

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : *Génétique, génomique et bioinformatique*

Par

Estelle LECLUZE

Analyse par RNA-seq de la différenciation des gonades fœtales humaines et de son altération par des perturbateurs endocriniens.

Thèse présentée et soutenue à Rennes, le 18 octobre 2018
Unité de recherche : Inserm U1085 IRSET, 9 avenue du Professeur Léon Bernard 35000 Rennes

Rapporteurs avant soutenance :

Marie-Christine CHABOISSIER
Yann GUIGUEN

Directeur de recherche, Inserm U1091 iBV
Directeur de recherche – INRA UR1037

Composition du Jury :

Président : Marie-Christine CHABOISSIER

Directeur de recherche, Inserm U1091 iBV

Examineurs : Marie-Christine CHABOISSIER
Yann GUIGUEN

Directeur de recherche, Inserm U1091 iBV
Directeur de recherche – INRA UR1037

Dir. de thèse : Bernard JÉGOU

Directeur de Recherche de classe exceptionnelle,
Inserm U1085 IRSET

Co-dir. de thèse : Frédéric CHALMEL

Chargé de recherche – Inserm U1085 IRSET

Remerciements

En premier lieu, je tiens bien évidemment à remercier chaleureusement le Professeur **Bernard Jégou**, directeur de l'Institut de Recherche en Santé, Environnement et Travail mais également directeur de la Recherche et de l'Innovation à l'École des Hautes Études en Santé Publique, de m'avoir accueillie au sein de son institut. Merci au Docteur **Nathalie Dejuçq-Rainsford**, directrice de l'équipe de recherche « Environnement viral et chimique, & Reproduction », pour m'avoir accueillie au sein de son équipe.

J'adresse mes sincères remerciements au Docteur **Marie-Christine Chaboissier**, directrice de recherche INSERM à l'institut Biologie Valrose, et au Docteur **Yann Guiguen**, directeur de recherche à l'INRA, pour avoir accepté d'être rapporteurs de cette thèse.

Je remercie encore une fois le Docteur **Bernard Jégou**, en tant que directeur de thèse, qui m'a guidé et conseillé pendant ces trois années, et qui figure parmi les personnes les plus inspirantes que je connaisse. Je tiens également à remercier le Docteur **Frédéric Chalmel**, mon directeur de thèse, pour ses conseils et son implication au quotidien à mes côtés.

Un grand merci au Docteur **Valérie Dupé**, tutrice de cette thèse, ainsi qu'aux Docteurs **Julien Bobe**, **Marie-Bérangère Troadec** et **Thierry Charlier**, membres de mon comité de thèse, pour leurs critiques constructives et leur bienveillance tout au long de ces trois années.

Je tiens également à remercier nos précieux collaborateurs, qui ont partagé avec nous leurs précieux échantillons, le Docteur **Paul Fowler** et son équipe, de l'université d'Aberdeen.

Je remercie également le **réseau doctoral de l'EHESP**, dont ses directeurs, Bernard Jégou et Judith Mueller, ainsi que Emmanuelle Guévara et Karine Laboux, de m'avoir offert l'opportunité de suivre de nombreuses formations interdisciplinaires de qualité. J'ai ainsi pu élargir mes connaissances au-delà de ma propre discipline, et apprécier l'étendue des domaines qui participent à l'effort commun d'améliorer les politiques de santé publique.

Je tiens à remercier l'ensemble de **mon équipe d'accueil**, avec qui je travaille depuis bientôt 4 ans. La diversité des travaux de recherches menées dans l'équipe m'a permis d'approfondir ma culture scientifique, et je vous remercie pour vos précieux conseils lors des réunions. Je voudrais remercier particulièrement Séverine, qui m'a éclairé de sa culture scientifique qui apparaît souvent sans limites. Avec Laetitia Lecante, Laurianne Lesné, Christèle Desdoit et Milissia Ben Maamar, elles prennent en charge ces échantillons humains si précieux. Sans cet investissement quotidien ma thèse n'aurait pas pu voir le jour. Pour cela, un grand merci !

Dans cette équipe, j'aimerais remercier le groupe avec qui j'ai interagi quotidiennement pendant les quatre dernières années. Bertrand Evrard, un collègue qui est toujours de bonne humeur, grand

sportif, et roi du labo. Aurélie Lardenois, experte du single-cell et de la capoeira, une force de la nature, possédant une montagne de compétences qui me fait rêver. Thomas Darde, ami fidèle et co-bureau, avec qui j'ai partagé tant de choses depuis toutes ces années. Antoine Rolland, mon co-encadrant officieux qui a été indispensable à mon travail quotidien. Je tiens à remercier une fois de plus Frédéric Chalmel, qui m'a encadré, soutenu et supporté tout au long de ma thèse. Je suis consciente que je te dois beaucoup, et je ne saurais exprimer toute ma reconnaissance en ces quelques lignes. Il serait difficile de remercier mes collègues de l'équipe sans apporter une dimension personnelle, car les relations que nous avons tous nouées ont été un soutien quotidien, au-delà des relations professionnelles. L'énergie de notre groupe n'a été que positive, et pour cela, je vous remercie tous.

J'aimerais également remercier tous les étudiants et stagiaires qui sont passés dans mon institut pendant ces trois années. Plus qu'un groupe de collègues, nos relations ont souvent dépassé le cadre du laboratoire. Tout d'abord, j'aimerais remercier Sadia, qui a impulsé cet esprit de groupe qui nous réunit. Je tiens à remercier Lauriane, Antonio, Charly, Laetitia, Salomé et Nathan, avec qui j'ai partagé de nombreux bons moments. Je tiens particulièrement à remercier Julie et Julie (« laquelle ? »), mes compères d'infortune, toujours présentes avec moi pendant ce dernier été de rédaction passé au labo. Vous êtes les prochaines, et je vous soutiendrai jusqu'à la fin ! Courage ! Je tiens particulièrement à remercier Léo, ma plus belle rencontre de ces trois années.

Au fil des années, j'ai eu la chance de rencontrer et de garder des amis fidèles, qui, bien qu'ils ne partagent pas ma vie quotidienne, sont des soutiens indéfectibles dont je ne saurais me passer. Je pense ainsi à Paul et à Sidwell, deux personnes que j'ai rencontrées pendant mes premières années à Rennes, et bien que je n'aie pas eu l'occasion de vous voir autant que je l'aurais voulu, je vous remercie de faire partie de ma vie.

Avant mes années à l'université, j'ai eu la chance de grandir à Cherbourg, et j'y ai rencontré les personnes qui forment maintenant la base sur laquelle je suis construite. Tout d'abord, je veux remercier Sabine et Sarah pour ce lien si particulier qui nous lie, et bien qu'on ne puisse plus se voir autant qu'auparavant, vous restez celles qui me manquent au quotidien ! Il me tarde de retrouver une vie sociale pour vous revoir ! Bien sûr, je voudrais remercier mes amis de toujours, « Cherbourg un jour, Cherbourg toujours » ! Nous nous connaissons depuis bientôt 10 ans, certains depuis 22 ans (!), et bien que l'on soit tous éparpillés aux quatre coins de la France, on ne s'est pas lâché depuis toutes ces années. Amélie, BenCoco, Charline, Dédé, Gégé, Grégoire, Jb, Maxime, Pauline, Sofia et Tuteur, il serait indécent d'expliquer combien vous m'êtes nécessaires dans ce paragraphe ;) Au-delà de tous ces bons moments (et mon dieu il y en a beaucoup !), nous avons toujours tout partagé, et je sais que cette amitié qui nous lie tous est indéfectible !

J'aimerais remercier celui sans qui rien n'aurait été possible, Thomas Darde ! Il serait difficile de résumer ici ces cinq années d'amitié ! Tu m'as soutenu plus que quiconque pendant cette thèse, qui

n'a pas toujours été facile pour moi. Mon mari de boulot, on pourrait dire ça ? Toujours de bonne humeur, toujours la crème de l'humour, indispensable pour une journée réussie ! J'espère t'avoir apporté au moins la moitié de ce que tu m'as apporté, ce serait déjà formidable !

Pour finir, j'aimerais évidemment remercier mes parents et ma sœur. Vous m'avez toujours soutenue, pendant toute ma scolarité, me soulageant des soucis de bien des manières. Je suis toujours restée mystérieuse sur mon travail, peut-être un peu trop, et j'espère que la lecture de ce manuscrit vous rendra fiers. Cherbourg restera toujours la ville où mes racines sont ancrées.

Enfin j'aimerais dire, à quiconque pourrait en douter, que ces trois années ont certainement été les plus difficiles jusqu'ici, mais je finis cette thèse avec la certitude que m'engager dans cette aventure a été la meilleure décision que j'ai prise dans ma vie.

Table des matières

Remerciements	I
Table des matières	V
Table des illustrations	VIII
Liste des abréviations et acronymes	XI
1. Introduction	1
1.1. La transcriptomique.....	1
1.1.1. Le dogme central de la biologie moléculaire	1
1.1.2. L'évolution du dogme et le transcriptome.....	3
1.1.2.1. Le génome et l'épigénome.....	3
1.1.2.2. Le transcriptome	8
1.1.2.2.1. Les ARN messagers	9
1.1.2.2.2. Les longs ARN non codants.....	9
1.1.2.2.3. Les petits ARN non codants.....	11
1.1.2.3. Le protéome	12
1.1.3. Les technologies liées à la transcriptomique	13
1.1.3.1. Les débuts de la transcriptomique : EST & SAGE.....	13
1.1.3.2. Les puces à ADN	13
1.1.3.3. Le séquençage à haut-débit et ultra-haut-débit	14
1.1.3.3.1. Seconde génération de séquenceurs	15
1.1.3.3.2. Troisième génération (PacBio, Nanopore).....	21
1.1.4. L'analyse bio-informatique des données de RNA-seq.....	25
1.1.4.1. Le contrôle qualité des données de séquençage.....	25
1.1.4.2. L'alignement des lectures sur le génome.....	25
1.1.4.3. L'assemblage des transcrits	26
1.1.4.4. La quantification des transcrits et la normalisation	27
1.1.4.5. L'analyse différentielle	28
1.1.4.6. La caractérisation de nouvelles isoformes et de nouveaux gènes	29
1.1.4.7. L'analyse fonctionnelle	30
1.1.4.8. L'analyse de réseaux.....	30
1.1.4.9. Intégration de données haut-débit : la multi-omique	31
1.1.4.9.1. RNA-seq et génomique	31
1.1.4.9.2. RNA-seq et épigénomique	31
1.1.4.9.3. RNA-seq et smallRNA-seq.....	32
1.1.5. Les bases de données et référentiels en transcriptomique	32
1.1.5.1. Les banques de données généralistes.....	32
1.1.5.1.1. Les banques de séquences nucléotidiques.....	32
1.1.5.1.2. Les banques de séquences protéiques	33
1.1.5.1.3. Les banques de données génomiques.....	33
1.1.5.2. Les banques de données dédiées à la reproduction.....	33
1.2. La différenciation gonadique chez les mammifères	37
1.2.1. Les gonades adultes et fonctions de reproduction	37
1.2.1.1. Structures des gonades adultes	37
1.2.1.2. La fonction exocrine des gonades adultes	38
1.2.1.3. La fonction endocrine des gonades adultes	39

