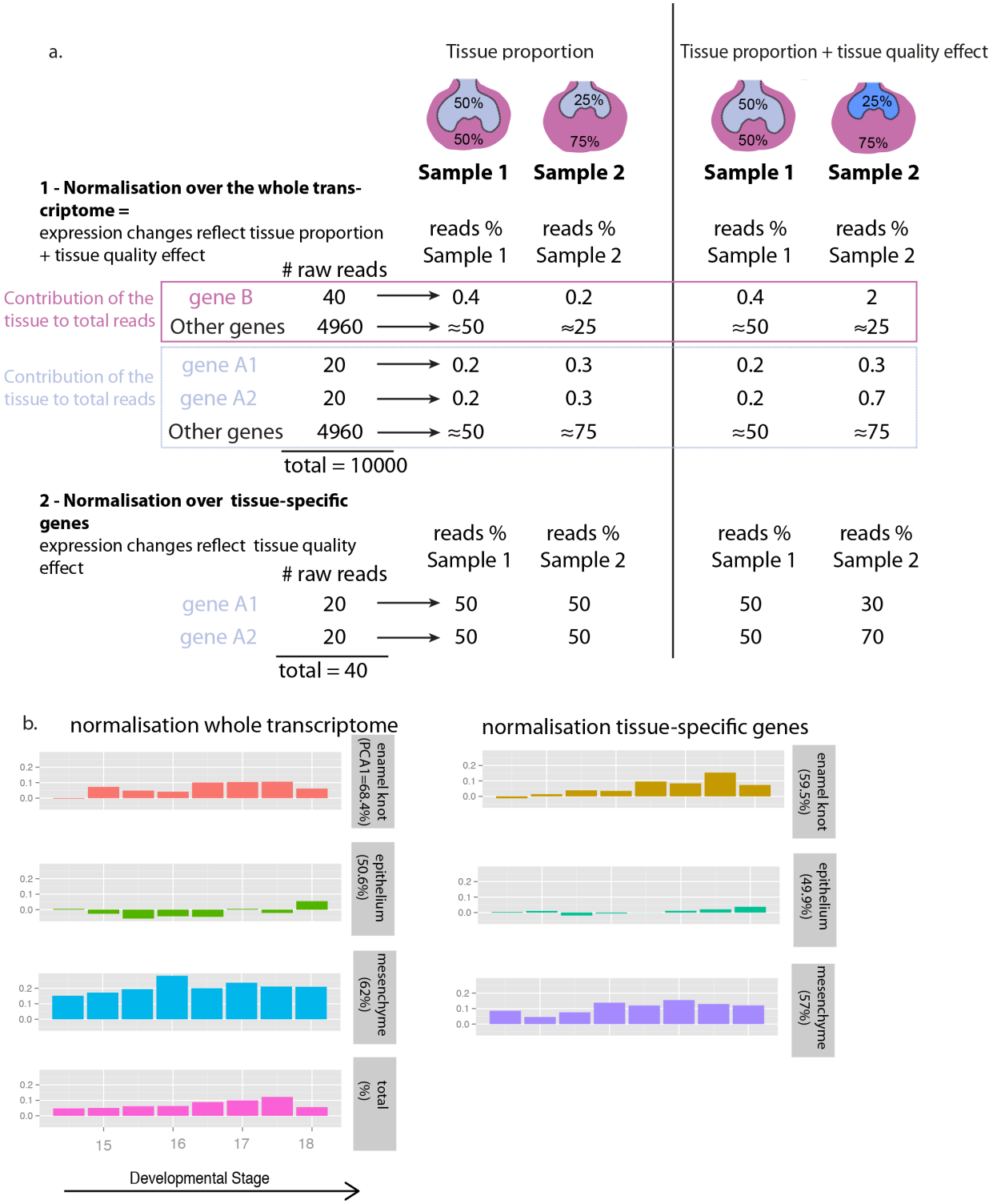
c. Measure of mesenchyme proportion from 5 pairs of 3D- reconstructed lower-upper molar germs

| | | measured volumes | | % of total volume | | |
|---|---|---|---|---|---|---|
| | | Epithelium volume | Total volume | Epithelium | mesenchyme | % Vmes(upper)-%Vmes(lower) |
| ED14.5 | lower | 5030708 | 14011360 | 35,9 | 64,1 | |
| | upper | 6833440 | 23084740 | 29,6 | 70,4 | 6,3 |
| ED14.5 | lower | 5984882 | 20390236 | 29,4 | 70,6 | |
| | upper | 7109809 | 26925336 | 26,4 | 73,6 | 3,0 |
| ED15.0 | lower | 5004660 | 12344049 | 40,5 | 59,5 | |
| | upper | 5154652 | 13540549 | 38,1 | 61,9 | 2,5 |
| ED15.0 | lower | 9356574 | 32496996 | 28,8 | 71,2 | |
| | upper | 10398992 | 40934108 | 25,4 | 74,6 | 3,4 |
| ED15.5 | lower | 14238123 | 36044588 | 39,5 | 60,5 | |
| | upper | 13448003 | 37602856 | 35,8 | 64,2 | 3,7 |

**SUP Fig 4 - Mesenchyme proportion are always larger in upper molar as measured from 3D reconstructions of dissected tooth germs.**




**SUP Fig 5 -Models of cusp patterning and expansion, in the lower and upper molar.**




**SUP Fig 6 - Heterochrony signals are visible in the transcriptomes of each tissue compartment.**

## CAPTIONS OF SUPPLEMENTARY FIGURES

**SUP Fig 5 -Models of cusp patterning and expansion, in the lower and upper molar.** Model of cusp expansion in the lower and upper molars were build using in situ data on the timing of cusp patterning. Once patterned, the territory of each cusp expands at the same speed. In the schematic representation, the average shade of grey of each tooth, at each developmental stage, gives a visual impression on the average degree of the expansion of the cusp territory, from unpatterned cusp (white) to just-patterned cusp (light grey) and fully expanded cusp (black). We computed cusp expansion in each tooth based on this model, 0 corresponding to no cusp patterned (all white), and 1 to a complete expansion (all black), which we will later refer to as "cusp tissue" proportion from model. **(a)** Left : Model applied to the upper and lower tooth, following the sequence of cusp addition that ranges from 1 to 8 cusps in upper molar, and from 1 to 6 cusps in lower molar (see Figure 5e); The time for complete expansion is set to 8 steps, so that the first cusp to pattern (at stage 14.5) is fully expanded at the last stage of the series (at stage ED 18.0). Middle : y-axis represents the estimated "cusp tissue" proportions from the model, x axis represents the position of the samples on the first PCA axis. P-value associated to a binomial regression = 10-24, McFadden's pseudo R squared = 0.71. Right : Offset of cusp proportion as estimated from the model (corresponding to Lower-Upper), for each developmental stage. With respect to the overall fit to PCA1 (middle column), none of the three alternative models (b-d) perform better than this main model (a), and they are all worse considering specifically upper/lower heterochrony. Legend for b-d follows (a). **(b)** First alternative model, in which the two supplementary cusps are omitted. This implies that the formation of extra cusps does not translate in a change in cusp proportions. Middle : P = 10-23, pseudo R squared = 0.69. **(c)** Second alternative model, in which upper tooth is progressively acquiring the extra space corresponding to the extra cusps. Middle: P = 10-22, pseudo R squared = 0.68.

**(d)** Third alternative model, in which the rate of maturation is different in the lower and upper molars. Right: The example shows a 50% slower rate in upper molar. Middle : P = 10-23, pseudo R squared = 0.69. Middle : We build 17 models by making the relative rates of maturation vary between upper and lower molars (from a ratio of upper/lower rates = 0.2 to upper/lower rates = 1.8, by steps of 0.1). We assessed each model through McFadden's pseudo R squared, which are represented in the middle panel. It is obvious that models with upper/lower ratio $< 1$ (that is, upper having a slower rate than lower) fit less well to the PCA. We also assessed each model through the correlation between the heterochrony estimated by the model and the heterochrony estimated by the PCA1 (Right). It is obvious that the heterochrony is best recapitulated by models with equals rates between upper and lower molars, or by models with slower maturation in lower molar. From the middle and the right panel taken together, it can be deduced that the best model(s) have roughly equals maturation rates in upper and lower molars. **(e)** Finite cusp expansion : Model like in (a) but with reducing the time for each cusp full expansion. We reduced this time, from 8 steps (corresponding to the entire time serie, therefore to the infinite maturation model mentioned above) down to 2 steps. The pattern of heterochrony observed with the PCA is best recapitulated by a finite model which allows a duration of 2 days and a half for each cusp to expand (5 days, exemplified here). Note that the heterochrony (right) is slightly better than the one observed in (a), in that it correlates better with PCA1 coordinates.

**SUP Fig 6 - Heterochrony signals are visible in the transcriptomes of each tissue compartment. (a)** Schematic explanation of the different effect measured when the normalization of RNAseq counts is performed on the whole transcriptome or over tissue specific genes only. Two examples are taken. **Left:** On the first example (tissue proportion, left) two samples are compared, for which cellular expression levels are the same (schematized here by the same pink and blue colors in both samples), but tissue proportions differ between samples (here, 50% epithelium in sample1, 25% epithelium in sample 2). Because of these differences in tissue proportions, the expression level of tissue-specific genes (like gene B, A1, A2 here) will be different between sample 1 and sample 2. The expression of gene A1, for instance, appears to be different when normalization is performed on the whole dataset, but is equal when normalization is done on tissue-specific genes. **Right:** On the second example (right), both tissue proportions and cellular expression levels differ between stages (schematized here by a different blue color in the samples). The expression of gene A1, for instance, remains different when normalisation is done on tissue- specific genes. **(b)** Heterochrony measured on PCA drawn using the whole dataset (total), and with 3 subsets of genes, each specific to one of 3 tissue compartments (336 epithelium markers, 421 EK markers and 566 mesenchyme markers). Normalization was done on the whole transcriptome (left) or on each set of markers (right, normalization on tissue-specific genes). For each dataset, the first axis ordered the samples with their developmental advancement. The height of each bar represents the heterochrony measured on the first PCA axis in each case (that is, coordinate of lower sample on PCA1 – coordinate of upper sample on PCA1, for each time point).

# Bibliographie


Åberg, Thomas, Wozney, John, & Thesleff, Irma. 1997. Expression patterns of bone morphogenetic proteins (Bmps) in the developing mouse tooth suggest roles in morphogenesis and cell differentiation. *Developmental Dynamics*, **210**(4), 383–396.