1.2.1.3.1.	La stéroïdogenèse.....	39
1.2.1.3.2.	Les hormones polypeptidiques.....	40
1.2.1.3.3.	L'axe hypothalamo-hypophysaire.....	43
1.2.2.	La gonade indifférenciée.....	45
1.2.2.1.	La crête génitale.....	45
1.2.2.2.	Origine et migration des cellules germinales.....	47
1.2.3.	La différenciation sexuelle.....	51
1.2.3.1.	Les cellules somatiques : origine de la différenciation.....	51
1.2.3.1.1.	Trajectoire mâle.....	51
1.2.3.1.2.	Trajectoire femelle.....	57
1.2.3.1.3.	L'antagonisme des voies mâles et femelles.....	59
1.2.3.2.	Devenir des structures embryonnaires génitales.....	62
1.2.4.	Le développement gonadique.....	65
1.2.4.1.	Développement testiculaire.....	65
1.2.4.1.1.	Les cordons sexuels : tubules séminifères en devenir.....	65
1.2.4.1.2.	Tissu interstitiel.....	69
1.2.4.1.3.	Régulation endocrine : l'axe hypothalamo-hypophysaire gonadique.....	76
1.2.4.2.	Développement ovarien.....	79
1.2.4.2.1.	Formation des cordons ovariens.....	79
1.2.4.2.2.	Formation des follicules primordiaux.....	83
1.2.4.2.3.	Régulation endocrine : un rôle à définir.....	84
1.2.5.	Les troubles du développement sexuel.....	87
1.2.5.1.	Concept de TDS.....	87
1.2.5.2.	Les anomalies du tractus urogénital mâle et désordres de la fertilité.....	88
1.2.5.2.1.	Cryptorchidie.....	88
1.2.5.2.2.	Hypospadias.....	88
1.2.5.2.3.	Les cancers du testicule.....	88
1.2.5.2.4.	La qualité spermatique.....	89
1.2.5.3.	Perturbation endocrinienne et les gonades fœtales.....	91
1.3.	Analyses transcriptomiques du développement gonadique.....	94
1.3.1.	Premières études du développement gonadiques grâce aux puces à ADN.....	94
1.3.2.	Etude des cellules germinales primordiales.....	95
1.3.3.	Atout du RNA-seq pour l'étude de la différenciation sexuelle chez les espèces non-modèles.....	96
1.3.4.	Etude du développement gonadique chez les espèces modèles : complémentarité du RNA-seq et scRNA-seq.....	96
2.	Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines.....	99
2.1.	Introduction.....	99
2.2.	Manuscrit en préparation.....	100
2.3.	Résultats.....	165
2.3.1.	L'assemblage des transcrits.....	165
2.3.2.	Le raffinement et le choix du seuil d'expression.....	167
2.3.3.	L'analyse d'expression différentielle.....	167
2.3.4.	La conservation du programme transcriptionnel chez le modèle murin.....	168
2.3.4.1.	Intégration des puces à ADN : limites.....	169
2.3.4.2.	Intégration de données de RNA-seq.....	169
2.3.4.2.1.	Analyse bio-informatique et conversion des données de RNA-seq souris vers le génome humain.....	169

2.3.4.2.2.	Association des stades de développement humains et souris et corrélation des profils d'expression des transcrits	170
2.4.	Discussion et Perspectives.....	170
2.4.1.	L'étude comparative des mécanismes de régulation transcriptionnelle de l'Homme et de la souris	171
2.4.2.	L'intégration des données de GWAS.....	172
2.4.3.	La caractérisation transcriptomique du développement des gonades fœtales humaines à l'échelle de la cellule unique : scRNA-seq.....	173
3.	Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain	179
3.1.	Introduction	179
3.1.1.	L'ibuprofène.....	180
3.1.2.	Le chlordécone	182
3.2.	Matériel et Méthodes.....	183
3.2.1.	La collecte des échantillons, culture et exposition	183
3.2.2.	L'extraction de l'ARN et séquençage	185
3.2.3.	Les analyses bio-informatiques	186
3.2.3.1.	L'alignement et quantification.....	186
3.2.3.2.	La filtration et analyse de l'expression différentielle.....	186
3.2.3.3.	Les analyses fonctionnelles	187
3.3.	Résultats	187
3.3.1.	Séquençage et alignement sur le génome humain	187
3.3.2.	Quantification et filtration statistique.....	188
3.3.3.	La signature toxicogénomique de l'ibuprofène.....	189
3.3.4.	La signature toxico-génomique du chlordécone.....	191
3.3.5.	La comparaison des effets transcriptionnels des deux substances.....	193
3.4.	Discussion et perspectives.....	194
4.	The Reprogenomic Viewer.....	199
4.1.	Introduction	199
4.2.	Objectifs	199
4.3.	Résultats	214
4.3.1.	Description de la première version de RGV.....	214
4.3.2.	La nouvelle stratégie d'intégration.....	214
4.3.3.	Les nouvelles études intégrées	214
4.3.4.	Les nouveaux outils de visualisation.....	215
4.4.	Discussion et Perspectives.....	215
5.	Revue de la littérature.....	228
6.	Conclusion.....	243
	Références	249
	Résumé	1

Table des illustrations

Figure 1.	Le dogme central de la biologie moléculaire.	2
Figure 2.	Organisation du génome, de l'épigénome, du transcriptome et du protéome et transfert d'information.	4
Figure 3.	Structure d'un gène eucaryote et régulation de son expression.	6
Figure 4.	Illustration du protocole de RNA-seq sur la plateforme Illumina.	16
Figure 5.	Protocole de scRNA-seq et description des différentes approches expérimentales disponibles à chaque étape (Kolodziejczyk et al., 2015).	18
Figure 6.	Principe du DGE-seq (adapté de Morrissy et al., 2009)	20
Figure 7.	Étapes de l'analyse bio-informatique de données de RNA-seq	24
Figure 8.	Anatomie du testicule (A) et de l'ovaire adulte (B).	36
Figure 9.	Fonction et régulation endocrine des gonades adultes.	42
Figure 10.	Schématisation de la formation de la crête génitale et mécanismes moléculaires en jeu.	46
Figure 11.	Différenciation de la gonade bipotentielle en testicule ou ovaire.	50
Figure 12.	Régulation transcriptionnelle de la différenciation sexuelle.	58
Figure 13.	(A) Schéma des différents compartiments tissulaires du testicule fœtal : les cordons testiculaires et le tissu interstitiel. (B) Rôle central de la cellule de Sertoli dans la différenciation des autres populations cellulaires du testicule et dans le développement testiculaire.	64
Figure 14.	Mécanismes de régulation de la différenciation des cellules de Leydig fœtales (adapté de Wen et al., 2016).	70
Figure 15.	Stéroïdogénèse dans la cellule de Leydig.	72
Figure 16.	Évolution du nombre de cellules de Leydig, de la quantité de testostérone (testiculaire et plasmatique) et des taux de gonadotrophines sériques au cours de la vie fœtale chez l'Homme.	74
Figure 17.	Formation des nids ovocytaires et des follicules primordiaux dans l'ovaire fœtal.	78
Figure 18.	Nombre de cellules germinales au cours de la vie fœtale jusqu'à la vie adulte	82
Figure 19.	Schéma du syndrome de dysgénésie testiculaire selon Skakkebaek <i>et al.</i> 2016	86
Figure 20.	Différents modes d'action des Perturbateurs endocriniens	90
Figure 21.	Visualisation de l'alignement des lectures séquencées sur le génome humain hg19 grâce à IGV.	165
Figure 22.	Représentation en « boîte à moustache » du maximum d'abondance des transcrits à travers les échantillons en fonction de leurs classcodes.	166
Figure 23.	Représentation en nuage de point des résultats de pré-séquencage des 90 000 cellules capturées à partir des 30 gonades fœtales humaines.	174
Figure 24.	Effet de l'ibuprofène sur le testicule fœtal humain en culture organotypique (issu de Ben Maamar et al., 2017).	184
Figure 25.	Effet du chlordécone sur le testicule fœtal humain en culture organotypique (issu de Gaudriault et al., 2017).	185
Figure 26.	Résultats du séquençage et de l'alignement sur le génome humain (hg38)	188
Figure 27.	Résultats de la quantification des gènes et de la filtration statistique.	189
Figure 28.	Présentation des groupes de gènes sur- ou sous exprimés dans le testicule fœtal après exposition à l'ibuprofène.	190
Figure 29.	Présentation des groupes de gènes sur- ou sous exprimés dans le testicule fœtal après exposition au chlordécone.	192
Figure 30.	Diagramme de Venn représentant les gènes communs aux deux signatures toxicogénomiques de l'ibuprofène (Ibu) et du chlordécone (Chl) dans le testicule fœtal humain.	193
Figure 31.	Expression de CYP11A1, CYP17A1, HSD17B3, et de l'AMH.	194

Tableau I.	Homologies entre les structures embryonnaires génitales et leurs dérivés différenciés dans chaque sexe (vestiges embryonnaires)	62
Tableau II.	Ensemble des études intégrées dans la nouvelle version de RGV.....	227

Liste des abréviations et acronymes

17 β -HSD	Voir HSD17B	CGH array	Comparative genomic hybridization array
3 β -HSD	Voir HSD3B	CGP	Cellule germinale primordiale
aa	Acide aminé	ChIP	Chromatin immunoprecipitation
ACP	Analyse en composante principale	ChIP-seq	Chromatin immunoprecipitation sequencing
ACTR1A	ARP1 actin-related protein 1A, centractin alpha	CITED2	Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2
ADS	Anomalies du développement du sexe	COL2A1	Collagen, type II, alpha 1
ADN	Acide désoxyribonucléique	COUP-TF2	Voir NR2F2
ADNc	ADN complémentaire	COX	Cyclooxygenase
AHR	Aryl hydrocarbon receptor	CREBBP	CREB binding protein
AKR1C3	Aldo-keto reductase family 1, member C3	CTNNB1	Catenin beta 1
AKR1C2	Aldo-keto reductase family 1, member C2	CXCL12	C-X-C motif chemokine ligand 12
AMH	Anti-Mullerian hormone	CXCR4	C-X-C motif chemokine receptor 4
AMHR2	Anti-Mullerian hormone receptor, type II	CYP11A1	Cytochrome P450, family 11, subfamily A, polypeptide 1
AMPc	Adénosine monophosphate cyclique	CYP17A1	Cytochrome P450, family 17, subfamily A, polypeptide 1
AR	Androgen receptor	CYP26B1	Cytochrome P450, family 26, subfamily B, polypeptide 1
ARX	Aristaless related homeobox	DAX1	Voir NR0B1
ARN	Acide ribonucléique	DAZL	Deleted in azoospermia-like
ARNm	ARN messenger	DDX4	DEAD box polypeptide 4
ARNc	ARN non codant	DDX53	DEAD (Asp-Glu-Ala-Asp) box polypeptide 53
ARNr	ARN ribosomique	DES	Diéthylstilbestrol
ARNt	ARN de transfert	DGE-seq	Digital Gene Expression
ARNInc	ARN long non codant	DHEA	Dehydroepiandrosterone
ARNpnc	ARN petits non codant	DHH	Desert hedgehog
ARNr	ARN ribosomique	DHT	Dihydrotestosterone
ARX	Aristaless related homeobox	DKK1	Dickkopf WNT signaling pathway inhibitor 1
BMP	Bone morphogenetic protein	DLK1	Delta-like 1 homolog
BMPR1A	Bone morphogenetic protein receptor, type IA	DMRT1	Doublesex and mab-3 related transcription factor 1
CAGE	Cap analysis gene expression	DMSO	Diméthylsulfoxyde
CBX2	Chromobox 2	eQTL	expression quantitative trait loci
CCNA1	Cyclin A1	EMX2	Empty spiracles homeobox 2
CDK6	Cyclin dependent kinase 6	ENCODE	Encyclopedia of DNA Elements
CDKN1B	Cyclin dependent kinase inhibitor 1B		
CDKN2B	Cyclin dependent kinase inhibitor 2B		