Åberg, Thomas, Cavender, Adriana, Gaikwad, Joel S, Bronckers, Antonius LJJ, Wang, Xiuping, Waltimo-Sirén, Janna, Thesleff, Irma, & D'Souza, Rena N. 2004a. Phenotypic changes in dentition of Runx2 homozygote-null mutant mice. *Journal of Histochemistry & Cytochemistry*, **52**(1), 131–139.

Åberg, Thomas, Wang, Xiu-Ping, Kim, Jung-Hwan, Yamashiro, Takashi, Bei, Marianna, Rice, Ritva, Ryoo, Hyun-Mo, & Thesleff, Irma. 2004b. Runx2 mediates FGF signaling from epithelium to mesenchyme during tooth morphogenesis. *Developmental Biology*, **270**(1), 76–93.

Abler, Lisa L., Mansour, Suzanne L., & Sun, Xin. 2009. Conditional gene inactivation reveals roles forFgf10andFgfr2in establishing a normal pattern of epithelial branching in the mouse lung. *Developmental Dynamics*, **238**(8), 1999–2013.

Abzhanov, Arhat. 2013. von Baer's law for the ages : lost and found principles of developmental evolution. *Trends in Genetics*, **29**(12), 712–722.

Abzhanov, Arhat, Protas, Merediddth, Grant, B. Rosemary, Grant, Peter R., & Tabin, Clifford J. 2004. Bmp4 and Morphological Variation of Beaks in Darwin's Finches. *Science*, **305**(5689), 1462–1465.

Abzhanov, Arhat, Kuo, Winston P., Hartmann, Christine, Grant, B. Rosemary, Grant, Peter R., & Tabin, Clifford J. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, **442**(7102), 563–567.

Alexa, Adrian, & Rahnenfuhrer, Jorg. 2016. *topGO : Enrichment Analysis for Gene Ontology*. R package version 2.24.0 (https://bioconductor.org/packages/release/bioc/vignettes/topGO/inst/doc/topGO.pdf).

Ambrosio, Andrea L., Taelman, Vincent F., Lee, Hojoon X., Metzinger, Carrie A., Coffinier, Catherine, & Robertis, E.M. De. 2008. Crossveinless-2 Is a BMP Feedback Inhibitor that Binds Chordin/BMP to Regulate Xenopus Embryonic Patterning. *Developmental Cell*, **15**(2), 248–260.

Amthor, Helge, Christ, Bodo, Rashid-Doubell, Fiza, Kemp, C.Fred, Lang, Emily, & Patel, Ketan. 2002. Follistatin Regulates Bone Morphogenetic Protein-7 (BMP-7) Activity to Stimulate Embryonic Muscle Growth. *Developmental Biology*, **243**(1), 115–127.

Anavy, L., Levin, M., Khair, S., Nakanishi, N., Fernandez-Valverde, S. L., Degnan, B. M., & Yanai, I. 2014. BLIND ordering of large-scale transcriptomic developmental timecourses. *Development*, **141**(5), 1161–1166.

Anders, S., Pyl, P. T., & Huber, W. 2014. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**(2), 166–169.

Angerer, Philipp, Haghverdi, Laleh, Büttner, Maren, Theis, Fabian J., Marr, Carsten, & Buettner, Florian. 2015. destiny : diffusion maps for large-scale single-cell data in R. *Bioinformatics*, **32**(8), 1241–1243.

Arendt, Detlev. 2008. The evolution of cell types in animals : emerging principles from molecular studies. *Nat Rev Genet*, **9**(11), 868–882.

Averof, Michalis, & Patel, Nipam H. 1997. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature*, **388**(June), 682–686.

Baer, Charles F., Miyamoto, Michael M., & Denver, Dee R. 2007. Mutation rate variation in multicellular eukaryotes : causes and consequences. *Nature Reviews Genetics*, **8**(8), 619–631.

Balic, Anamaria, & Thesleff, Irma. 2015. Chapter Seven - Tissue Interactions Regulating Tooth Development and Renewal. *Pages 157 – 186 of :* Chai, Yang (ed), *Craniofacial Development*. Current Topics in Developmental Biology, vol. 115. Academic Press.

Bates, C. M. 2011. Role of fibroblast growth factor receptor signaling in kidney development. *AJP : Renal Physiology*, **301**(2), F245–F251.

Bei, Marianna, & Maas, Richard. 1998. FGFs and BMP4 induce both Msx1-independent and Msx1-dependent signaling pathways in early tooth development. *Development*, **125**(21), 4325–4333.

Bei, Marianna, Kratochwil, Klaus, & Maas, Richard L. 2000. BMP4 rescues a non-cell-autonomous function of Msx1 in tooth development. *Development*, **127**(21), 4711–4718.

Bertrand, Stéphanie, Brunet, Frédéric G., Escriva, Hector, Parmentier, Gilles, Laudet, Vincent, & Robinson-Rechavi, Marc. 2004. Evolutionary Genomics of Nuclear Receptors : From Twenty-Five Ancestral Genes to Derived Endocrine Systems. *Molecular Biology and Evolution*, **21**(10), 1923–1937.

Beverdam, Annemiek, Merlo, Giorgio R., Paleari, Laura, Mantero, Stefano, Genova, Francesca, Barbieri, Ottavia, Janvier, Philippe, & Levi, Giovanni. 2002. Jaw transformation with gain of symmetry afterDlx5/Dlx6 inactivation : Mirror of the past ? *genesis*, **34**(4), 221–227.

Bolker, Jessica A. 2000. Modularity in development and why it matters to evo-devo. *American Zoologist*, **40**(5), 770–776.

Brawand, David, Soumillon, Magali, Necsulea, Anamaria, Julien, Philippe, Csárdi, Gábor, Harrigan, Patrick, Weier, Manuela, Liechti, Angélica, Aximu-Petri, Ayinuer, Kircher, Martin, Albert, Frank W., Zeller, Ulrich, Khaitovich, Philipp, Grützner, Frank, Bergmann, Sven, Nielsen, Rasmus, Pääbo, Svante, & Kaessmann, Henrik. 2011. The evolution of gene expression levels in mammalian organs. *Nature*, **478**(7369), 343–348.

Bray, Nicolas L, Pimentel, Harold, Melsted, Páll, & Pachter, Lior. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, **34**(5), 525–527.

Breschi, Alessandra, Djebali, Sarah, Gillis, Jesse, Pervouchine, Dmitri D., Dobin, Alex, Davis, Carrie A., Gingeras, Thomas R., & Guigó, Roderic. 2016. Gene-specific patterns of expression variation across organs and species. *Genome Biol*, **17**(1).

Briggs, Derek E. G., Clarkson, Euan N. K., & Aldridge, Richard J. 1983. The conodont animal. *Lethaia*, **16**(1), 1–14.

Buchtová, Marcela, Handrigan, Gregory R., Tucker, Abigail S., Lozanoff, Scott, Town, Liam, Fu, Katherine, Diewert, Virginia M., Wicking, Carol, & Richman, Joy M. 2008. Initiation and patterning of the snake dentition are dependent on Sonic Hedgehog signaling. *Developmental Biology*, **319**(1), 132–145.

Burrow, C. J. 2003. Comment on "Separate Evolutionary Origins of Teeth from Evidence in Fossil Jawed Vertebrates". *Science*, **300**(5626), 1661b–1661.

Butler, P. M. 1956. The ontogeny of molar pattern. *Biological Reviews*, **31**(1), 30–69.

Butler, Percy Milton. 1995. Ontogenetic aspects of dental evolution. *International Journal of Developmental Biology*, **39**(1), 25–34.

Butler, PM. 1990. Early trends in the evolution of tribosphenic molars. *Biological Reviews*, **65**(4), 529–552.

Cai, Jinglei, Cho, Sung-Won, Kim, Jae-Young, Lee, Min-Jung, Cha, Yoon-Geun, & Jung, Han-Sung. 2007. Patterning the size and number of tooth and its cusps. *Developmental Biology*, **304**(2), 499–507.

Cam, Yves, Fausser, Jean-Luc, Vonesch, Jean-Luc, Peterkova, Renata, Peterka, Miroslav, Halaskova, Milada, & Lesot, Herve. 2002. Asymmetrical morphogenesis and medio-lateral positioning of molars during mouse development. *Eur J Oral Sci*, **110**(1), 35–43.

Caparros-Martin, J. A., Valencia, M., Reytor, E., Pacheco, M., Fernandez, M., Perez-Aytes, A., Gean, E., Lapunzina, P., Peters, H., Goodship, J. A., & Ruiz-Perez, V. L. 2012. The ciliary Evc/Evc2 complex interacts with Smo and controls Hedgehog pathway activity in chondrocytes by regulating Sufu/Gli3 dissociation and Gli3 trafficking in primary cilia. *Human Molecular Genetics*, **22**(1), 124–139.

Carkett, Martin D, & Logan, Malcolm PO. 2011. 1, 2, 3 : Counting the fingers on a chicken wing. *Genome Biol*, **12**(10), 130.

Carroll, Sean B. 2008. Evo-Devo and an Expanding Evolutionary Synthesis : A Genetic Theory of Morphological Evolution. *Cell*, **134**(1), 25–36.

Catón, Javier, & Tucker, Abigail S. 2009. Current knowledge of tooth development : patterning and mineralization of the murine dentition. *Journal of Anatomy*, **214**(4), 502–515.

Chan, Esther T, Quon, Gerald T, Chua, Gordon, Babak, Tomas, Trochesset, Miles, Zirngibl, Ralph A, Aubin, Jane, Ratcliffe, Michael JH, Wilde, Andrew, Brudno, Michael, Morris, Quaid D, & Hughes, Timothy R. 2009. Conservation of core gene expression in vertebrate tissues. *Journal of Biology*, **8**(3), 33.

Charles, C., Lazzari, V., Tafforeau, P., Schimmang, T., Tekin, M., Klein, O., & Viriot, L. 2009a. Modulation of Fgf3 dosage in mouse and men mirrors evolution of mammalian dentition. *Proceedings of the National Academy of Sciences*, **106**(52), 22364–22368.

Charles, Cyril, Pantalacci, Sophie, Peterkova, Renata, Tafforeau, Paul, Laudet, Vincent, & Viriot, Laurent. 2009b. Effect ofedaLoss of Function on Upper Jugal Tooth Morphology. *The Anatomical Record : Advances in Integrative Anatomy and Evolutionary Biology*, **292**(2), 299–308.

Chen, Jianquan, Lan, Yu, Baek, Jin-A, Gao, Yang, & Jiang, Rulang. 2009. Wnt/$\beta$-catenin signaling plays an essential role in activation of odontogenic mesenchyme during early tooth development. *Developmental Biology*, **334**(1), 174–185.

Chen, Miao-Hsueh, Wilson, Christopher W., & Chuang, Pao-Tien. 2007. SnapShot : Hedgehog Signaling Pathway. *Cell*, **130**(2), 386–386.e2.