ESR1	Estrogen receptor 1	LDL	Low-density lipoprotein
ESR2	Estrogen receptor 2	LHCGR	Luteinizing hormone/choriogonadotropin receptor
EST	Expressed Sequenced Tag	LHX9	Lim homeobox protein 9
FACS	Fluorescence-activated cell sorting	LIN28A	Lin-28 homolog A
FGF9	Fibroblast growth factor 9	L-PGDS	Lipocalin-type prostaglandin D2 synthase (#PTGDS)
FGFR2	Fibroblast growth factor receptor 2	LRP5	LDL receptor related protein 5
FOXL2	Forkhead box L2	LRP6	LDL receptor related protein 6
FPKM	Fragments per kilobase of exon model per million mapped reads	MAGE-A4	Melanoma associated antigen A4
FSH	Follicle stimulating hormone	MAMLD1	Mastermind like domain containing 1
FSHR	Follicle stimulating hormone receptor	MAP3K1	Mitogen-activated protein kinase kinase kinase 1
FST	Follistatin	MAP3K4	Mitogen-activated protein kinase kinase kinase 4
GADD45G	Growth arrest and DNA damage inducible gamma	MAPK11	Mitogen-activated protein kinase 11
GATA4	GATA binding protein 4	MAPK14	Mitogen-activated protein kinase 14
GFP	Green Fluorescence Protein	miARN	microARN
GnRH	Gonadotropin-releasing hormone	NANOG	Nanog homeobox
hCG	Human chorionic gonadotropin	NANOS2	Nanos homolog 2
hCGPLCs	Human primordial germ cells-like	NCBI	National Center of Biotechnology Information
HDL	High-density lipoprotein	NODAL	Nodal growth differentiation factor
hESCs	Human embryonic stem cells	NR0B1	Nuclear receptor subfamily 0, group B, member 1
hLH	Human luteinizing hormone	NR2F2	Nuclear receptor subfamily 2, group F, member 2
HMG	High-mobility group	NR5A1	Nuclear receptor subfamily 5, group A, member 1
HSD17B	Hydroxysteroid (17-beta) dehydrogenase	Nt	Nucléotides
HSD3B	Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase	NUMB	NUMB, endocytic adaptor protein
IGF1 & 2	Insulin-like growth factor 1 & 2	NTKR	Neurotrophic tyrosine kinase receptor
IGF1R	Insulin-like growth factor 1 receptor	OMS	Organisation mondiale de la santé
INHA	Inhibin, alpha	PAX2	Paired box 2
INHBB	Inhibin subunit beta B	pb	Paires de bases
INSL3	Insulin-like 3	PCR	Polymerase chain reaction
INSR	Insulin receptor	PDB	Protein Data Bank
INSRR	Insulin receptor related receptor	PDGF	Platelet-derived growth factor
JAG1	Jagged 1	PDGFRA	Platelet-derived growth factor receptor, alpha polypeptide
Jpc	Jour post-conception	PE	Perturbateur endocrinien
Kb	Kilobases		
KDM3A	Lysine demethylase 3A		
KIT	KIT proto-oncogene receptor tyrosine kinase		
KITLG	KIT ligand		
KLF4	Kruppel like factor 4		

PGD2 Prostaglandin D2
 piARN ARN interagissant avec la protéine PIWI
 PIT Proteomics informed by Transcriptomics
 POD1 Voir TCF1
 polyA Poly-adénylé
 POU5F1 POU domain, class 5, transcription factor 1
 PRDM14 PR/SET domain 14
 préARNm ARN pré-messager
 pré-miARN précurseur des miARN
 PTCH1 Patched homolog 1
 PTGDS Prostaglandin D2 synthase
 RISC RNA-silencing complex
 RNA-seq RNA sequencing
 RPKM Reads per kilobase of exon model per million reads
 RSPO1 R-spondin 1
 RXFP2 Relaxin/insulin-like family peptide receptor 2
 SA Semaine d'aménorrhée
 SAGE Serial Analysis of Gene Expression
 SCARB1 Scavenger receptor class B, member 1
 SCF Stem cell factor
 scRNA-seq Single-cell RNA-seq
 SD Semaine de développement
 SF1 Steroidogenic factor 1
 SIX1 SIX homeobox 1
 SIX4 SIX homeobox 4
 Smart-seq Switch Mechanism at the 5' End of RNA Templates sequencing
 SNP Single nucleotide polymorphisme
 SOX10 Sry-like HMG-box protein 10
 SOX17 Sry-like HMG-box protein 17
 SOX2 Sry-like HMG-box protein 2
 SOX8 Sry-like HMG-box protein 8
 SOX9 Sry-like HMG-box protein 9
 SP1 Sp1 transcription factor
 SRB1 Voir SCARB1
 SRY Sex determining region chromosome Y

STAR Steroidogenic acute regulatory protein
 STRA8 Stimulated by retinoic acid 8
 TCF21 Transcription factor 21
 TDS Testicular Dysgenesis Syndrome
 TES Testis-specific enhancer
 TFAP2C Transcription factor AP-2 gamma
 TGFβ Transforming growth factor, beta
 TMM Trimmed mean of M-values
 TPM Transcripts per million
 TSS Transcription start site
 VEGFA Vascular endothelial growth factor A
 WGCNA Weighted Correlation Network Analysis
 WNT4 Wingless-type MMTV integration site family, member 4
 WT1 Wilms tumor 1
 ZFPM1 zinc finger protein, FOG family member 1

1. Introduction

1.1. La transcriptomique

1.1.1. Le dogme central de la biologie moléculaire

L'hérédité a suscité l'intérêt des scientifiques de tout âge (Lherminier, 1998). Le terme « gène » a été inventé pour désigner les unités définissant un caractère et responsables de son hérédité (Johannsen, 1909). Au début du 20^{ème} siècle, les scientifiques pensaient que l'information génétique était portée par les protéines (Le Guyader, 2001). À la même époque, une des caractéristiques différenciant les animaux et les plantes était la composition de leurs tissus, le premier contenant de l'acide nucléique de thymus (acide thymonucléique) et le second contenant de l'acide nucléique de levure, ainsi nommés selon les premiers échantillons à partir desquels ils ont été purifiés (Cammack et al., 2006). Lorsque l'acide nucléique de levure a été détecté dans des cellules animales, ces deux molécules ont été renommées selon leurs structures biochimiques : l'acide thymonucléique devint l'acide désoxyribonucléique, l'ADN et l'acide nucléique de levure devint l'acide ribonucléique, l'ARN (Allen, 1941). À la fin des années 1920, Phœbus Aaron Levene avait identifié ses quatre composants de l'ADN : la cytosine, l'adénine, la guanine et la thymine (Tipson, 1957). Les bases de la biologie moléculaire sont posées grâce au concept « un gène – une enzyme » proposé par Beadle et Tatum en 1941 (Beadle and Tatum, 1941; Strauss, 2016). Trois ans plus tard, Oswald Avery découvre que les gènes sont portés par l'ADN (Avery et al., 1944), dont la structure en double hélice sera mise en évidence par Watson et Crick en 1953 (Watson and Crick, 1953), inspirés par les travaux de Rosalind Franklin (Franklin and Gosling, 1953a, 1953b) (Stasiak, 2001). C'est en 1956 que l'on retrouve les premières traces du « dogme central de la biologie moléculaire » émis par Crick, qui fait état de l'organisation du flux d'information entre ces entités (Crick, 1958).

L'ADN est donc une macromolécule double brin, présent dans le noyau des cellules et composée d'acides nucléiques, l'adénine, la cytosine, la guanine et la thymine. L'ordre de ces quatre nucléotides définit la séquence et de ce fait, l'information génétique. La séquence d'un brin est complémentaire de l'autre brin, grâce à la complémentarité des bases : l'adénine et la cytosine s'associeront avec la thymine et la guanine respectivement. Celle-ci sera transcrite par l'ARN polymérase pour former une molécule d'ARN simple brin, composé des mêmes nucléotides de l'ADN à l'exception de la thymine qui est où le nucléotide terminal porte un groupement phosphate

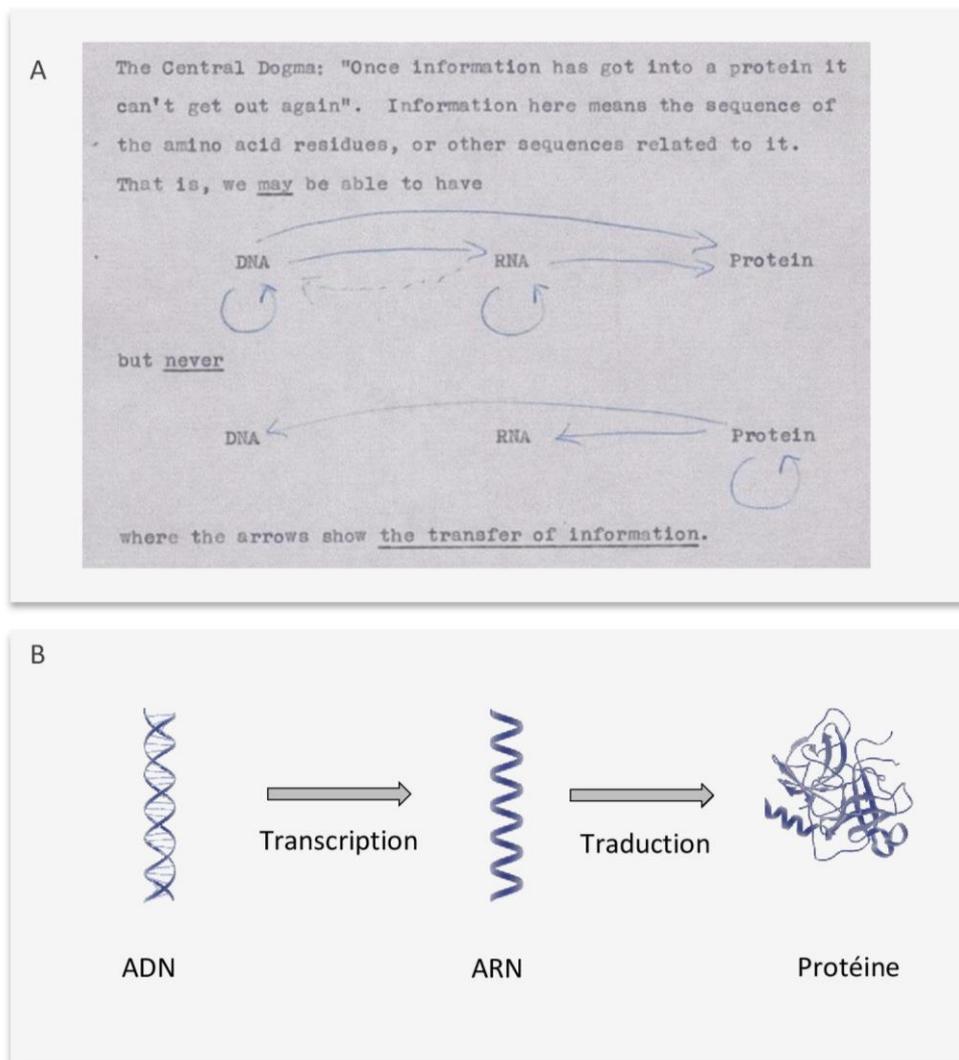


Figure 1. Le dogme central de la biologie moléculaire.

(A) Première explicitation du dogme central, note non publiée de Crick en 1956. Credit: Wellcome Library, London. (B) Schématisation du dogme central de la biologie moléculaire. L'information génétique, responsable de l'hérédité des caractères, est portée par l'ADN. L'ADN peut être dupliqué, pour permettre la transmission de l'information qu'il contient et être transcrit en ARN qui sera lui-même traduit en protéine.

(sur son carbone n 5), vers l'extrémité 3', où le dernier nucléotide porte un groupement hydroxyle (OH sur son carbone n 3). Le sens de transcription est donc important et cela induit que deux gènes peuvent être localisés chacun sur un brin dans une même région génomique. Le site d'initiation de la transcription (*transcription start site* = TSS) va définir le cadre de lecture de l'ADN, qui regroupe par triplets consécutifs les nucléotides de la séquence. Dans le cas de gène codant pour des protéines, ces triplets correspondent à des codons, qui vont définir la séquence en acides aminés de la protéine. Le cadre ouvert de lecture désigne le cas où l'on retrouve un codon d'initiation susceptible d'être reconnu par un ribosome sur l'ARN et un codon-stop, terminateur de la traduction. Dans ce cas, l'ARN transcrit dans ce cadre pourra potentiellement être traduit en protéine.

Après son adressage dans le cytoplasme, l'ARN sera traduit en protéine par un ribosome. En 1960, Francois Jacob et Jacques Monod démontrent que l'ARN messager (ARNm) constitue bien l'intermédiaire entre l'ADN et la synthèse des protéines, découverte pour laquelle ils recevront le prix Nobel en 1965 (Brenner et al., 1961; 1965). Chaque triplet de ribonucléotides est appelé codons, correspondant à un acide aminé (aa) donné, constituants de base des protéines. Les codons permettent également de guider la lecture par le ribosome du codon initiateur (AUG) jusqu'au codon terminateur (codon stop). La liaison séquentielle des acides aminés par des liaisons peptidiques permet donc la construction de la protéine (Lodish et al., 2000). Les protéines sont des macromolécules polypeptidiques qui assurent l'activité enzymatique, la structure, la mobilité, la réceptivité et la régulation des fonctions d'une cellule (Lodish et al., 2000). La séquence en aa va définir les caractéristiques biochimiques et structurales de la protéine et donc définir sa fonction.