Chen, Y, Bei, Marianna, Woo, Ian, Satokata, Ichiro, & Maas, Richard. 1996. Msx1 controls inductive signaling in mammalian tooth morphogenesis. *Development*, **122**(10), 3035–3044.

Cho, S.-W., Kwak, S., Woolley, T. E., Lee, M.-J., Kim, E.-J., Baker, R. E., Kim, H.-J., Shin, J.-S., Tickle, C., Maini, P. K., & Jung, H.-S. 2011. Interactions between Shh, Sostdc1 and Wnt signaling and a new feedback loop for spatial patterning of the teeth. *Development*, **138**(9), 1807–1816.

Cho, Sung-Won, Lee, Hyun-A, Cai, Jinglei, Lee, Min-Jung, Kim, Jae-Young, Ohshima, Hayato, & Jung, Han-Sung. 2007. The primary enamel knot determines the position of the first buccal cusp in developing mice molars. *Differentiation*, **75**(5), 441–451.

Clevers, Hans, & Nusse, Roel. 2012. Wnt/$\beta$-Catenin Signaling and Disease. *Cell*, **149**(6), 1192–1205.

Cobourne, Martyn T, & Sharpe, Paul T. 2003. Tooth and jaw : molecular mechanisms of patterning in the first branchial arch. *Archives of Oral Biology*, **48**(1), 1–14.

Cohn, Martin J. 2011. Development of the external genitalia : Conserved and divergent mechanisms of appendage patterning. *Dev. Dyn.*, **240**(5), 1108–1115.

Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F., & Zucker, S. W. 2005. Geometric diffusions as a tool for harmonic analysis and structure definition of data : Diffusion maps. *Proceedings of the National Academy of Sciences*, **102**(21), 7426–7431.

Coifman, R. R., Kevrekidis, I. G., Lafon, S., Maggioni, M., & Nadler, B. 2008. Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems. *Multiscale Modeling & Simulation*, **7**(2), 842–864.

Coin, ROMUALD, Lesot, HERVÉ, Vonesch, JL, Haikel, YOUSSEF, & Ruch, JEAN-VICTOR. 1999. Aspects of cell proliferation kinetics of the inner dental epithelium during mouse molar and incisor morphogenesis : a reappraisal of the role of the enamel knot area. *International Journal of Developmental Biology*, **43**(3), 261–267.

Coin, Romuald, Kieffer, Sandrine, Lesot, Hervé, Vonesch, Jl, & Ruch, Jean-victor. 2000. Inhibition of apoptosis in the primary enamel knot does not affect specific tooth crown morphogenesis in the mouse. *International Journal of Developmental Biology*, **44**(4), 389–396.

Comte, Aurélie, Roux, Julien, & Robinson-Rechavi, Marc. 2010. Molecular signaling in zebrafish development and the vertebrate phylotypic period. *Evolution & Development*, **12**(2), 144–156.

Dailey, Lisa, Ambrosetti, Davide, Mansukhani, Alka, & Basilico, Claudio. 2005. Mechanisms underlying differential responses to FGF signaling. *Cytokine & Growth Factor Reviews*, **16**(2), 233–247.

Darwin, Charles. 1859. *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life /althop.* Smithsonian Institution.

Dassule, Hélène R, Lewis, Paula, Bei, Marianna, Maas, Richard, & McMahon, Andrew P. 2000. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development*, **127**(22), 4775–4785.

Davidson, Eric H. 2010. Emerging properties of animal gene regulatory networks. *Nature*, **468**(7326), 911–920.

De la Porte, J, Herbst, BM, Hereman, W, & Van Der Walt, SJ. 2008. An introduction to diffusion maps. *In : The 19th Symposium of the Pattern Recognition Association of South Africa.*

de Pinna, Mario C. C. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics*, **7**(4), 367–394.

Delafontaine, Julien. 2015. *L'analyse en composantes principales (avec R). http://bioinfo-fr.net/lanalyse-en-composantes-principales-avec-r?s=analyse+en+composante+principale*.

Depew, M. J. 2002. Specification of Jaw Subdivisions by Dlx Genes. *Science*, **298**(5592), 381–385.

Domazet-Lošo, Tomislav, & Tautz, Diethard. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*, **468**(7325), 815–818.

Dover, G. A. 1992. Observing Development Through Evolutionary Eyes : A Practical Approach. *Bioessays*, **Vol. 14**(4), 281–287.

Dray, Sthephane, & Dufour, Anne-Beatrice. 2007. The ade4 package : implementing the duality diagram for ecologists. *Journal of statistical software*, **22**(1), 1–20.

Drost, H.-G., Gabel, A., Grosse, I., & Quint, M. 2015. Evidence for Active Maintenance of Phylotranscriptomic Hourglass Patterns in Animal and Plant Embryogenesis. *Molecular Biology and Evolution*, **32**(5), 1221–1231.

Duboule, Denis. 1994. Temporal colinearity and the phylotypic progression : a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development*, **1994**(Supplement), 135–142.

Duboule, Denis, & Morata, Gines. 1994. Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends in Genetics*, **10**(10), 358–364.

Duret, Laurent, & Mouchiroud, MDominique. 2000. Determinants of substitution rates in mammalian genes : expression pattern affects selection intensity but not mutation rate. *Molecular Biology and Evolution*, **17**(1), 68–70.

Dutko, James A., & Mullins, Mary C. 2011. SnapShot : BMP Signaling in Development. *Cell*, **145**(4), 636–636.e2.

Enard, Wolfgang, Khaitovich, Philipp, Klose, Joachim, Zöllner, Sebastian, Heissig, Florian, Giavalisco, Patrick, Nieselt-Struwe, Kay, Muchmore, Elaine, Varki, Ajit, Ravid, Rivka, Doxiadis, Gaby M., Bontrop, Ronald E., & Pääbo, Svante. 2002. Intra- and Interspecific Variation in Primate Gene Expression Patterns. *Science*, **296**(5566), 340–343.

Evans, Alistair R., Wilson, Gregory P., Fortelius, Mikael, & Jernvall, Jukka. 2006. High-level similarity of dentitions in carnivorans and rodents. *Nature*, **445**(7123), 78–81.

Ewing, Rob M., Kahla, Alia Ben, Poirot, Olivier, Lopez, Fabrice, Audic, Stéphane, & Claverie, Jean-Michel. 1999. Large-Scale Statistical Analyses of Rice ESTs Reveal Correlated Patterns of Gene Expression. *Genome Research*, **9**(10), 950–959.

Fausser, Jean-Luc, Schlepp, Odile, Aberdam, Daniel, Meneguzzi, Guerrino, Ruch, Jean Victor, & Lesot, Herve. 1998. Localization of antigens associated with adherens junctions, desmosomes, and hemidesmosomes during murine molar morphogenesis. *Differentiation*, **63**(1), 1–11.

Fraser, Gareth J, Hulsey, C. Darrin, Bloomquist, Ryan F, Uyesugi, Kristine, Manley, Nancy R, & Streelman, J. Todd. 2009. An Ancient Gene Network Is Co-opted for Teeth on Old and New Jaws. *PLoS Biology*, **7**(2), 233–247.

Fraser, Gareth J., Cerny, Robert, Soukup, Vladimir, Bronner-Fraser, Marianne, & Streelman, J. Todd. 2010. The odontode explosion : The origin of tooth-like structures in vertebrates. *Bioessays*, **32**(9), 808–817.

Freese, Jennifer L., Pino, Darya, & Pleasure, Samuel J. 2010. Wnt signaling in development and disease. *Neurobiology of Disease*, **38**(2), 148–153.

Félix, Marie-Anne. 1999. Evolution of developmental mechanisms in nematodes. *J. Exp. Zool.*, **285**(1), 3–18.

Gaujoux, R., & Seoighe, C. 2013. CellMix : a comprehensive toolbox for gene expression deconvolution. *Bioinformatics*, **29**(17), 2211–2212.

Gaujoux, Renaud. 2013 (June). *An introduction to gene expression deconvolution and theCellMix package. A Comprehensive Framework for Gene Expression Deconvolution.* http ://-web.cbio.uct.ac.za/ renaud/CRAN/web/CellMix/vignettes/Introduction.pdf. Consulted : 16/10/2016.

Gaunt, W.A. 1955. The development of the molar pattern of the mouse (Mus musculus). *Cells Tissues Organs*, **24**(3-4), 249–268.

Gilad, Yoav, & Mizrahi-Man, Orna. 2015. A reanalysis of mouse ENCODE comparative gene expression data. *F1000Research*, may.

Gilbert, Scott F. 2003. The morphogenesis of evolutionary developmental biology. *International Journal of Developmental Biology*, **47**(7-8), 467–477.

Gilbert, Scott F., Opitz, John M., & Raff, Rudolf A. 1996. Resynthesizing Evolutionary and Developmental Biology. *Developmental Biology*, **173**(2), 357–372.

Goltsev, Yury, & Papatsenko, Dmitri. 2009. Time warping of evolutionary distant temporal gene expression data based on noise suppression. *BMC Bioinformatics*, **10**(1), 353.

Goto, Asami, Yamazaki, Naohiro, & Nogawa, Hiroyuki. 2014. Characterization of FGF Family Growth Factors Concerning Branching Morphogenesis of Mouse Lung Epithelium. *Zoological Science*, **31**(5), 267–273.

Goudemand, N., Orchard, M. J., Urdy, S., Bucher, H., & Tafforeau, P. 2011. Synchrotron-aided reconstruction of the conodont feeding apparatus and implications for the mouth of the first vertebrates. *Proceedings of the National Academy of Sciences*, **108**(21), 8720–8724.

Grabherr, Manfred G, Haas, Brian J, Yassour, Moran, Levin, Joshua Z, Thompson, Dawn A, Amit, Ido, Adiconis, Xian, Fan, Lin, Raychowdhury, Raktima, Zeng, Qiandong, Chen, Zehua, Mauceli, Evan, Hacohen, Nir, Gnirke, Andreas, Rhind, Nicholas, di Palma, Federica, Birren, Bruce W, Nusbaum, Chad, Lindblad-Toh, Kerstin, Friedman, Nir, & Regev, Aviv. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**(7), 644–652.

Gritli-Linde, A., Bei, Marianna, Maas, Richard, Zhang, Xiaoyan M., Linde, Anders, & McMahon, Andrew P. 2002. Shh signaling within the dental epithelium is necessary for cell proliferation, growth and polarization. *Development*, **129**(23), 5323–5337.

Grüneberg, Hans. 1965. Genes and genotypes affecting the teeth of the mouse. *Development*, **14**(2), 137–159.