1.1.2. L'évolution du dogme et le transcriptome

Le dogme central est resté cohérent avec les résultats expérimentaux pendant de nombreuses années, mais l'évolution des techniques de biologie moléculaire a permis de faire évoluer cette vision simplifiée du flux de l'information génétique (Le Guyader, 2001) (0).

1.1.2.1. Le génome et l'épigénome

Ces dernières décennies, les avancées technologiques de séquençage ont permis l'analyse à grande échelle de la séquence de l'ADN (voir chapitre «Les technologies liées à la transcriptomique»). L'ensemble de l'information génétique contenue dans les cellules d'un organisme a donc pu être décrite, que l'on désigne maintenant comme le génome. La mise en place de projets de grande envergure (Moraes and Góes, 2016), comme le Human Genome Project (Baetu, 2012), ou ENCODE (Gerstein et al., 2007) ont permis des avancées sans précédent dans la compréhension de la structure des génomes, mais paradoxalement, a brouillé peu à peu le concept de gène.

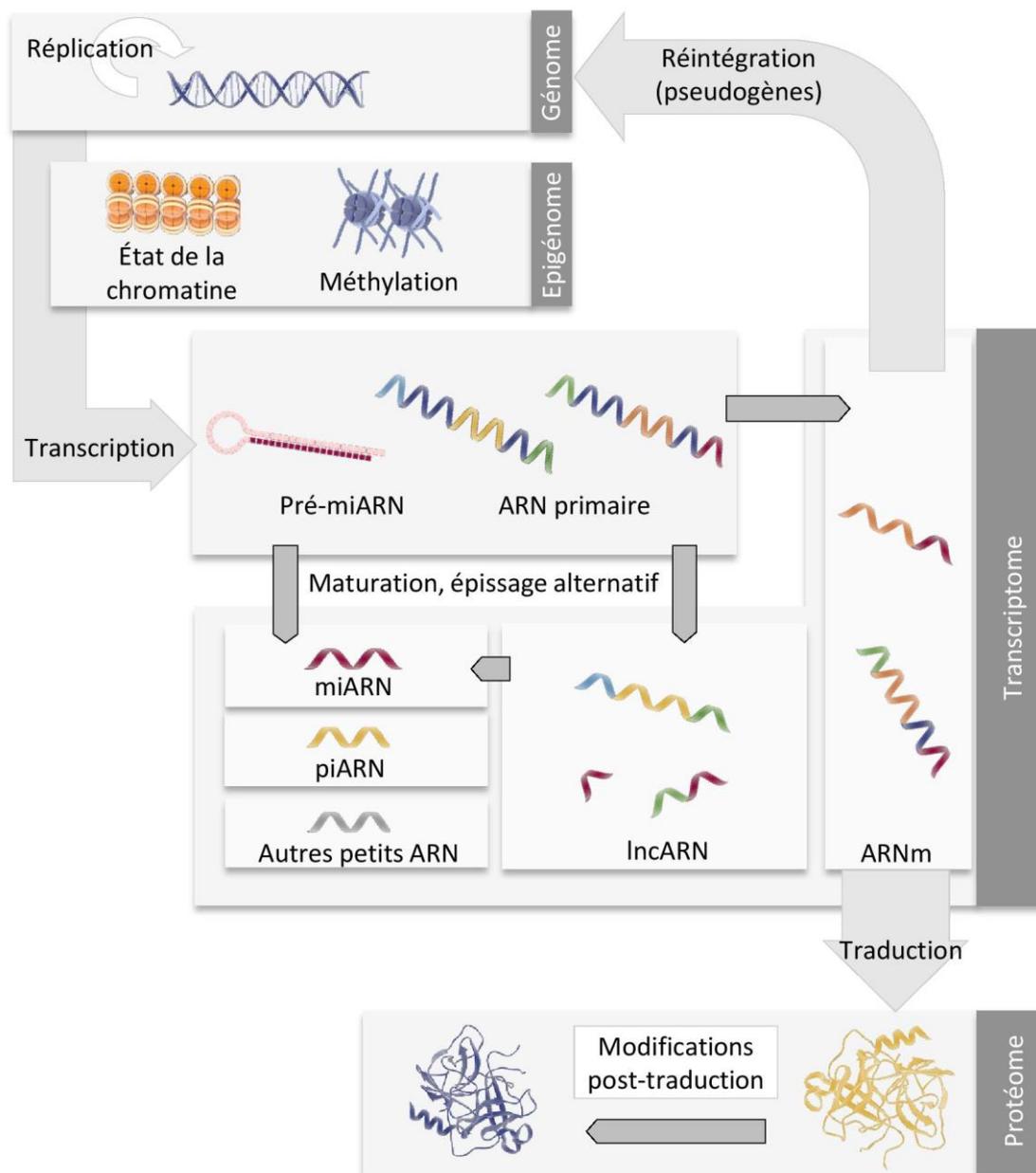


Figure 2. Organisation du génome, de l'épigénome, du transcriptome et du protéome et transfert d'information.

Aujourd'hui, il n'existe aucun consensus sur sa définition (Le Guyader, 2001). Il peut être généralement défini comme « *une union de séquences génomiques codant pour un ensemble cohérent de produits fonctionnels potentiellement concordants* » (Gerstein et al., 2007), ou plus précisément, comme « *l'unité de base de l'hérédité, consistant en un segment d'ADN organisé de manière linéaire le long du chromosome. Un gène code pour une protéine spécifique, un segment de protéine ou un ARN non codant* » (NCBI, National Center of Biotechnology Information). Chez l'Homme, le génome comprend environ 20 000 gènes codant pour une protéine et 22 000 gènes non codants (Zerbino et al., 2018).

Si un gène correspond à une union de séquences permettant la production d'un produit fonctionnel, alors les régions génomiques qui participent à la régulation de son expression font partie intégrante de celui-ci. D'une part, le promoteur d'un gène correspond à une région généralement proximale qui ne sera pas transcrite en ARN. Chez les eucaryotes, il est composé de la boîte « TATA », région riche en thymine et en adénine, située à 25-30 nucléotides en amont du TSS, qui sera reconnue par l'ARN polymérase (Krebs et al., 2017). D'autres régions plus ou moins éloignées du TSS peuvent également être reconnues par des facteurs de transcription et leur fixation conditionnera la formation du complexe d'initiation. Des cofacteurs activateurs (*enhancers*) ou inhibiteurs (*silencers*) entrent également en jeu, afin de promouvoir ou d'inhiber l'expression d'un gène (Spilianakis et al., 2005). Ils peuvent se lier à des promoteurs distaux spécifiques d'un gène, parfois localisés à plusieurs milliers de nucléotides du promoteur proximal. Même dans ce cas, les courbures appliquées à l'ADN permettent l'interaction des cofacteurs avec le complexe d'initiation et peuvent induire sa stabilisation ou déstabilisation, promouvant ou inhibant la transcription.

Les standards de la biologie moléculaire désignent le brin portant le gène transcrit comme le brin « sens » et le brin complémentaire comme le brin « antisens ». Le brin antisens du gène sera utilisé comme matrice par l'ARN polymérase afin de définir la séquence de nucléotides complémentaires à incorporer pour former le brin d'ARN primaire, copie du gène sens. Des sites alternatifs d'initiation de la transcription peuvent exister pour un même gène, aboutissant à la production d'ARN tronqués et de fait, à des protéines différentes ou même des ARN non codants (Krebs et al., 2017). De plus, sur une même région génomique, plusieurs gènes chevauchants peuvent être situés dans différents cadres de lecture, sur le brin sens ou antisens, codant pour des ARN fondamentalement différents (Contreras et al., 1977; Krebs et al., 2017).

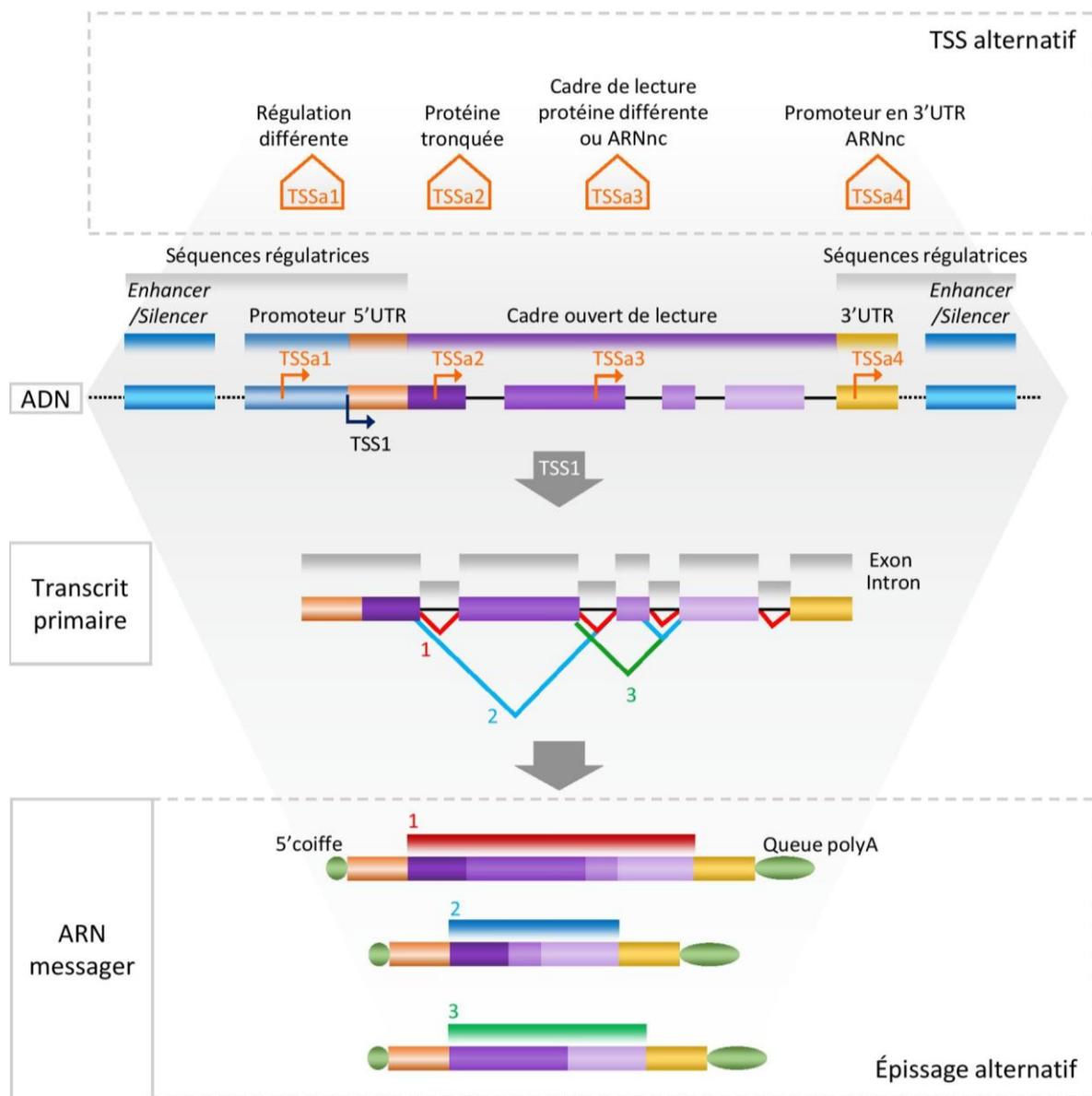


Figure 3. Structure d'un gène eucaryote et régulation de son expression.

Chez les eucaryotes, les gènes codants peuvent être schématiquement décomposés en 2 parties : le promoteur d'une part, qui est constitué de séquences régulatrices, et la séquence transcrite d'autre part. Classiquement un promoteur est constitué d'une boîte TATA et d'un ou plusieurs site(s) de fixation pour des facteurs de transcription, *enhancer* ou *silencer*, localisé(s) en amont ou en aval du site d'initiation de la transcription (TSS = *Transcription start site*). La transcription d'un gène va générer un transcrit primaire composé d'une succession d'introns et d'exons. Ce transcrit primaire va ensuite subir une maturation : les introns vont être éliminés par épissage, son extrémité 5' va être modifiée par addition d'une coiffe et son extrémité 3' va subir un clivage et être polyadénylée. L'ARNm mature ainsi généré contient des extrémités 5' et 3' non codantes (UTR = *Untranslated regions*) et un cadre de lecture ouvert ou région *Untranslated regions*) et un cadre de lecture ouvert ou région codante. Les possibilités de régulation de l'expression d'un gène sont multiples chez les eucaryotes. Tout d'abord l'utilisation de sites d'initiation de la transcription (TSS= *Transcription start site*) alternatifs permet de générer plusieurs transcrits primaires pour un même gène. Les différences de séquences de ces transcrits vont soit affecter l'extrémité 5' non traduite (5'UTR) des ARNm correspondants, autorisant alors des régulations différentes de leur traduction, soit affecter la séquence codante des ARNm, permettant alors la synthèse de protéines différentes, voir même d'ARN non codants (ARNnc). De plus, par le processus d'épissage alternatif, à un même transcrit primaire va pouvoir correspondre plusieurs ARNm. Ces ARNm peuvent présenter une stabilité

Le gène est également défini par sa structure en mosaïque (Berget et al., 1977), composée

Le gène est également défini par sa structure en mosaïque (Berget et al., 1977), composée d'exons et d'introns. Au sein de la région transcrite en ARN primaire, seuls les exons peuvent être potentiellement conservés lors de la maturation du transcrit, les introns seront quant à eux éliminés par le processus d'épissage. Chez les eucaryotes, la plupart des gènes sont multi-exoniques, contenant également un ou plusieurs introns. Cependant, 3% des gènes humains sont mono-exoniques et ne contiennent pas d'introns (Grzybowska, 2012).

Les régions intergéniques contiennent également de nombreux éléments tels que les pseudogènes, désignés à l'origine comme des « gènes fossiles ». D'une part, ces pseudogènes peuvent provenir de la duplication d'une région génomique (pseudogène *nonprocessed*). Dans ce cas, ils peuvent conserver leur région promotrice ou trouver de nouveaux promoteurs, ce qui modifie leur régulation. D'autre part, ils peuvent être le fruit de la réintégration d'un ARNm dans l'ADN, ce qui modifie leur structure, ne contenant plus que des exons (pseudogène *processed*). On retrouve environ 15 000 pseudogènes dans le génome humain, ce qui contribue à l'augmentation de sa taille (Zerbino et al., 2018). Cependant, la plupart de ces nouveaux gènes perdent leur fonctionnalité à cause de mutations, pouvant empêcher leur transcription ou inclure un codon-stop précoce dans leur séquence (revue dans (Li et al., 2013b; Shi et al., 2016; Tutar, 2012). Ainsi, certains d'entre eux sont transcrits et produisent un ARN non codant, qui peut participer à la régulation de l'expression des gènes (voir chapitre « Les longs ARN non codants ») (Li et al., 2013b; Shi et al., 2016). En revanche, certains d'entre semblent conserver un cadre ouvert de lecture et peuvent toujours être traduits en protéines (revue dans Xu and Zhang, 2016).

Le génome comprend également des séquences d'ADN qui sont présentes en plusieurs copies, appelés éléments répétés. Les différentes classes d'éléments répétés représentent 5 % du génome humain (Richard et al., 2008). D'une part, l'ADN satellite correspond à des séquences de faible complexité accumulées au niveau des centromères et des télomères des chromosomes (extrémités et point de contact des chromatides). Il existe également des séquences dites « égoïstes », capables de se déplacer ou de se dupliquer au sein du génome : les transposons et rétrotransposons. Certains éléments transposables peuvent produire une classe de petits ARNs, les piARNs (voir « Les petits ARN non codants »). Ceux-ci sont impliqués dans la régulation du développement de la drosophile (Rouget et al., 2010). Globalement, la régulation transcriptionnelle et post-transcriptionnelle des transposons et rétrotransposons est principalement contrôlée par la méthylation de l'ADN et par les piARN. L'inhibition de leur expression est nécessaire afin d'éviter la duplication et l'insertion anarchique de ces séquences, ce qui pourrait entraîner des effets délétères au sein du génome. Les éléments transposables semblent impliqués dans la sélection des ovocytes fœtaux murins qui seront dégradés par attrition (Malki et al., 2014), et dans la régulation de l'expression de certains gènes lors de la spermatogenèse (Davis et al., 2017). Les conséquences de cette activité

d'insertion et de duplication dans l'évolution des génomes n'ont pas encore été décryptées (Sotero-Caio et al., 2017).

Par rapport au dogme central de la biologie moléculaire initialement énoncé, le flux de l'information génétique est bidirectionnel, illustré par la capacité de certains ARN à être intégré au génome. De plus, le statut de l'ARN en tant que messenger entre l'information d'un gène vers une protéine est infirmé, car les TSS et l'épissage alternatif permettent à un gène de produire plusieurs ARN ou protéines fonctionnelles (Figure 3).

L'épigénome définit les processus induisant des modifications héréditaires de l'expression des gènes qui ne sont pas dues à une modification de l'ADN (Carlberg and Molnár, 2018). Il contribue directement à la régulation de l'expression des gènes. Cela peut inclure des modifications biochimiques de l'ADN, des protéines qui lui sont associées ou des processus de régulation médiés par l'ARN (revu dans Allis and Jenuwein, 2016). L'ADN s'enroule autour d'histones pour former une structure compacte appelée chromatine. Celle-ci peut être dans un état condensé, l'hétérochromatine, qui bloque l'accès des différents facteurs initiant la transcription de leur gène cible. L'euchromatine désigne l'état décondensé de cette structure et rend donc accessible l'ADN pour la transcription. Les modifications biochimiques des histones peuvent moduler l'état de condensation de l'ADN autour de ces protéines : la méthylation induit généralement la fermeture de la chromatine et l'acétylation provoque son ouverture. Les différentes modifications biochimiques sont appelées marques d'histone. Par exemple, la marque H3K9me3 désigne une triméthylation de la lysine 9 et est associée à la répression de l'expression. La modification de la chromatine peut également être médiée par de petits ARN (voir chapitre « Les petits ARN non codants ») (Grewal and Elgin, 2007). La méthylation directe de l'ADN, en particulier la 5-méthylcytosine, induit la répression de l'expression des gènes (Allis and Jenuwein, 2016).

1.1.2.2. Le transcriptome

Le transcriptome peut être défini comme l'ensemble des transcrits (ARN) présent dans un échantillon biologique, que ce soit à l'échelle d'un organisme (bactérie, organisme complexe), d'un organe, d'une population cellulaire ou d'une cellule unique. Il est donc directement dépendant du génome et de l'épigénome. Cette grande interdépendance des différentes strates organisant l'activité d'une cellule permet la grande plasticité du transcriptome au sein du règne animal. D'une part, les ARN messenger (codants) pourront être traduits en protéines et d'autre part, les ARN non codants (ARNnc) sont des entités fonctionnelles à part entière (Cech and Steitz, 2014). Parmi les ARNnc, les ARN ribosomiques (ARNr) participent au complexe en charge de la lecture de l'ARNm et de l'assemblage des acides aminés et les ARN de transfert (ARNt), capables de porter un acide aminé

correspondant à un codon spécifique. Ces deux classes participent donc à la traduction et ne seront pas développées dans ce manuscrit.

1.1.2.2.1. Les ARN messagers

Pour mieux appréhender la complexité de la définition du gène, l'initiation de la transcription et la production de l'ARN primaire, ou ARN pré-messager (préARNm), ont déjà été décrits (voir «Le dogme central de la biologie moléculaire»). Ces préARNm subiront différentes étapes de maturation. Tout d'abord, une coiffe sera ajoutée à leur extrémité 5' pour éviter leur dégradation par les enzymes nucléaires, notamment les exonucléases.

Les événements d'épissage alternatifs permettront l'excision des introns et de certains exons. En effet, ce processus peut conserver une combinaison donnée d'exons, qui définira l'isoforme d'un transcrit. Suite à ce mécanisme, certains gènes pourront produire des ARN sans aucun exon en commun et donc des produits fonctionnels fondamentalement différents. Les jonctions intron-exons contiennent des séquences nucléotidiques conservées qui pourront être reconnues par le complexe protéique en charge de l'épissage, le spliceosome. Il est difficile de prédire quelle isoforme sera transcrite dans une cellule ou un tissu donné, car les mécanismes régulant l'épissage sont encore mal connus (De Conti et al., 2013). Par abus de langage, les exons sont souvent désignés comme les parties codantes du gène qui peuvent être conservées après épissage, mais les mêmes processus ont lieu dans la maturation de certains ARNnc (Quinn and Chang, 2016). Dans le génome humain, environ deux tiers des gènes sont capables de telles recombinaisons lors de leur transcription (Tress et al., 2017). Dans le cas d'un gène codant pour une protéine, les différentes isoformes d'ARNm produites pourront aboutir à des protéines tronquées, portant des domaines communs, ou même complètement différentes. L'ajout d'une queue polyA à l'extrémité 3' des ARNm permettra leur adressage du noyau vers le cytoplasme, ainsi que leur stabilisation et leur traduction.

Cependant, une étude transcriptomique comparative de différents tissus humains a démontré que l'épissage alternatif affectait davantage les ARNnc que les ARN codants et que les sites alternatifs d'initiation de la transcription ou de polyadénylation contribuaient davantage à la diversité des isoformes (Reyes and Huber, 2018).

1.1.2.2.2. Les longs ARN non codants

Les ARN longs non codants (ARNlnc) constituent la majorité des ARNnc, d'une longueur de 200 nt ou plus. La concomitance de l'émergence des technologies de séquençage et de l'intérêt grandissant de la communauté scientifique pour ces entités moléculaires nouvellement découvertes ont permis la caractérisation rapide des ARNlnc dans les génomes de nombreuses espèces (Uszczyńska-Ratajczak et al., 2018). Ils sont notamment répertoriés dans la base de données

NONCODE (Fang et al., 2018). Cela a permis de mettre en évidence que la classe des ARNlnc n'a pas de réelle pertinence biologique, car ils peuvent être produits à partir de divers éléments génomiques et sont impliqués dans une myriade de fonctions (Marchese et al., 2017 ; Wang and Chang, 2011 ; Wu et al., 2017).

Bien que certains d'entre eux suivent les mêmes étapes de biogenèse que les ARNm, il a été démontré que leur niveau d'expression était globalement plus faible, mais également plus spécifique d'un tissu (Ulitsky and Bartel, 2013). Comme énoncé lors de la définition du gène, les gènes non codants peuvent se trouver dans des régions intragéniques, c'est-à-dire chevauchant d'autres gènes, dans un intron ou sur le brin antisens. Les éléments transposables et les pseudogènes peuvent également être une source de ARNlnc (Johnson and Guigó, 2014) (voir « Le dogme central de la biologie moléculaire »).

Généralement, les ARNlnc sont impliqués dans la régulation de l'expression des gènes par des voies variées (revue dans Hu et al., 2012 ; Quinn and Chang, 2016 ; Shi et al., 2013 ; Yang et al., 2014). Leur transcription peut indirectement influencer l'expression des gènes voisins, en maintenant la structure de la chromatine ouverte (régulation positive), ou grâce à l'occupation stérique de leur ARN polymérase (régulation négative). Au niveau transcriptionnel, ils participent également au remodelage de la chromatine en guidant les acteurs de la modification des histones vers une région spécifique. Ils peuvent également se fixer aux facteurs de transcription, enhancer ou silencer et induire leur recrutement vers le gène cible ou au contraire, leur séquestration. Au niveau post-transcriptionnel, ils peuvent s'hybrider sur d'autres ARN. Ce processus est en général observé lorsque la région codant pour l'ARNlnc se trouve en antisens d'un gène codant. Les ARN sont donc complémentaires, leur appariement pourra avoir des conséquences variées sur la maturation, la stabilité et la traduction de l'ARN cible. L'hybridation imparfaite d'un ARNlnc pourra également induire la dégradation de sa cible (Quinn and Chang, 2016). Au contraire, ils peuvent également prévenir la dégradation de l'ARN médiée par les microARN (miARN, voir « Les petits ARN non codants »), en se fixant à ceux-ci, agissant comme une « éponge ». Au niveau traductionnel, ils peuvent masquer les sites de fixation des miARN sur un ARNm cible et ainsi protéger sa traduction. Paradoxalement, certains ARNlnc peuvent être transformés en miARN lors de leur biogenèse (Quinn and Chang, 2016). Ils peuvent également se lier à des protéines, afin de participer à la formation de complexe protéique ou moduler leur activité.