Grün, Dominic, & van Oudenaarden, Alexander. 2015. Design and Analysis of Single-Cell Sequencing Experiments. *Cell*, **163**(4), 799–810.

Haara, O., Harjunmaa, E., Lindfors, P. H., Huh, S.-H., Fliniaux, I., Aberg, T., Jernvall, J., Ornitz, D. M., Mikkola, M. L., & Thesleff, I. 2012. Ectodysplasin regulates activator-inhibitor balance in murine tooth development through Fgf20 signaling. *Development*, **139**(17), 3189–3199.

Haas, Brian J, Papanicolaou, Alexie, Yassour, Moran, Grabherr, Manfred, Blood, Philip D, Bowden, Joshua, Couger, Matthew Brian, Eccles, David, Li, Bo, Lieber, Matthias, MacManes, Matthew D, Ott, Michael, Orvis, Joshua, Pochet, Nathalie, Strozzi, Francesco, Weeks, Nathan, Westerman, Rick, William, Thomas, Dewey, Colin N, Henschel, Robert, LeDuc, Richard D, Friedman, Nir, & Regev, Aviv. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**(8), 1494–1512.

Halder, Georg, Callaerts, Patrick, & Gehring, Walter J. 1995a. Induction of Ectopic Eyes by Targeted Expression of the *eyeless* Gene in Drosophila. *Science*, **267**(Mar.), 1788–1792.

Halder, Georg, Callaerts, Patrick, & Gehring, Walter J. 1995b. New perspectives on eye evolution. *Current Opinion in Genetics & Development*, **5**(5), 602–609.

Hall, Brian K. 2003. Descent with modification : the unity underlying homology and homoplasy as seen through an analysis of development and evolution. *Bio. Rev.*, **78**(3), 409 :433.

Hammond, Stephanie, Kaplarevic, Mihailo, Borth, Nicole, Betenbaugh, Michael J., & Lee, Kelvin H. 2011. Chinese hamster genome database : An online resource for the CHO community at www.CHOgenome.org. *Biotechnology and Bioengineering*, **109**(6), 1353–1356.

Handrigan, G. R., Leung, K. J., & Richman, J. M. 2010. Identification of putative dental epithelial stem cells in a lizard with life-long tooth replacement. *Development*, **137**(21), 3545–3549.

Handrigan, Gregory R., & Richman, Joy M. 2010. A network of Wnt, hedgehog and BMP signaling pathways regulates tooth replacement in snakes. *Developmental Biology*, **348**(1), 130–141.

Harada, Hidemitsu, Toyono, Takashi, Toyoshima, Kuniaki, Yamasaki, Masahiro, Itoh, Nobuyuki, Kato, Shigeaki, Sekine, Keisuke, & Ohuchi, Hideyo. 2002. FGF10 maintains stem cell compartment in developing mouse incisors. *Development*, **129**(6), 1533–1541.

Harjunmaa, Enni, Kallonen, Aki, Voutilainen, Maria, Hämäläinen, Keijo, Mikkola, Marja L., & Jernvall, Jukka. 2012. On the difficulty of increasing dental complexity. *Nature*, **483**(7389), 324–327.

Harjunmaa, Enni, Seidel, Kerstin, Häkkinen, Teemu, Renvoisé, Elodie, Corfe, Ian J., Kallonen, Aki, Zhang, Zhao-Qun, Evans, Alistair R., Mikkola, Marja L., Salazar-Ciudad, Isaac, Klein, Ophir D., & Jernvall, Jukka. 2014. Replaying evolutionary transitions from the dental fossil record. *Nature*, jul.

Haycraft, Courtney J., Banizs, Boglarka, Aydin-Son, Yesim, Zhang, Qihong, Michaud, Edward J., & Yoder, Bradley K. 2005. Gli2 and Gli3 Localize to Cilia and Require the Intraflagellar Transport Protein Polaris for Processing and Function. *PLoS Genetics*, **1**(4), e53.

Headon, Denis J., Emmal, Stephanie A., Ferguson, Betsy M., Tucker, Abigail S., Justice, Monica J., Sharpe, Paul T., Zonana, Jonathan, & Overbeek, Paul A. 2001. Gene defect in ectodermal dysplasia implicates a death domain adapter in development. *Nature*, **414**(6866), 913–916.

Hoekstra, Hopi E., & Coyne, Jerry A. 2007. THE LOCUS OF EVOLUTION : EVO DEVO AND THE GENETICS OF ADAPTATION. *Evolution*, **61**(5), 995–1016.

Hoorebeke, E. Pauwels AND D. Van Loo AND P. Cornillie AND L. Brabant AND L. Van. 2013. An exploratory study of contrast agents for soft tissue visualization by means of high resolution X-ray computed tomography imaging. *Journal of Microscopy*, **250**(1), 21–31.

Hu, Bing, Nadiri, Amal, Kuchler-Bopp, Sabine, Perrin-Schmitt, Fabienne, Peters, Heiko, & Lesot, Hervé. 2006. Tissue Engineering of Tooth Crown, Root, and Periodontium. *Tissue Engineering*, **12**(8), 2069–2075.

Hu, Jimmy Kuang-Hsien, Mushegyan, Vagan, & Klein, Ophir D. 2013. On the cutting edge of organ renewal : Identification, regulation, and evolution of incisor stem cells. *genesis*, **52**(2), 79–92.

Hunter, John P, & Jernvall, Jukka. 1995. The hypocone as a key innovation in mammalian evolution. *Proceedings of the National Academy of Sciences*, **92**(23), 10718–10722.

Huysseune, Ann, Sire, Jean-Yves, & Witten, P. Eckhard. 2009. Evolutionary and developmental origins of the vertebrate dentition. *Journal of Anatomy*, **214**(4), 465–476.

Ikeya, M., Kawada, M., Kiyonari, H., Sasai, N., Nakao, K., Furuta, Y., & Sasai, Y. 2006. Essential pro-Bmp roles of crossveinless 2 in mouse organogenesis. *Development*, **133**(22), 4463–4473.

Ingham, Philip W., Nakano, Yoshiro, & Seger, Claudia. 2011. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nature Reviews Genetics*, **12**(6), 393–406.

Irie, Naoki, & Kuratani, Shigeru. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Communications*, **2**(mar), 248.

Irie, Naoki, & Sehara-Fujisawa, Atsuko. 2007. The vertebrate phylotypic stage and an early bilaterian-related stage in mouse embryogenesis defined by genomic information. *BMC Biology*, **5**(1), 1.

Ishida, Kentaro, Murofushi, Mayumi, Nakao, Kazuhisa, Morita, Ritsuko, Ogawa, Miho, & Tsuji, Takashi. 2011. The regulation of tooth morphogenesis is associated with epithelial cell proliferation and the expression of Sonic hedgehog through epithelial–mesenchymal interactions. *Biochemical and Biophysical Research Communications*, **405**(3), 455–461.

Janvier, Philippe. 2013. Palaeontology : Inside-out turned upside-down. *Nature*, **502**(7472), 457–458.

Järvinen, E., Salazar-Ciudad, I., Birchmeier, W., Taketo, M. M., Jernvall, J., & Thesleff, I. 2006. Continuous tooth generation in mouse is induced by activated epithelial Wnt/$\beta$-catenin signaling. *Proceedings of the National Academy of Sciences*, **103**(49), 18627–18632.

Järvinen, Elina, Vlimki, Kaisa, Pummila, Marja, Thesleff, Irma, & Jernvall, Jukka. 2008. The taming of the shrew milk teeth. *Evolution & Development*, **10**(4), 477–486.

Järvinen, Elina, Tummers, Mark, & Thesleff, Irma. 2009. The role of the dental lamina in mammalian tooth replacement. *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, **312B**(4), 281–291.

Jernvall, J., & Thesleff, I. 2012. Tooth shape formation and tooth renewal : evolving with the same signals. *Development*, **139**(19), 3487–3497.

Jernvall, J., Aberg, T., Kettunen, P., Keranen, S., & Thesleff, I. 1998. The life history of an embryonic signaling center : BMP4 induces p21 and is associated with apoptosis in the mouse tooth enamel knot. *Development*, **125**(2), 161–169.

Jernvall, Jukka, & Thesleff, Irma. 2000. Reiterative signaling and patterning during mammalian tooth morphogenesis. *Mechanisms of Development*, **92**(1), 19–29.

Jernvall, Jukka, Kettunen, Paivi, Karavanova, Irina, Martin, Lawrence B., & Thesleff, Irma. 1994. Evidence for the role of the enamel knot as a control center in mammalian tooth cusp formation : non-dividing cells express growth stimulating Fgf4 gene. *International Journal of Developmental Biology*, **38**(3), 463–469.

Jia, S., Zhou, J., Gao, Y., Baek, J.-A., Martin, J. F., Lan, Y., & Jiang, R. 2012. Roles of Bmp4 during tooth morphogenesis and sequential tooth formation. *Development*, **140**(2), 423–432.

Jia, Shihai, Kwon, Hyuk-Jae Edward, Lan, Yu, Zhou, Jing, Liu, Han, & Jiang, Rulang. 2016. Bmp4-Msx1 signaling and Osr2 control tooth organogenesis through antagonistic regulation of secreted Wnt antagonists. *Developmental Biology*, **420**(1), 110–119.

Johnston, Daniel St. 2015. The Renaissance of Developmental Biology. *PLOS Biology*, **13**(5), e1002149.

Jussila, M., & Thesleff, I. 2012. Signaling Networks Regulating Tooth Organogenesis and Regeneration, and the Specification of Dental Mesenchymal and Epithelial Cell Lineages. *Cold Spring Harbor Perspectives in Biology*, **4**(4), a008425–a008425.

Juuri, E., Jussila, M., Seidel, K., Holmes, S., Wu, P., Richman, J., Heikinheimo, K., Chuong, C.-M., Arnold, K., Hochedlinger, K., Klein, O., Michon, F., & Thesleff, I. 2013. Sox2 marks epithelial competence to generate teeth in mammals and reptiles. *Development*, **140**(7), 1424–1432.

Kalinka, Alex T., Varga, Karolina M., Gerrard, Dave T., Preibisch, Stephan, Corcoran, David L., Jarrells, Julia, Ohler, Uwe, Bergman, Casey M., & Tomancak, Pavel. 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature*, **468**(7325), 811–814.

Kassai, Yoshiaki, Munne, Pauliina, Hotta, Yuhei, Penttilä, Enni, Kavanagh, Kathryn, Ohbayashi, Norihiko, Takada, Shinji, Thesleff, Irma, Jernvall, Jukka, & Itoh, Nobuyuki. 2005. Regulation of Mammalian Tooth Cusp Patterning by Ectodin. *Science*, **309**(5743), 2067–2070.