Chez les mammifères, les ARNlnc sont notamment impliqués dans la pluripotence des cellules embryonnaires, dans la différenciation cellulaire, la régulation du cycle cellulaire et la réponse immunitaire (Hu et al., 2012 ; Yang et al., 2014). Chez l'homme, au même titre que les ARNm, les ARNlnc sont associés à de nombreux désordres et pathologies (Chen et al., 2013).

Pour finir, dans une séquence nucléotidique aléatoire, un codon initiateur – séquence ATG- apparaît tout les 64 nucléotides et un codon-stop apparaîtra dans le même cadre de lecture tout les 21,3 nucléotides en moyenne. Des études statistiques ont démontré que même les ARNnc de plus de 100 nucléotides peuvent contenir des cadres ouverts de lecture (Dinger et al., 2008 ; Inagaki et al., 2016). De récentes études ont mis en évidence l'existence de gènes codant pour de petits peptides dans le génome eucaryote (Frith et al., 2006 ; Hanada et al., 2007 ; Kastenmayer et al., 2006 ; Oyama et al., 2004), qui peuvent avoir une fonction biologique (Hashimoto et al., 2008). Une étude de protéomique à grande échelle a également identifiée des peptides comme le fruit de la traduction de pseudogènes ayant conservés un cadre ouvert de lecture (Kim et al., 2014; Xu and Zhang, 2016). Dès lors, il est probable que certains gènes annotés comme ARNlnc ou pseudogènes non codants peuvent correspondre à des gènes codant pour de petits peptides et il est nécessaire de développer des approches permettant d'évaluer avec plus acuité le caractère non codant d'un ARN (Kageyama et al., 2011) (voir chapitre « La caractérisation de nouvelles isoformes et de nouveaux gènes »).

1.1.2.2.3. Les petits ARN non codants

Les ARN non codants d'une longueur inférieure à 200 nucléotides sont désignés sous le nom de petits ARN non codant (ARN petits non codant, ARNpnc). Cette classe d'ARN n'a pas réellement de pertinence biologique, car elle regroupe de très nombreuses sous-classes, présentant chacune des caractéristiques différentes, impliquées dans des processus de régulation différents, revus dans (Luo et al., 2016; Pek and Okamura, 2015).

Parmi les ARNpnc les mieux caractérisés, les micro ARN (miARN) sont des molécules endogènes composées de 21 à 24 nucléotides (Liu et al., 2014). Ils sont généralement produits par la transformation de certains ARNlnc du noyau en précurseur pré-miARN par la protéine Drosha. Dans le cytoplasme, le pré-miARN est clivé en miARN double brins mature par la protéine Dicer. L'un des brins sera ensuite incorporé dans le complexe protéique RISC (*RNA-silencing complex*), ce qui permettra de le guider vers son ARNm complémentaire. Ainsi, l'ARNm cible pourra être dégradé, ou sa traduction inhibée (Liu et al., 2014).

Les ARN interagissant avec la protéine PIWI (piARN) sont une autre classe de ARNpnc connus pour leur rôle important pendant la spermatogenèse (Luo et al., 2016). Ils sont produits à partir d'éléments répétés du génome et sont composés de 24 à 30 nucléotides. Ils ont la capacité d'interagir avec les protéines de la famille des PIWI, ce qui leur permet d'induire la dégradation d'ARN (Fu and Wang, 2014).

1.1.2.3. Le protéome

Au même titre que le transcriptome désigne l'ensemble des transcrits exprimés dans un échantillon, le protéome regroupe l'ensemble des protéines qu'il contient.

Leur synthèse est assurée par les ribosomes, à partir des ARNm. Leur séquence en acides aminés définit leur structure primaire, base de leur fonction biologique (Berg et al., 2002). La structure secondaire correspond aux arrangements locaux des acides aminés, pouvant former des structures en hélices α ou feuillettes β , stabilisés par des liaisons hydrogènes. La structure tertiaire représente le repliement général de la protéine et l'assemblage de plusieurs unités protéiques aboutit à la structure quaternaire. Les modifications post-traductionnelles qui peuvent intervenir sur la structure biochimique des peptides et donc modifier la stabilité de leur structure et leur fonction. La structure des protéines peut en général être prédite par des outils de modélisation à partir de leur séquence en acides aminés (Pavlopoulou and Michalopoulos, 2011). Cependant, il est intéressant de noter que jusqu'aujourd'hui, 44 à 54 % du protéome eucaryote est distingué comme le protéome « noir », dont la structure n'est pas identifiable par les approches expérimentales classiques ou les modélisations par homologie et dont la fonction reste difficile à définir (Perdigão et al., 2015).

De fait, les isoformes d'ARNm peuvent aboutir à des protéines possédant différentes séquences polypeptidiques et donc à un équilibre physico-chimique modifié menant à diverses structures et fonctions. Une étude transcriptomique comparative des différents tissus humains a permis de mettre en lumière que l'épissage alternatif affectait davantage les ARNnc que les gènes codants, ce qui explique la faible complexité du protéome au regard de la myriade d'isoformes potentielles d'ARNm (Reyes and Huber, 2018).

Le dogme central de la biologie moléculaire énoncé par Crick en 1970 (Crick, 1970) est donc obsolète aujourd'hui. Cette vision initiale du flux de l'information génétique, construite sur les connaissances de l'époque, a été totalement bouleversée. Comme décrit dans cette introduction, il est possible pour un ARN d'être intégré au génome pour potentiellement devenir un nouveau gène fonctionnel. De plus l'association d'un gène avec une protéine fonctionnelle s'est grandement complexifiée, un gène pouvant produire de nombreux produits fondamentalement différents par leur nature (ARN ou protéine) ou par leur composition (isoformes de transcrits, domaines protéiques différents). Dès lors, les noms des gènes, historiquement dérivés de leur fonction, ont maintenant une pertinence limitée quant à leur description et la nomenclature utilisée aujourd'hui a dû s'adapter (Wain et al., 2002).

1.1.3. Les technologies liées à la transcriptomique

Frederick Sanger développa la première méthode de séquençage de l'ADN en 1977 (Sanger et al., 1977), 24 ans après la découverte de sa structure en double hélice (Watson and Crick, 1953). Cette découverte recevra le prix Nobel trois ans plus tard et reste encore aujourd'hui la 4^{ème} publication scientifique la plus citée de tous les temps (selon le Science Citation Index). Enfin, le code génétique pouvait être décrypté.

1.1.3.1. Les débuts de la transcriptomique : EST & SAGE

La première technologie dérivée de la méthode Sanger a été le séquençage d'EST, *Expressed Sequenced Tag*, qui consistait au séquençage d'ADN complémentaire (ADNc) cloné dans des plasmides de bactéries (Adams et al., 1991; Putney et al., 1983). Rapidement, la méthode SAGE, Serial Analysis of Gene Expression, a permis de tendre vers une caractérisation à plus grande échelle de l'ADN, mais également à la quantification des gènes et transcrits d'intérêt (Velculescu et al., 1995). C'est à cette époque que la notion d'assemblage de transcrits est apparue, où les « tags » séquencés étaient concaténés en contigs. Bien que ces technologies figurent comme une ouverture vers la transcriptomique haut-débit, elles demeuraient très coûteuses en temps, en expériences et en main d'œuvre.

1.1.3.2. Les puces à ADN

La technologie des puces à ADN est la première à avoir réellement permis d'étudier à haut-débit le transcriptome, car elles permettent d'estimer simultanément la concentration de milliers de transcrits au sein d'un échantillon donné.

Elles reposent toutes sur la capacité que possède l'ADN simple brin (dénaturé) à s'hybrider spontanément avec une séquence d'ADN complémentaire. Bien que plusieurs technologies coexistent, il existe un principe commun : des fragments d'ADN, appelés sondes nucléotidiques, sont fixés sur une surface solide, composée de verre, de silicium ou de plastique. Selon la technologie utilisée, la synthèse des sondes et la fabrication de la puce ne suivront pas le même procédé.

Les puces à ADN « spottés » sont ainsi désignées car les sondes sont déposées sur la surface dans des microgouttelettes puis fixées au support. Ces sondes peuvent être des banques d'ADNc, ou des produits amplification d'ADN génomique. Cette méthode permet de personnaliser la banque de gènes interrogés par la puce, grâce à sa mise en place simplifiée en laboratoire. La technologie la plus démocratisée aujourd'hui est celle des puces à oligonucléotides synthétisées « *in situ* » par lithographie, sur les puces Affimetrix (Lipshutz et al., 1999), ou par impression « jet d'encre » (Hughes et al., 2001), sur les puces Agilent par exemple. Cette méthode de synthèse industrielle

permet d'optimiser la fiabilité et la reproductibilité des puces, ainsi qu'une plus grande densité des sondes.

Comme décrit précédemment, les puces permettent de quantifier l'expression relative d'un gène grâce à l'hybridation d'ADN dénaturé à sa sonde complémentaire. De fait, les échantillons d'ARN doivent subir une amplification puis une rétrotranscription avant d'obtenir leurs ADNc correspondants, puis sont classiquement marqués par un fluorochrome. Lorsque l'ADNc trouvera sa sonde correspondante, ils s'hybrideront et le signal fluorescent persistant après lavage de la puce sera capturé par un scanner haute résolution. Le niveau de fluorescence associé à chaque sonde déterminera le nombre de molécules d'ADN hybridées et donc, leur quantité relative (Heller, 2002; Nelson, 2001; Pozhitkov et al., 2007; Schena et al., 1995). Ainsi, la conception des sondes est cruciale pour la spécificité des hybridations, ce qui limite l'ensemble des gènes étudiés à ceux ayant déjà été caractérisés. Cette technologie ne permet donc pas d'identifier de nouveaux transcrits ou de nouveaux gènes. De plus, une hybridation non spécifique des sondes peut entraîner des erreurs dans la quantification (Pozhitkov et al., 2007)

Historiquement, l'expression d'un gène était évaluée par une sonde représentant une région de celui-ci, mais les dernières applications de la technologie permettent d'étudier la composition exonique des ARN (*exon arrays*), les polymorphismes mononucléotidiques (*single nucleotide polymorphisme*, SNP) (LaFramboise, 2009), les variations du nombre de copies d'un gène dans l'ADN (*comparative genomic hybridization array*, CGH array) (Shinawi and Cheung, 2008), ou les interactions protéines-ADN (ChIP-on-chip) (Aparicio et al., 2004 ; Bulyk, 2006). Le *Tiling array* est une technologie de puce permettant l'étude de nouveaux transcrits ou de nouveaux gènes. La différence avec les puces à ADN classiques réside dans la conception des sondes. Celles-ci ne correspondent plus à des séquences de gènes connus, mais à chaque segment de l'ADN d'un génome, qu'il soit décrit comme une région transcrite ou non (Mockler et al., 2005; Samanta et al., 2007; Yazaki et al., 2007). Cependant, une telle technologie est coûteuse, difficile à mettre en place et peu adaptée aux génomes eucaryotes complexes.

Même si le nombre d'articles référencés dans Pubmed utilisant la technologie des puces à ADN diminue depuis 2015, certainement du au développement d'autres technologies de transcriptomique, elle reste une manière rapide d'analyser l'expression de gènes au sein d'un échantillon, à un cout réduit.

1.1.3.3. Le séquençage à haut-débit et ultra-haut-débit

Comme décrit dans 1.1.2, le transcriptome est une composante centrale dans la régulation de l'activité biologique d'une cellule. En théorie, il serait donc nécessaire de connaître chaque transcrit pour aboutir à une compréhension exhaustive de la fonction d'une cellule. En pratique, le séquençage

de l'ARN (*RNA sequencing*, RNA-seq) est aujourd'hui la technologie la plus utilisée pour l'analyse de transcriptome à grande échelle (Morozova et al., 2009; Wang et al., 2009). Elle permet de s'affranchir de la connaissance *a priori* des gènes étudiés, en déterminant la séquence de chaque transcrite et son abondance dans un échantillon (voir « L'analyse bio-informatique des données de RNA-seq »). Bien que l'exhaustivité ne soit pas encore atteinte, le RNA-seq a sans nul doute révolutionné notre capacité à décrire la complexité du transcriptome au cours de la dernière décennie.

1.1.3.3.1. Seconde génération de séquenceurs

1.1.3.3.1.1. Le RNA-seq

Depuis 2009, plusieurs générations de séquenceurs se sont succédées, chacune apportant des améliorations notables, développées autour d'une méthodologie commune, le séquençage Sanger. Actuellement, le séquenceur Illumina HiSeq (Bentley et al., 2008) est la technologie la plus utilisée, son fonctionnement est décrit dans ce paragraphe.