Kelley, Rusty, Ren, Rongqin, Pi, Xinchun, Wu, Yaxu, Moreno, Isabel, Willis, Monte, Moser, Martin, Ross, Malcolm, Podkowa, Monika, Attisano, Liliana, & Patterson, Cam. 2009. A concentration-dependent endocytic trap and sink mechanism converts Bmper from an activator to an inhibitor of Bmp signaling. *J Cell Biol*, **184**(4), 597–609.

Keränen, SVE, Åberg, Thomas, Kettunen, Päivi, Thesleff, Irma, & Jernvall, Jukka. 1998. Association of developmental regulatory genes with the development of different molar tooth shapes in two species of rodents. *Development Genes and Evolution*, **208**(9), 477–486.

Kettunen, Päivi, & Thesleff, Irma. 1998. Expression and function of FGFs-4, -8, and -9 suggest functional redundancy and repetitive use as epithelial signals during tooth morphogenesis. *Developmental Dynamics*, **211**(3), 256–268.

Kettunen, Päivi, Laurikkala, Johanna, Itäranta, Petri, Vainio, Seppo, Itoh, Nobuyuki, & Thesleff, Irma. 2000. Associations of FGF-3 and FGF-10 with signaling networks regulating tooth morphogenesis. *Developmental Dynamics*, **219**(3), 322–332.

Khaitovich, Philipp, Weiss, Gunter, Lachmann, Michael, Hellmann, Ines, Enard, Wolfgang, Muetzel, Bjoern, Wirkner, Ute, Ansorge, Wilhelm, & Pääbo, Svante. 2004. A Neutral Model of Transcriptome Evolution. *PLoS Biology*, **2**(5), e132.

Khaitovich, Philipp, Hellmann1, Ines, Enard, Wolfgang, Nowick, Katja, Leinweber, Marcus, Franz, Henriette, Weiss, Gunter, Lachmann, Michael, & Pääbo, Svante. 2005. Parallel Patterns of Evolution in the Genomes and Transcriptomes of Humans and Chimpanzees. *Science*, **309**(5742), 1850–1854.

Khaitovich, Philipp, Enard, Wolfgang, Lachmann, Michael, & Pääbo, Svante. 2006. Evolution of primate gene expression. *Nat Rev Genet*, **7**(9), 693–702.

Kim, Daehwan, Pertea, Geo, Trapnell, Cole, Pimentel, Harold, Kelley, Ryan, & Salzberg, Steven L. 2013. TopHat2 : accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, **14**(4), R36.

King, M., & Wilson, A. 1975. Evolution at two levels in humans and chimpanzees. *Science*, **188**(4184), 107–116.

Kiontke, Karin, Barrière, Antoine, Kolotuev, Irina, Podbilewicz, Benjamin, Sommer, Ralf, Fitch, David H.A., & Félix, Marie-Anne. 2007. Trends, Stasis, and Drift in the Evolution of Nematode Vulva Development. *Current Biology*, **17**(22), 1925–1937.

Klingenberg, Christian Peter. 1998. Heterochrony and allometry : the analysis of evolutionary change in ontogeny. *Biological Reviews of the Cambridge Philosophical Society*, **73**(01), 79–123.

Kondo, Takashi, Zákány, József, Innis, Jeffrey W., & Duboule, Denis. 1997. Of fingers, toes and penises. *Nature*, **390**(6655), 29–29.

Kratochwil, K, Dull, M, Farinas, I, Galceran, J, & Grosschedl, R. 1996. Lef1 expression is activated by BMP-4 and regulates inductive tissue interactions in tooth and hair development. *Genes & Development*, **10**(11), 1382–1394.

Krumlauf, Robb. 1992. Evolution of the vertebrateHox homeobox genes. *Bioessays*, **14**(4), 245–252.

Kurimoto, K. 2006. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Research*, **34**(5), e42–e42.

Kusserow, Arne, Pang, Kevin, Sturm, Carsten, Hrouda, Martina, Lentfer, Jan, Schmidt, Heiko A., Technau, Ulrich, von Haeseler, Arndt, Hobmayer, Bert, Martindale, Mark Q., & Holstein, Thomas W. 2005. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*, **433**(7022), 156–160.

Kwon, H. J. E., Park, E. K., Jia, S., Liu, H., Lan, Y., & Jiang, R. 2015. Deletion of Osr2 Partially Rescues Tooth Development in Runx2 Mutant Mice. *Journal of Dental Research*, **94**(8), 1113–1119.

Kyrylkova, Kateryna, Kyryachenko, Sergiy, Biehs, Brian, Klein, Ophir, Kioussi, Chrissa, & Leid, Mark. 2012. BCL11B Regulates Epithelial Proliferation and Asymmetric Development of the Mouse Mandibular Incisor. *PLoS ONE*, **7**(5), e37670.

Land, M F, & Fernald, R D. 1992. The Evolution of Eyes. *Annu. Rev. Neurosci.*, **15**(1), 1–29.

Landin, Maria A, Nygård, Ståle, Shabestari, Maziar G, Babaie, Eshrat, Reseland, Janne E, & Osmundsen, Harald. 2015. Mapping the global mRNA transcriptome during development of the murine first molar. *Frontiers in Genetics*, **6**(feb).

Langmead, Ben, & Salzberg, Steven L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4), 357–359.

Laugel-Haushalter, Virginie, Paschaki, Marie, Thibault-Carpentier, Christelle, Dembelé, Doulaye, Dollé, Pascal, & Bloch-Zupan, Agnès. 2013. Molars and incisors : show your microarray IDs. *BMC Research Notes*, **6**(1), 113.

Laurikkala, Johanna, Mikkola, Marja, Mustonen, Tuija, Åberg, Thomas, Koppinen, Petra, Pispa, Johanna, Nieminen, Pekka, Galceran, Juan, Grosschedl, Rudolf, & Thesleff, Irma. 2001. TNF Signaling via the Ligand–Receptor Pair Ectodysplasin and Edar Controls the Function of Epithelial Signaling Centers and Is Regulated by Wnt and Activin during Tooth Organogenesis. *Developmental Biology*, **229**(2), 443–455.

Lazzari, Vincent. 2006 (December). *Origine et évolution du plan dentaire murin chez les muroida (rodentia, mammalia). Apports de la paléontologie, de la morphologie fonctionnelle et du développemet dans l'étude d'une innovation évolutive.* phdthesis, Université Montpellier II.

Lazzari, Vincent, Charles, Cyril, Tafforeau, Paul, Vianey-Liaud, Monique, Aguilar, Jean-Pierre, Jaeger, Jean-Jacques, Michaux, Jacques, & Viriot, Laurent. 2008. Mosaic Convergence of Rodent Dentitions. *PLoS ONE*, **3**(10), e3607.

Lemos, Bernardo, Meiklejohn, Colin D., Cáceres, Mario, & Hartl, Daniel L. 2005. RATES OF DIVERGENCE IN GENE EXPRESSION PROFILES OF PRIMATES, MICE, AND FLIES : STABILIZING SELECTION AND VARIABILITY AMONG FUNCTIONAL CATEGORIES. *Evolution*, **59**(1), 126–137.

Lesot, H, Vonesch, JL, Peterka, M, Tureckova, J, Peterkova, R, & Ruch, JV. 1996. Mouse molar morphogenesis revisited by three-dimensional reconstruction. II. Spatial distribution of mitoses and apoptosis in cap to bell staged first and second upper molar teeth. *International Journal of Developmental Biology*, **40**(5), 1017–1031.

Lesot, H, Hovorakova, M, Peterka, M, & Peterkova, R. 2014. Three-dimensional analysis of molar development in the mouse from the cap to bell stage. *Australian Dental Journal*, **59**(feb), 81–100.

Levin, Michal, Hashimshony, Tamar, Wagner, Florian, & Yanai, Itai. 2012. Developmental Milestones Punctuate Gene Expression in the Caenorhabditis Embryo. *Developmental Cell*, **22**(5), 1101–1108.

Levin, Michal, Anavy, Leon, Cole, Alison G., Winter, Eitan, Mostov, Natalia, Khair, Sally, Senderovich, Naftalie, Kovalev, Ekaterina, Silver, David H., Feder, Martin, Fernandez-Valverde, Selene L., Nakanishi, Nagayasu, Simmons, David, Simakov, Oleg, Larsson, Tomas, Liu, Shang-Yun, Jerafi-Vider, Ayelet, Yaniv, Karina, Ryan, Joseph F., Martindale, Mark Q., Rink, Jochen C., Arendt, Detlev, Degnan, Sandie M., Degnan, Bernard M., Hashimshony, Tamar, & Yanai, Itai. 2016. The mid-developmental transition and the evolution of animal body plans. *Nature*, **531**(7596), 637–641.

Lewis, E. B. 1978. A gene complex controlling segmentation in Drosophila. *Nature*, **276**(5688), 565–570.

Lewis, RA, Wakimoto, BT, Denell, RE, & Kaufman, TC. 1980. Genetic analysis of the Antennapedia gene complex (Ant-C) and adjacent chromosomal regions of Drosophila melanogaster. II. Polytene Chromosome segments 84A–84B1, 2. *Genetics*, **95**(2), 383–397.

Li, Bo, & Dewey, Colin N. 2011. RSEM : accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**(1), 323.

Li, Jingyuan, Feng, Jifan, Liu, Yang, Ho, Thach-Vu, Grimes, Weston, Ho, Hoang Anh, Park, Shery, Wang, Songlin, & Chai, Yang. 2015. BMP-SHH Signaling Network Controls Epithelial Stem Cell Fate via Regulation of Its Niche in the Developing Tooth. *Developmental Cell*, **33**(2), 125–135.

Liao, Ben-Yang, & Zhang, Jianzhi. 2005. Evolutionary Conservation of Expression Profiles Between Human and Mouse Orthologous Genes. *Molecular Biology and Evolution*, **23**(3), 530–540.

Lin, Congxing, Yin, Yan, Bell, Sheila M., Veith, G. Michael, Chen, Hong, Huh, Sung-Ho, Ornitz, David M., & Ma, Liang. 2013. Delineating a Conserved Genetic Cassette Promoting Outgrowth of Body Appendages. *PLoS Genetics*, **9**(1), e1003231.

Lin, Shin, Lin, Yiing, Nery, Joseph R., Urich, Mark A., Breschi, Alessandra, Davis, Carrie A., Dobin, Alexander, Zaleski, Christopher, Beer, Michael A., Chapman, William C., Gingeras, Thomas R., Ecker, Joseph R., & Snyder, Michael P. 2014. Comparison of the transcriptional landscapes between human and mouse tissues. *Proceedings of the National Academy of Sciences*, **111**(48), 17224–17229.

Liu, C., Gu, S., Sun, C., Ye, W., Song, Z., Zhang, Y., & Chen, Y. 2013. FGF signaling sustains the odontogenic fate of dental mesenchyme by suppressing -catenin signaling. *Development*, **140**(21), 4375–4385.

Liu, Fei, Chu, Emily Y., Watt, Brenda, Zhang, Yuhang, Gallant, Natalie M., Andl, Thomas, Yang, Steven H., Lu, Min-Min, Piccolo, Stefano, Schmidt-Ullrich, Ruth, Taketo, Makoto M., Morrisey, Edward E., Atit, Radhika, Dlugosz, Andrzej A., & Millar, Sarah E. 2008. Wnt/$\beta$-catenin signaling directs multiple stages of tooth morphogenesis. *Developmental Biology*, **313**(1), 210–224.

Logan, Catriona Y., & Nusse, Roel. 2004. the wnt signaling pathway in development and desease. *Annual Review of Cell and Developmental Biology*, **20**(1), 781–810.

Love, Michael I, Huber, Wolfgang, & Anders, Simon. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**(12).

Luo, Zhe-Xi. 2007. Transformation and diversification in early mammal evolution. *Nature*, **450**(7172), 1011–1019.

Lynch, Michael, & Hill, William G. 1986. Phenotypic Evolution by Neutral Mutation. *Evolution*, **40**(5), 915.

Maaten, Laurens van der, & Hinton, Geoffrey. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, **9**(November), 2579–2605.

Makarenkova, H. P., Hoffman, M. P., Beenken, A., Eliseenkova, A. V., Meech, R., Tsau, C., Patel, V. N., Lang, R. A., & Mohammadi, M. 2009. Differential Interactions of FGFs with Heparan Sulfate Control Gradient Formation and Branching Morphogenesis. *Science Signaling*, **2**(88), ra55–ra55.

Mandler, Markus, & Neubüser, Annette. 2001. FGF Signaling Is Necessary for the Specification of the Odontogenic Mesenchyme. *Developmental Biology*, **240**(2), 548–559.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. 2008. RNA-seq : An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18**(9), 1509–1517.

Matalova, E., Antonarakis, G.S., Sharpe, P.T., & Tucker, A.S. 2005. Cell lineage of primary and secondary enamel knots. *Dev. Dyn.*, **233**(3), 754–759.

Matalova, Eva, Sharpe, Paul T., Lakhani, Saquib A., Roth, Kevin A., Flavell, Richard A., Setkova, Jana, Misek, Ivan, & Tucker, Abigail S. 2006. Molar tooth development in caspase-3 deficient mice. *The International Journal of Developmental Biology*, **50**(Next).

McCollum, Melanie, & Sharpe, Paul T. 2001. Evolution and development of teeth. *Journal of Anatomy*, **199**(1-2), 153–159.

McGinnis, William, Garber, Richard L., Wirz, Johannes, Kuroiwa, Atsushi, & Gehring, Walter J. 1984. A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. *Cell*, **37**(2), 403–408.

Metscher, Brian D. 2009. MicroCT for comparative morphology : simple staining methods allow high-contrast 3D imaging of diverse non-mineralized animal tissues. *BMC Physiology*, **9**(1), 11.

Miletich, Isabelle, Buchner, Georg, & Sharpe, Paul T. 2005. Barx1 and evolutionary changes in feeding. *Journal of Anatomy*, **207**(5), 619–622.

Mina, M, & Kollar, EJ. 1987. he induction of odontogenesis in non-dental mesenchyme combined with early murine mandibular arch epithelium. *Archives of Oral Biology*, **32**(2), 123–127.

Minoux, M., & Rijli, F. M. 2010. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development*, **137**(16), 2605–2621.

Morgan, T. H. 1932. The rise of genetics. *Science*, **76**(1969), 261–267.

Moser, M., Binder, O., Wu, Y., Aitsebaomo, J., Ren, R., Bode, C., Bautch, V. L., Conlon, F. L., & Patterson, C. 2003. BMPER, a Novel Endothelial Cell Precursor-Derived Protein, Antagonizes Bone Morphogenetic Protein Signaling and Endothelial Cell Differentiation. *Molecular and Cellular Biology*, **23**(16), 5664–5679.

Munne, P. M., Tummers, M., Jarvinen, E., Thesleff, I., & Jernvall, J. 2009. Tinkering with the inductive mesenchyme : Sostdc1 uncovers the role of dental mesenchyme in limiting tooth induction. *Development*, **136**(3), 393–402.

Murdock, Duncan J. E., Dong, Xi-Ping, Repetski, John E., Marone, Federica, Stampanoni, Marco, & Donoghue, Philip C. J. 2013. The origin of conodonts and of vertebrate mineralized skeletons. *Nature*, **502**(7472), 546–549.

Müller, G. B. 2007. Evo-devo : extending the evolutionary sythesis. *Nature Reviews Genetics*, **8**(12), 943 :949.

Nadiri, A., Kuchler–Bopp, S., Perrin–Schmitt, F., & Lesot, H. 2006. Expression patterns of BMPRs in the developing mouse molar. *Cell and Tissue Research*, **324**(1), 33–40.

Nakatomi, M., Hovorakova, M., Gritli-Linde, A., Blair, H. J., MacArthur, K., Peterka, M., Lesot, H., Peterkova, R., Ruiz-Perez, V. L., Goodship, J. A., & Peters, H. 2013. Evc Regulates a Symmetrical Response to Shh Signaling in Molar Development. *Journal of Dental Research*, **92**(3), 222–228.

Necsulea, Anamaria, & Kaessmann, Henrik. 2014. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat Rev Genet*, **15**(11), 734–748.

Nichols, S. A., Dirks, W., Pearse, J. S., & King, N. 2006. Early evolution of animal cell signaling and adhesion genes. *Proceedings of the National Academy of Sciences*, **103**(33), 12451–12456.

Niswander, L, & Martin, GAIL R. 1992. Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development*, **114**(3), 755–768.

Obara, Nobuko, & Lesot, Hervé. 2007. Asymmetrical growth, differential cell proliferation, and dynamic cell rearrangement underlie epithelial morphogenesis in mouse molar development. *Cell and Tissue Research*, **330**(3), 461–473.

Ogawa, T., Kapadia, H., Feng, J. Q., Raghow, R., Peters, H., & D'Souza, R. N. 2006. Functional Consequences of Interactions between Pax9 and Msx1 Genes in Normal and Abnormal Tooth Development. *Journal of Biological Chemistry*, **281**(27), 18363–18369.

Ohazama, A., Haycraft, C. J., Seppala, M., Blackburn, J., Ghafoor, S., Cobourne, M., Martinelli, D. C., Fan, C.-M., Peterkova, R., Lesot, H., Yoder, B. K., & Sharpe, P. T. 2009. Primary cilia regulate Shh activity in the control of molar tooth number. *Development*, **136**(6), 897–903.

Ornitz, David M., & Itoh, Nobuyuki. 2015. The Fibroblast Growth Factor signaling pathway. *Wiley Interdisciplinary Reviews : Developmental Biology*, **4**(3), 215–266.

Osborn, Henry F. 1887. On the structure and classification of the Mesozoic Mammalia. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **nnd**, 282–292.

Osborn, Henry Fairfield. 1888. The evolution of mammalian molars to and from the tritubercular type. *The American Naturalist*, **22**(264), 1067–1079.

O'Connell, Daniel J, Ho, Joshua WK, Mammoto, Tadanori, Turbe-Doan, Annick, O'Connell, Joyce T, Haseley, Psalm S, Koo, Samuel, Kamiya, Nobuhiro, Ingber, Donald E, Park, Peter J, *et al.* . 2012. A Wnt-Bmp Feedback Circuit Controls Intertissue Signaling Dynamics in Tooth Organogenesis. *Science Signaling*, **5**(206), ra4–ra4.

Paaby, A. B., & Rockman, M. V. 2014. Cryptic genetic variation : evolution's hidden substrate. *Nature Reviews Genetics*, **15**(4), 247 :258.

Panafieu, Jean-Baptiste de, & Gries, Patrick. 2007. *Evolution*. Editions Xavier Barral, MNHN, ISBN : 2915173745.

Pantalacci, Sophie, & Sémon, Marie. 2014. Transcriptomics of developing embryos and organs : A raising tool for evo-devo. *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, **324**(4), 363–371.

Pantalacci, Sophie, Guéguen, Laurent, Petit, Coraline, Lambert, Anne, Peterkova, Renata, & Sémon, Marie. 2016. Transcriptomic signatures shaped by cell proportions shed light on comparative developmental biology. *Genome Biology*, Dec.

Papatsenko, Dmitri, Levine, Michael, & Goltsev, Yury. 2011. Clusters of Temporal Discordances Reveal Distinct Embryonic Patterning Mechanisms in Drosophila and Anopheles. *PLoS Biology*, **9**(1), e1000584.

Pavlicev, Mihaela, & Wagner, Günter P. 2012. A model of developmental evolution : selection, pleiotropy and compensation. *Trends in Ecology & Evolution*, **27**(6), 316–322.

Perié, Leïla, & Duffy, Ken R. 2016. Retracing the in vivo haematopoietic tree using single-cell methods. *FEBS Lett*, jul.

Peterka, M., Lesot, H., & Peterková, R. 2002. Body Weight in Mouse Embryos Specifies Staging of Tooth Development. *Connective Tissue Research*, **43**(2-3), 186–190.

Peterkova, R, Peterka, M, Viriot, L, & Lesot, H. 2002. Dentition development and budding morphogenesis. *Journal of craniofacial genetics and developmental biology*, **43**(2-3), 120–128.

Peters, H., Neubuser, A., Kratochwil, K., & Balling, R. 1998. Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes & Development*, **12**(17), 2735–2747.

Piasecka, Barbara, Lichocki, Paweł, Moretti, Sébastien, Bergmann, Sven, & Robinson-Rechavi, Marc. 2013. The Hourglass and the Early Conservation Models—Co-Existing Patterns of Developmental Constraints in Vertebrates. *PLoS Genetics*, **9**(4), e1003476.

Picelli, Simone, Faridani, Omid R, Björklund, Åsa K, Winberg, Gösta, Sagasser, Sven, & Sandberg, Rickard. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, **9**(1), 171–181.

Pimentel, Harold J, Bray, Nicolas, Puente, Suzette, Melsted, Páll, & Pachter, Lior. 2016. Differential analysis of RNA-Seq incorporating quantification uncertainty. *bioRXiv*, jun.

Pispa, Johanna, & Thesleff, Irma. 2003. Mechanisms of ectodermal organogenesis. *Developmental Biology*, **262**(2), 195–205.