Dans un échantillon, l'ensemble des ARN totaux est extrait, puis plusieurs protocoles d'enrichissement permettent de sélectionner la fraction d'intérêt. Afin d'améliorer considérablement la sensibilité du séquençage, il est nécessaire d'enrichir l'échantillon en ARN polyadénylés, ce qui permet de sélectionner les ARNm et quelques ARNnc (~40 %). Il est également possible d'effectuer une déplétion en ARN ribosomique (ARNr), qui représente environ 90 % des transcrits d'une cellule.

À la différence du premier protocole, celui-ci permet de conserver également les ARN non polyadénylés, comprenant l'ensemble des ARNnc et est donc privilégié lors de l'étude des ARNnc. Une sélection par la taille pourra orienter l'étude transcriptomique vers du smallRNA-seq, variante du séquençage classique permettant de caractériser les petits ARNnc, ou du RNA-seq « classique » pour étudier les grands ARN (>200 nucléotides), ARNm et ARNnc.

Le protocole de RNA-seq requiert une fragmentation de l'ARN en morceaux de taille aléatoire, puis une rétrotranscription, pour générer l'ADNc correspondant. À cette étape, il est possible d'utiliser un protocole permettant de conserver l'information sur le sens de transcription de l'ARN (séquençage brin spécifique). Le séquençage requiert alors la ligation d'adaptateurs permettant la reconnaissance de l'ADN à séquencer. Ces adaptateurs peuvent également servir de code-barres, spécifiques d'un échantillon donné et ainsi effectuer l'analyse de plusieurs d'entre eux au cours de la même procédure de séquençage. Enfin, l'ADN est amplifié par PCR (*Polymerase Chain Reaction*). Étant donné la taille variable des fragments d'ADNc, cette dernière étape peut introduire un biais en augmentant artificiellement la proportion de certains fragments au sein de l'échantillon.

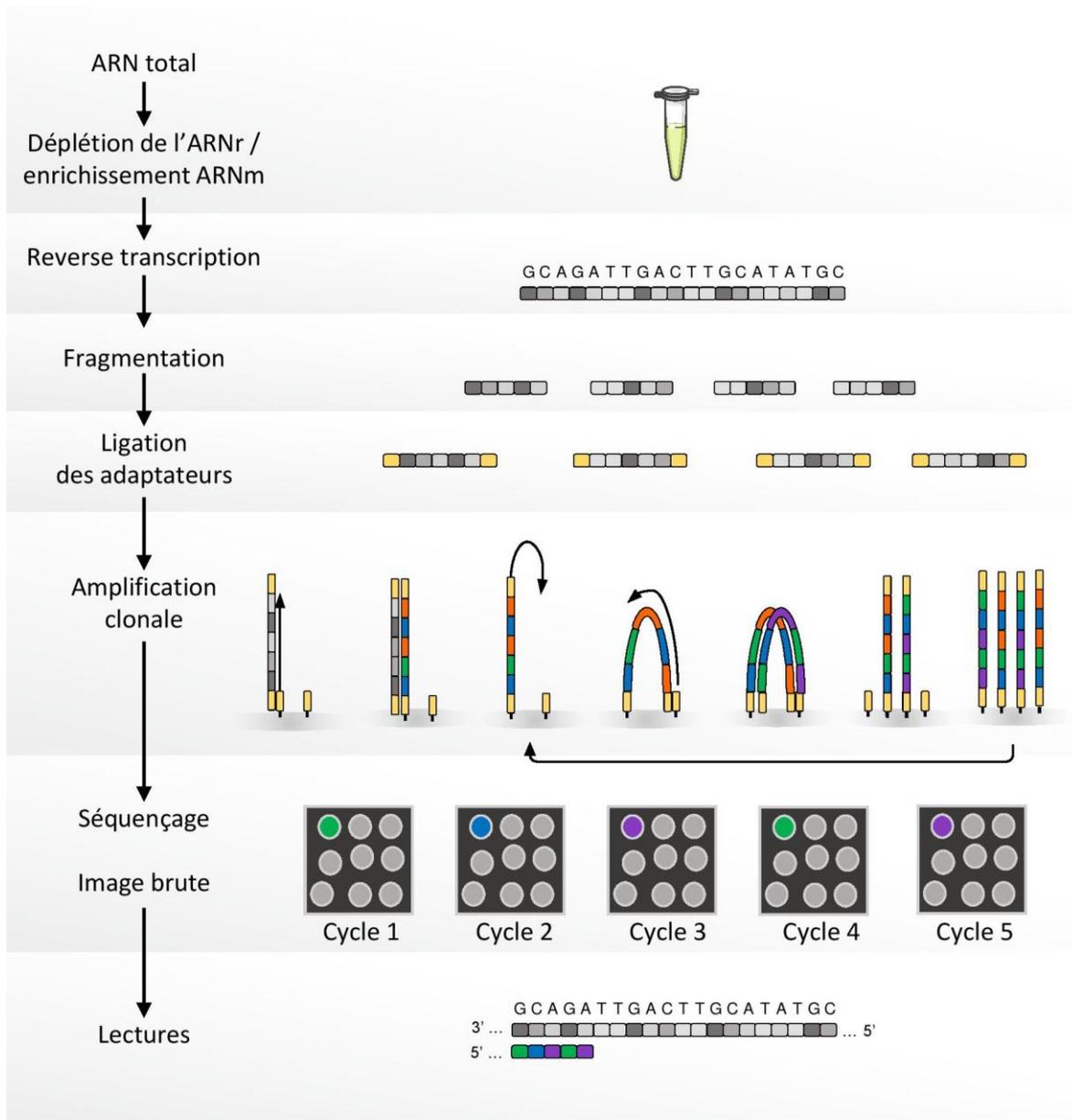


Figure 4. Illustration du protocole de RNA-seq sur la plateforme Illumina.

Les adaptateurs permettent la fixation aléatoire de chaque fragment sur la *flowcell* (support de séquençage). Une étape d'amplification permet d'augmenter le nombre de copies de chaque fragment et de créer des regroupements (généralement appelés *clusters*) d'une même molécule, ce qui optimisera la capture du signal de séquençage. La lecture de chaque fragment se fait base par base, de manière simultanée. À chaque cycle, des nucléotides couplés à des fluorochromes sont ajoutés sur la *flowcell* et lorsque l'amorce de séquençage incorporera le nucléotide complémentaire à la base en cours de lecture, la fluorescence émise permettra son identification. À chaque fin de cycle, les molécules fluorescentes sont lavées et de nouveaux nucléotides marqués sont ajoutés afin de commencer le cycle suivant. La capture d'image permet d'enregistrer la séquence des nucléotides incorporés au sein d'un *cluster*. Actuellement, il est possible d'effectuer 200 cycles subséquents, générant ainsi des lectures (courtes séquences, généralement appelées *reads*) d'une longueur maximum de 200 bases. La longueur des lectures dépend du protocole et de la technologie utilisée, mais plus la longueur de celle-ci est importante, plus les erreurs de séquençage sont susceptibles d'intervenir. Les fragments d'ADNc sont généralement plus longs, ce qui ne permet pas leur séquençage complet. Il est possible de séquencer un fragment d'ADNc à partir d'une seule ou des deux extrémités. Ainsi, deux séquences pairées sont obtenues pour un fragment, ce qui permettra de faciliter les futures analyses bio-informatiques, notamment l'alignement, l'assemblage et l'identification des différentes isoformes d'un même gène.

Le RNA-seq a permis des avancées spectaculaires dans le domaine de la transcriptomique. Depuis, une multitude de technologies haut-débit ont vu le jour et permettent de dépasser les verrous technologiques inhérents au séquençage classique. Ces approches sont complémentaires, car elles abordent l'étude du transcriptome de manière différente.

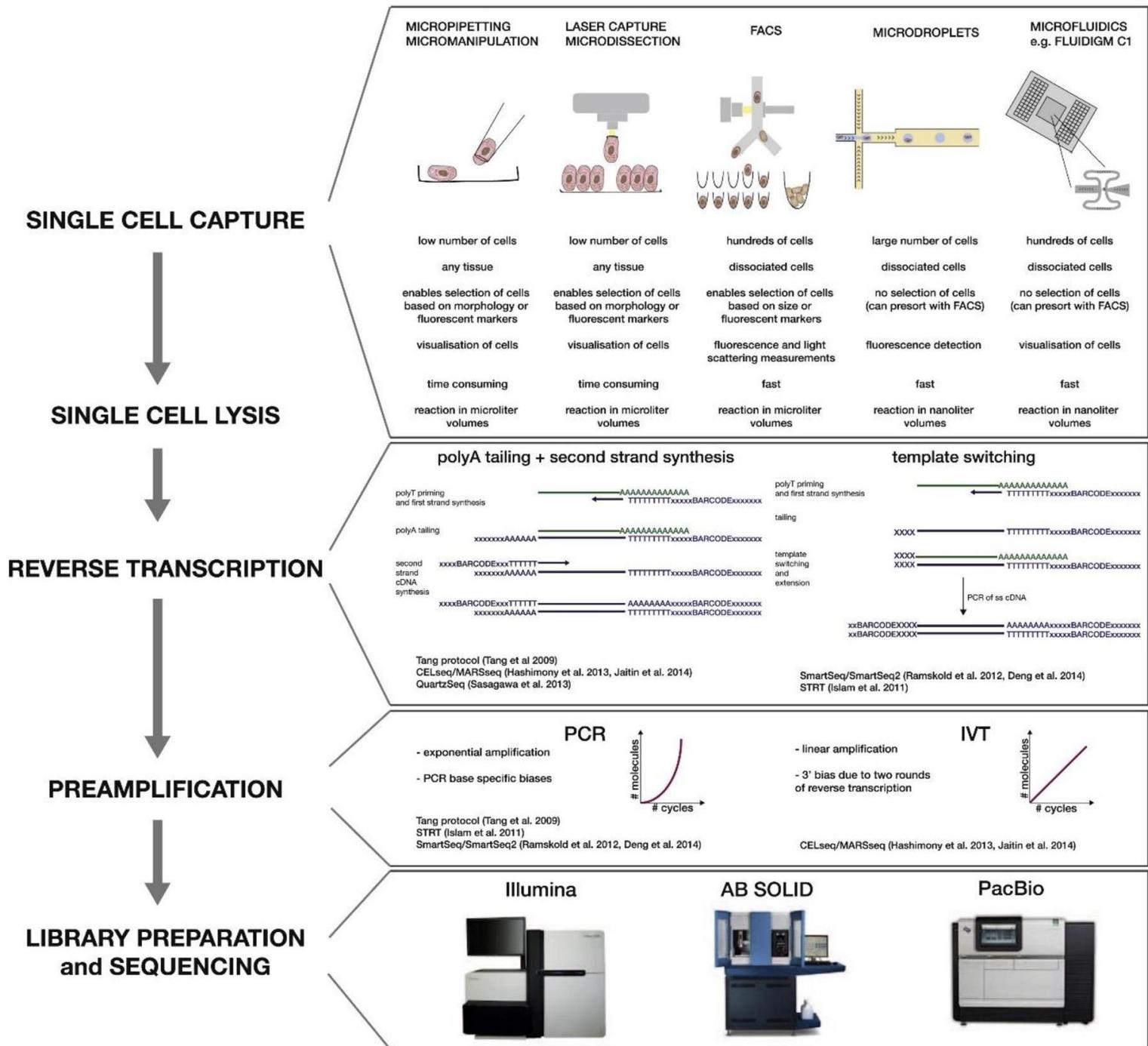


Figure 5. Protocole de scRNA-seq et description des différentes approches expérimentales disponibles à chaque étape (Kolodziejczyk et al., 2015).

1.1.3.3.1.2. RNA-seq à l'échelle de la cellule unique, scRNA-seq

Les études transcriptomiques à l'échelle d'un tissu, de cellules purifiées ou de lignées, ne permettent pas d'appréhender la plasticité du transcriptome à l'échelle de la cellule unique. Dès lors, il est difficile de caractériser de nouveaux types cellulaires, la dynamique du programme transcriptionnel lors de la différenciation de lignées, ou des populations hétérogènes, comme des tumeurs cancéreuses (Hedlund and Deng, 2018; Kolodziejczyk et al., 2015).