Pispa, Johanna, Jung, Han-Sung, Jernvall, Jukka, Kettunen, Päivi, Mustonen, Tuija, Tabata, Makoto J., Kere, Juha, & Thesleff, Irma. 1999. Cusp Patterning Defect in Tabby Mouse Teeth and Its Partial Rescue by FGF. *Developmental Biology*, **216**(2), 521–534.

Porntaveetus, Thantrira, Otsuka-Tanaka, Yoko, Basson, M. Albert, Moon, Anne M., Sharpe, Paul T., & Ohazama, Atsushi. 2011. Expression of fibroblast growth factors (Fgfs) in murine tooth development. *Journal of Anatomy*, **218**(5), 534–543.

Prud'homme, Benjamin, & Gompel, Nicolas. 2010. Evolutionary biology : Genomic hourglass. *Nature*, **468**(7325), 768–769.

Qian, Mengjia, Wang, Diane C, Chen, Hao, & Cheng, Yunfeng. 2016. Detection of single cell heterogeneity in cancer. *Seminars in Cell & Developmental Biology*, sep.

Qiao, Jizeng, Bush, Kevin T., Steer, Dylan L., Stuart, Robert O., Sakurai, Hiroyuki, Wachsman, William, & Nigam, Sanjay K. 2001. Multiple fibroblast growth factors support growth of the ureteric bud but have different effects on branching morphogenesis. *Mechanisms of Development*, **109**(2), 123–135.

Qiu, Mengsheng, Bulfone, Alessandro, Ghattas, Ingrid, Meneses, Juanito J., Christensen, Lars, Sharpe, Paul T., Presley, Robert, Pedersen, Roger A., & Rubenstein, John L.R. 1997. Role of the Dlx Homeobox Genes in Proximodistal Patterning of the Branchial Arches : Mutations of Dlx-1, Dlx-2, and Dlx-1 and -2 Alter Morphogenesis of Proximal Skeletal and Soft Tissue Structures Derived from the First and Second Arches. *Developmental Biology*, **185**(2), 165–184.

Quint, Marcel, Drost, Hajk-Georg, Gabel, Alexander, Ullrich, Kristian Karsten, Bönn, Markus, & Grosse, Ivo. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature*, **490**(7418), 98–101.

R Core Team. 2016. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raff, Rudolf A. 2000. Evo-devo : the evolution of a new discipline. *Nat Rev Genet*, **1**(1), 74–79.

Reisz, Robert R. 2006. Origin of dental occlusion in tetrapods : signal for terrestrial vertebrate evolution? *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, **306B**(3), 261–277.

Rentzsch, F., Zhang, Jinli, Kramer, Carina, Sebald, Walter, & Hammerschmidt, Matthias. 2006. Crossveinless 2 is an essential positive feedback regulator of Bmp signaling during zebrafish gastrulation. *Development*, **133**(5), 801–811.

Richardson, M. K., Hanken, James, Gooneratne, Mayoni L., Pieau, Claude, Raynaud, Albert, Selwood, Lynne, & Wright, Glenda M. 1997. There is no highly conserved embryonic stage in the vertebrates : implications for current theories of evolution and development. *Anatomy and Embryology*, **196**(2), 91–106.

Richardson, Michael K., & Keuck, Gerhard. 2002. Haeckel's ABC of evolution and development. *Biol. Rev.*, **77**(4), 495–528.

Rider, Christopher C., & Mulloy, Barbara. 2010. Bone morphogenetic protein and growth differentiation factor cytokine families and their protein antagonists. *Biochemical Journal*, **429**(1), 1–12.

Rifkin, Scott A., Houle, David, Kim, Junhyong, & White, Kevin P. 2005. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature*, **438**(7065), 220–223.

Ringnér, Markus. 2008. What is principal component analysis? *Nature Biotechnology*, **26**(3), 303–304.

Robertis, E.M. De. 2008. Evo-Devo : Variations on Ancestral Themes. *Cell*, **132**(2), 185–195.

Rodrigues, H. Gomes, Marangoni, P., Sumbera, R., Tafforeau, P., Wendelen, W., & Viriot, L. 2011. Continuous dental replacement in a hyper-chisel tooth digging rodent. *Proceedings of the National Academy of Sciences*, **108**(42), 17355–17359.

Rohatgi, R., Milenkovic, L., & Scott, M. P. 2007. Patched1 Regulates Hedgehog Signaling at the Primary Cilium. *Science*, **317**(5836), 372–376.

Romero, Irene Gallego, Ruvinsky, Ilya, & Gilad, Yoav. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet*, **13**(7), 505–516.

Rossant, Cyrille. 2015. *An illustrated introduction to the t-SNE algorithm. https://www.oreilly.com/learning/an-illustrated-introduction-to-the-t-sne-algorithm.*

Roux, Julien, & Robinson-Rechavi, Marc. 2008. Developmental Constraints on Vertebrate Genome Evolution. *PLoS Genetics*, **4**(12), e1000311.

Roux, Julien, Rosikiewicz, Marta, & Robinson-Rechavi, Marc. 2015. What to compare and how : Comparative transcriptomics for Evo-Devo. *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, **324**(4), 372–382.

Ruvinsky, I., & Gibson-Brown, J. J. 2000. Genetic and developmental bases of serial homology in vertebrate limb evolution. *Development*, **127**(24), 5233–5244.

Sadier, Alexa, Viriot, Laurent, Pantalacci, Sophie, & Laudet, Vincent. 2014. The ectodysplasin pathway : from diseases to adaptations. *Trends in Genetics*, **30**(1), 24–31.

Salazar-Ciudad, I., & Jernvall, J. 2002. A gene network model accounting for development and evolution of mammalian teeth. *Proceedings of the National Academy of Sciences*, **99**(12), 8116–8120.

Salazar-Ciudad, Isaac. 2008. Tooth Morphogenesis in vivo, in vitro, and in silico. *Current Topics in Developmental Biology*, 341–371.

Salazar-Ciudad, Isaac. 2012. Tooth patterning and evolution. *Current Opinion in Genetics & Development*, **22**(6), 585–592.

Salazar-Ciudad, Isaac, & Jernvall, Jukka. 2010. A computational model of teeth and the developmental origins of morphological variation. *Nature*, **464**(7288), 583–586.

Sanchez-Freire, Veronica, Ebert, Antje D, Kalisky, Tomer, Quake, Stephen R, & Wu, Joseph C. 2012. Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. *Nature Protocols*, **7**(5), 829–838.

Sansom, I., Smith, M., Armstrong, H., & Smith, M. 1992. Presence of the earliest vertebrate hard tissue in conodonts. *Science*, **256**(5061), 1308–1311.

Sasai, Noriaki, & Briscoe, James. 2012. Primary cilia and graded Sonic Hedgehog signaling. *Wiley Interdisciplinary Reviews : Developmental Biology*, **1**(5), 753–772.

Schmitt, RÉGINE, Lesot, HERVÉ, Vonesch, JL, & Ruch, JEAN-VICTOR. 1999. Mouse odontogenesis in vitro : the cap-stage mesenchyme controls individual molar crown morphogenesis. *International Journal of Developmental Biology*, **43**(3), 255–260.

Schultz, Julia A., & Martin, Thomas. 2014. Function of pretribosphenic and tribosphenic mammalian molars inferred from 3D animation. *Naturwissenschaften*, **101**(10), 771–781.

Scialdone, Antonio, Tanaka, Yosuke, Jawaid, Wajid, Moignard, Victoria, Wilson, Nicola K., Macaulay, Iain C., Marioni, John C., & Göttgens, Berthold. 2016. Resolving early mesoderm diversification through single-cell expression profiling. *Nature*, **535**(7611), 289–293.

Scott, Matthew P., & Weiner, Amy J. 1984. Structural relationships among genes that control development : Sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proc Natl Acad Sci USA*, **81**(July), 4115–4119.

Sebbagh, Michael, & Borg, Jean-Paul. 2014. Insight into planar cell polarity. *Experimental Cell Research*, **328**(2), 284–295.

Semenov, Mikhail V., Habas, Raymond, MacDonald, Bryan T., & He, Xi. 2007. SnapShot : Noncanonical Wnt Signaling Pathways. *Cell*, **131**(7), 1378.e1–1378.e2.

Serpe, Mihaela, Umulis, David, Ralston, Amy, Chen, Jun, Olson, David J., Avanesov, Andrei, Othmer, Hans, O'Connor, Michael B., & Blair, Seth S. 2008. The BMP-Binding Protein Crossveinless 2 Is a Short-Range, Concentration-Dependent, Biphasic Modulator of BMP Signaling in Drosophila. *Developmental Cell*, **14**(6), 940–953.

Setkova, J., Matalova, E., Sharpe, P.T., Misek, I., & Tucker, A.S. 2007. Primary enamel knot cell death in Apaf-1 and caspase-9 deficient mice. *Archives of Oral Biology*, **52**(1), 15–19.

Shubin, Neil, Tabin, Cliff, & Carroll, Sean. 2009. Deep homology and the origins of evolutionary novelty. *Nature*, **457**(7231), 818–823.

Sidow, Arend. 1996. Gen(om)e duplications in the evevolution early vertebrates. *Current Opinion in Genetics & Development*, **6**(6), 715–722.

Slack, J. M. W. 2002. Conrad Hal Waddington : the last Renaissance biologist? *Nature Reviews Genetics*, **3**(11), 889–895.

Smith, Moya Meredith. 2003. Vertebrate dentitions at the origin of jaws : when and how pattern evolved. *Evol Dev*, **5**(4), 394–413.

Smith, Moya Meredith, & Johanson, Zerina. 2003. Separate Evolutionary Origins of Teeth from Evidence in Fossil Jawed Vertebrates. *Science*, **299**(5610), 1235–1236.

Steppan, Scott, Adkins, Ronald, & Anderson, Joel. 2004. Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes. *Systematic Biology*, **53**(4), 533–553.

Stock, D. W. 2001. The genetic basis of modularity in the development and evolution of the vertebrate dentition. *Philosophical Transactions of the Royal Society B : Biological Sciences*, **356**(1414), 1633–1653.

Stock, David W., Weiss, Kenneth M., & Zhao, Zhiyong. 1997. Patterning of the mammalian dentition in development and evolution. *Bioessays*, **19**(6), 481–490.

Su, A. I., Cooke, M. P., Ching, K. A., Hakak, Y., Walker, J. R., Wiltshire, T., Orth, A. P., Vega, R. G., Sapinoso, L. M., Moqrich, A., Patapoutian, A., Hampton, G. M., Schultz, P. G., & Hogenesch, J. B. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the National Academy of Sciences*, **99**(7), 4465–4470.

Tchitchek, Nicolas, Safronetz, David, Rasmussen, Angela L., Martens, Craig, Virtaneva, Kimmo, Porcella, Stephen F., Feldmann, Heinz, Ebihara, Hideki, & Katze, Michael G. 2014. Sequencing, Annotation and Analysis of the Syrian Hamster (Mesocricetus auratus) Transcriptome. *PLoS ONE*, **9**(11), e112617.

Thesleff, I., Keranen, S., & Jernvall, J. 2001. Enamel Knots as Signaling Centers Linking Tooth Morphogenesis and Odontoblast Differentiation. *Advances in Dental Research*, **15**(1), 14–18.

Thomas, Bethan L, Tucker, Abigail S, Qui, M, Ferguson, Christine A, Hardcastle, Zoë, Rubenstein, JL, & Sharpe, Paul T. 1997. Role of Dlx-1 and Dlx-2 genes in patterning of the murine dentition. *Development*, **124**(23), 4811–3818.

Thompson, Thomas B., Lerch, Thomas F., Cook, Robert W., Woodruff, Teresa K., & Jardetzky, Theodore S. 2005. The Structure of the Follistatin :Activin Complex Reveals Antagonism of Both Type I and Type II Receptor Binding. *Developmental Cell*, **9**(4), 535–543.

Tims, HW Marett. 1903. The evolution of the teeth in the Mammalia. *Journal of anatomy and physiology*, **37**(2), 131–149.

Togo, Yumiko, Takahashi, Katsu, Saito, Kazuyuki, Kiso, Honoka, Tsukamoto, Hiroko, Huang, Boyen, Yanagita, Motoko, Sugai, Manabu, Harada, Hidemitsu, Komori, Toshihisa, Shimizu, Akira, MacDougall, Mary, & Bessho, Kazuhisa. 2016. Antagonistic Functions of USAG-1 and RUNX2 during Tooth Development. *PLOS ONE*, **11**(8), e0161067.

Trapnell, C., Pachter, L., & Salzberg, S. L. 2009. TopHat : discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**(9), 1105–1111.

True, John R., & Haag, Eric S. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev*, **3**(2), 109–119.

Tschopp, Patrick, Sherratt, Emma, Sanger, Thomas J., Groner, Anna C., Aspiras, Ariel C., Hu, Jimmy K., Pourquié, Olivier, Gros, Jérôme, & Tabin, Clifford J. 2014. A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature*, **516**(7531), 391–394.

Tucker, A. S. 1998. Transformation of Tooth Type Induced by Inhibition of BMP Signaling. *Science*, **282**(5391), 1136–1138.

Tucker, Abigail, & Sharpe, Paul. 2004. The cutting-edge of mammalian development : how the embryo makes teeth. *Nat Rev Genet*, **5**(7), 499–508.

Tucker, Abigail S., & Fraser, Gareth J. 2014. Evolution and developmental diversity of tooth regeneration. *Seminars in Cell & Developmental Biology*, **25-26**(jan), 71–80.

Tucker, Abigail S, Headon, Denis J, Schneider, Pascal, Ferguson, Betsy M, Overbeek, Paul, Tschopp, Jurg, & Sharpe, Paul T. 2000. Edar/Eda interactions regulate enamel knot formation in tooth morphogenesis. *Development*, **127**(21), 4691–4700.

Tucker, A.S, Headon, D.J, Courtney, J.-M, Overbeek, P, & Sharpe, P.T. 2004. The activation level of the TNF family receptor, Edar, determines cusp number and tooth number during tooth development. *Developmental Biology*, **268**(1), 185–194.

Tummers, Mark, & Thesleff, Irma. 2009. The importance of signal pathway modulation in all aspects of tooth development. *Journal of Experimental Zoology Part B : Molecular and Developmental Evolution*, **312B**(4), 309–319.

Turner, Nicholas, & Grose, Richard. 2010. Fibroblast growth factor signalling : from development to cancer. *Nature Reviews Cancer*, **10**(2), 116–129.

Vaahtokari, Anne, Aberg, T, & Thesleff, Irma. 1996. Apoptosis in the developing tooth : association with an embryonic signaling center and suppression by EGF and FGF-4. *Development*, **122**(1), 121–129.

Vainio, Seppo, Karavanova, Irina, Jowett, Adrian, & Thesleff, Irma. 1993. Identification of BMP-4 as a signal mediating secondary induction between epithelial and mesenchymal tissues during early tooth development. *Cell*, **75**(1), 45–58.

Viriot, L, Peterkova, Renata, Vonesch, Jl, Peterka, Miroslav, Ruch, Jv, & Lesot, H. 1997. Mouse molar morphogenesis revisited by three-dimensional reconstruction. III. Spatial distribution of mitoses and apoptoses up to bell-staged first lower molar teeth. *International Journal of Developmental Biology*, **41**(5), 679–690.

Vopalensky, P., Pergner, J., Liegertova, M., Benito-Gutierrez, E., Arendt, D., & Kozmik, Z. 2012. Molecular analysis of the amphioxus frontal eye unravels the evolutionary origin of the retina and pigment cells of the vertebrate eye. *Proceedings of the National Academy of Sciences*, **109**(38), 15383–15388.

Wagner, Günter P. 1996. Homologues, Natural Kinds and the Evolution of Modularity. *American Zoologist*, **36**(1), 36–43.

Wagner, Günter P. 2007. The developmental genetics of homology. *Nat Rev Genet*, **8**(6), 473–479.

Wagner, Günter P., Pavlicev, Mihaela, & Cheverud, James M. 2007. The road to modularity. *Nat Rev Genet*, **8**(12), 921–931.

Wallace, Arthur. 2002. The emerging conceptual framework of evolutionary developmental biology. *Nature*, **415**(6873), 757–764.

Wang, X-P, Åberg, T, James, MJ, Levanon, D, Groner, Y, & Thesleff, I. 2005. Runx2 (Cbfa1) inhibits Shh signaling in the lower but not upper molars of mouse embryos and prevents the budding of putative successional teeth. *Journal of dental research*, **84**(2), 138–143.

Wang, Xiu-Ping, Suomalainen, Marika, Jorgez, Carolina J., Matzuk, Martin M., Werner, Sabine, & Thesleff, Irma. 2004a. Follistatin Regulates Enamel Patterning in Mouse Incisors by Asymmetrically Inhibiting BMP Signaling and Ameloblast Differentiation. *Developmental Cell*, **7**(5), 719–730.

Wang, Xiu-Ping, Suomalainen, Marika, Jorgez, Carolina J., Matzuk, Martin M., Wankell, Miriam, Werner, Sabine, & Thesleff, Irma. 2004b. Modulation of activin/bone morphogenetic protein signaling by follistatin is required for the morphogenesis of mouse molar teeth. *Developmental Dynamics*, **231**(1), 98–108.

Wang, Zhe, Young, Rebecca L., Xue, Huiling, & Wagner, Günter P. 2011. Transcriptomic analysis of avian digits reveals conserved and derived digit identities in birds. *Nature*, **477**(7366), 583–586.

Wang, Zhong, Gerstein, Mark, & Snyder, Michael. 2009. RNA-Seq : a revolutionary tool for transcriptomics. *Nat Rev Genet*, **10**(1), 57–63.

White, A. K., VanInsberghe, M., Petriv, O. I., Hamidi, M., Sikorski, D., Marra, M. A., Piret, J., Aparicio, S., & Hansen, C. L. 2011. High-throughput microfluidic single-cell RT-qPCR. *Proceedings of the National Academy of Sciences*, **108**(34), 13999–14004.

Wickham, Hadley. 2009. *ggplot2 : elegant graphics for data analysis*. Springer Science & Business Media.

Wotton, Karl R, Jiménez-Guri, Eva, Crombach, Anton, Janssens, Hilde, Alcaine-Colet, Anna, Lemke, Steffen, Schmidt-Ott, Urs, & Jaeger, Johannes. 2015. Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly Megaselia abdita. *eLife*, **4**(jan).

Wray, Gregory A. 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, **8**(3), 206–216.

Yamada, G., Suzuki, K., Haraguchi, R., Miyagawa, S., Satoh, Y., Kamimura, M., Nakagata, N., Kataoka, H., Kuroiwa, A., & Chen, Y. 2006. Molecular genetic cascades for external genitalia formation : An emerging organogenesis program. *Dev. Dyn.*, **235**(7), 1738–1752.

Yanai, I., & Hunter, C. P. 2009. Comparison of diverse developmental transcriptomes reveals that coexpression of gene neighbors is not evolutionarily conserved. *Genome Research*, **19**(12), 2214–2220.

Yanai, Itai, Graur, Dan, & Ophir, Ron. 2004. Incongruent Expression Profiles between Human and Mouse Orthologous Genes Suggest Widespread Neutral Evolution of Transcription Control. *OMICS : A Journal of Integrative Biology*, **8**(1), 15–24.

Yoshiba, Kunihiko, Yoshiba, Nagako, Aberdam, Daniel, Meneguzzi, Guerrino, Perrin-Schmitt, Fabienne, Stoetzel, Corinne, Ruch, Jean Victor, & Lesot, Hervé. 1998. Expression and localization of laminin-5 subunits during mouse tooth development. *Developmental Dynamics*, **211**(2), 164–176.

Young, Nathan M., & HallgrÍmsson, Benedikt. 2005. Serial homology and the evolution mammalian limb covariation structure. *Evol*, **59**(12), 2691.

Young, Rebecca L., Bever, Gabe S., Wang, Zhe, & Wagner, Günter P. 2011. Identity of the avian wing digits : Problems resolved and unsolved. *Dev. Dyn.*, **240**(5), 1042–1053.

Zakin, Lise, Metzinger, Carrie A., Chang, Ellen Y., Coffinier, Catherine, & Robertis, E.M. De. 2008. Development of the vertebral morphogenetic field in the mouse : Interactions between Crossveinless-2 and Twisted Gastrulation. *Developmental Biology*, **323**(1), 6–18.

Zhang, Z., Lan, Y., Chai, Y., & Jiang, R. 2009. Antagonistic Actions of Msx1 and Osr2 Pattern Mammalian Teeth into a Single Row. *Science*, **323**(5918), 1232–1234.

Zhao, Xiang, Zhang, Zunyi, Song, Yiqiang, Zhang, Xiaoyun, Zhang, Yanding, Hu, Yuping, Fromm, Sigurd H, & Chen, YiPing. 2000. Transgenically ectopic expression of Bmp4 to the Msx1 mutant dental mesenchyme restores downstream gene expression but represses Shh and Bmp2 in the enamel knot of wild type tooth germ. *Mechanisms of Development*, **99**(1-2), 29–38.

Zhong, Yi, Wan, Ying-Wooi, Pang, Kaifang, Chow, Lionel ML, & Liu, Zhandong. 2013. Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC Bioinformatics*, **14**(1), 89.