Durant la dernière décennie, des méthodes de capture de cellules uniques ont été développées puis adaptées au RNA-seq. La dissociation du tissu à analyser est nécessaire pour libérer les cellules de la matrice extracellulaire et de leur adhésion aux cellules avoisinantes. Plusieurs approches vont permettre l'isolement des cellules, telles que la microfluidique (Klein et al., 2015; Zilionis et al., 2016), les microgouttes (Macosko et al., 2015) ou des approches dérivées du FACS (*Fluorescence-activated cell sorting*) (Jaitin et al., 2014). Les étapes subséquentes de reverse transcription, de l'amplification de l'ADNc et de préparation de bibliothèques sont similaires au RNA-seq classique, à ceci près que chaque cellule est considérée comme un échantillon. De fait, la quantité de matériel génétique à séquencer est très limitée. La quantité de transcrits présente dans une seule cellule est estimée entre 100 000 et 300 000 molécules d'ARN et seule une fraction d'entre eux seront efficacement rétrotranscrits en ADNc (Marinov et al., 2014; Pollen et al., 2014). Comme pour le RNA-seq classique, l'amplification par PCR peut introduire un biais dans la quantification. Ces deux facteurs pris en compte, la sensibilité du séquençage scRNA-seq est limitée et ne peut être corrigée par une profondeur augmentée (c.-à-d. plus de lectures séquencées). Cependant, différents protocoles de séquençage adaptés ont vu le jour et sont répertoriés dans (Hedlund and Deng, 2018). Par exemple, les protocoles Smart-seq (Ramsköld et al., 2012) et Smart-seq2 (Picelli et al., 2013) permettent d'optimiser le séquençage à partir de quantité d'ARN limitée. Le Drop-seq est une méthode qui permet d'isoler jusqu'à 10 000 cellules d'un échantillon, chacune dans une gouttelette contenant les code-barres pour l'identification des cellules et des ARN, ainsi que les réactifs nécessaires à la transcription inverse (Macosko et al., 2015). L'autre challenge actuel est de développer les outils bio-informatiques adaptés pour surpasser les variabilités et biais introduits par les différentes étapes du séquençage et ainsi décrire précisément la dynamique réelle du transcriptome entre les cellules (Stegle et al., 2015).

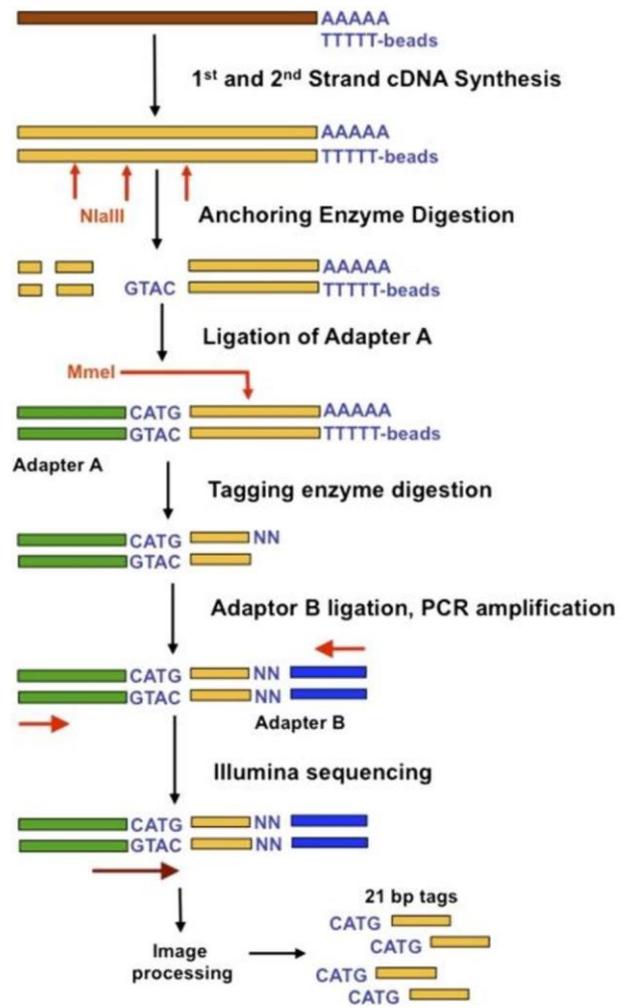


Figure 6. Principe du DGE-seq (adapté de Morrissy et al., 2009)

1.1.3.3.1.3. DGE-seq

Le DGE-seq, *Digital Gene Expression*, est une méthode de séquençage adaptée du SAGE (Asmann et al., 2009; Morrissy et al., 2009). Les ARN polyadénylés sont fixés sur des billes puis rétrotranscrits pour générer l'ADNc. Ils sont ensuite digérés par une enzyme agissant sur des sites de restriction fréquents, NlaIII. Seuls les fragments d'ADN fixés aux billes seront conservés pour la suite de la procédure. Des adaptateurs sont alors fixés sur cette extrémité 5', eux même spécifiques d'une seconde enzyme de restriction qui clivera molécules d'ADN en fragments de 21pb. Un second adaptateur est alors lié à l'extrémité 5'. L'ADN ainsi obtenu sera amplifié par PCR puis séquencé.

En raison de la nature du protocole, le sens de transcription de l'ARN d'origine est conservé et seulement l'extrémité 3' est séquencée. Cependant, la longueur réduite des lectures limite les erreurs inhérentes au séquençage. Cette nouvelle approche est plus économique que le RNA-seq classique (Chen et al., 2015) et permet donc d'augmenter le nombre de réplicats et d'échantillons analysés.

1.1.3.3.2. Troisième génération (PacBio, Nanopore)

Lors du RNA-seq classique, l'amplification des transcrits par PCR peut introduire un biais dans la quantification après le séquençage (Parekh et al., 2016). De plus, la fragmentation des transcrits implique un assemblage post-séquençage pour l'identification des isoformes, ce qui peut aboutir à des transcrits chimériques (Phillippy, 2017). Le développement d'outils bio-informatiques réduisant les artefacts d'assemblages reste un challenge, en particulier pour l'assemblage *de novo* chez les organismes ne disposant pas d'un génome ou d'un transcriptome de référence (voir « L'assemblage des transcrits »).

Les technologies de séquençage troisième génération ont pour vocation de surmonter ce type de limitations, pour rendre la caractérisation et la quantification des isoformes encore plus précises. Ces nouvelles méthodes ne nécessitent pas d'étapes d'amplification du matériel génétique à séquencer et permettent le séquençage de fragment significativement plus long, voire des transcrits entiers de plusieurs kilobases (Garalde et al., 2018; Laver et al., 2015; Lu et al., 2016; Rhoads and Au, 2015). Aujourd'hui, deux technologies onéreuses sont commercialisées : PacBio (Rhoads and Au, 2015) et Nanopore (Lu et al., 2016; Wanunu, 2012). La première utilise une molécule d'ADN polymérase fixée au fond d'un puits, qui synthétisera l'ADNc d'un brin d'ADN (ARN) natif en incorporant successivement l'un des quatre nucléotides marqués par un fluorochrome spécifique. La seconde technologie utilise les propriétés biochimiques spécifiques de chaque nucléotide pour détecter leur séquence. Une molécule d'ADNc est d'abord prise en charge par un adaptateur, qui guidera son passage à travers un pore, celui-ci détectera le changement électrique induit par chaque base.

Le taux d'erreurs de séquençage reste important (~10 %), mais les optimisations actuelles tendent vers une meilleure spécificité. À l'avenir, les différents transcrits pourront être plus précisément caractérisés et quantifiés. De plus, il apparaît que les modes de capture développés permettent également d'appréhender les modifications épigénétiques d'un nucléotide pendant le même processus (Schatz, 2017). Les futures technologies pourraient donc prétendre à des approches multi-omiques du séquençage.

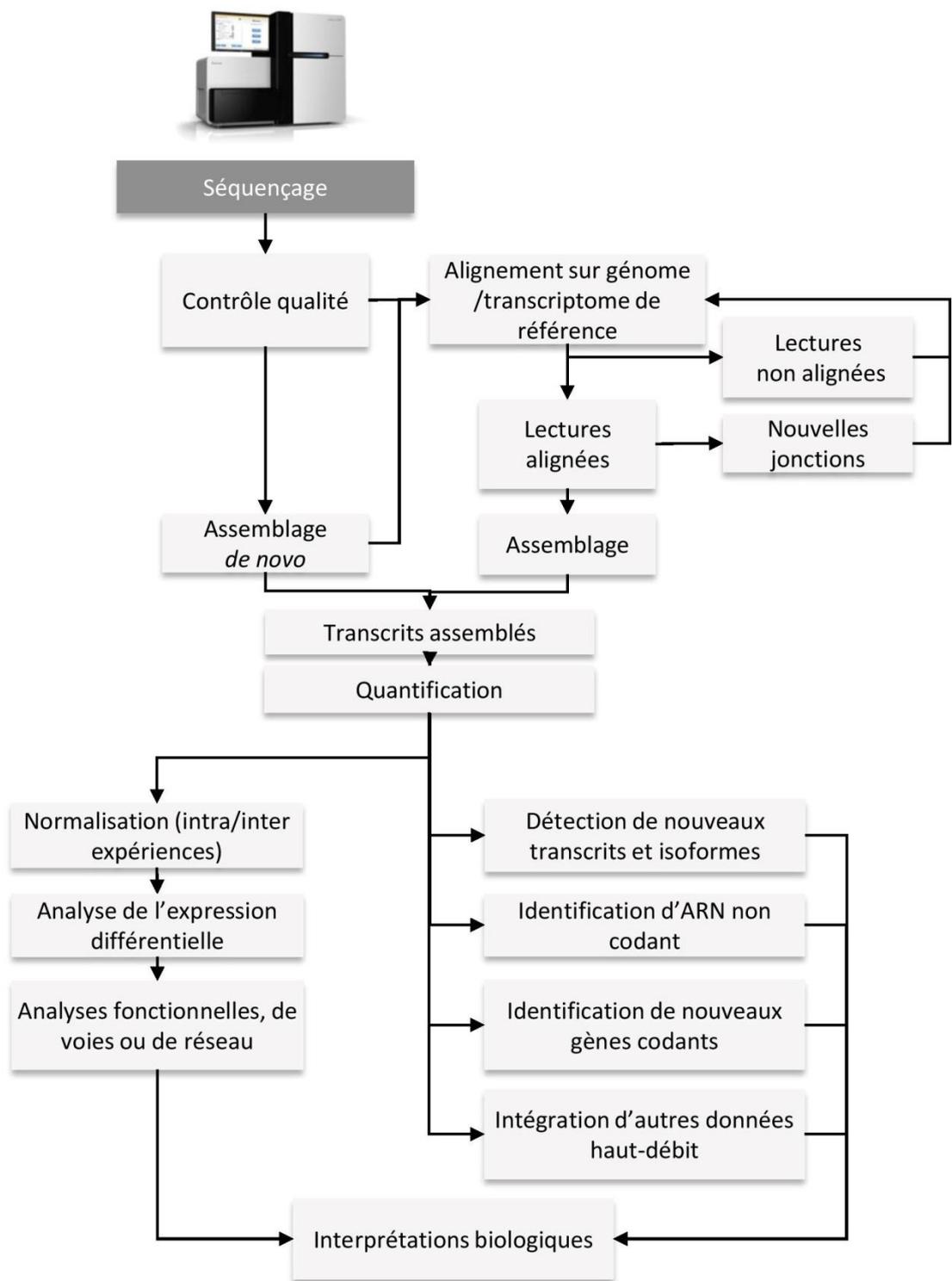


Figure 7. Étapes de l'analyse bio-informatique de données de RNA-seq

1.1.4. L'analyse bio-informatique des données de RNA-seq

De manière générale, tant au niveau du design expérimental, de la préparation des échantillons (Hardwick et al., 2017) que des protocoles de séquençage (Conesa et al., 2016; Han et al., 2015b; Hrdlickova et al., 2017), il existe de nombreux paramètres à optimiser pour que les données générées soient adaptées à la question biologique posée en amont (Head et al., 2014). Bien qu'il existe énormément d'outils bio-informatiques permettant l'analyse des données de transcriptomique, les protocoles mis en place doivent être pensés en fonction de l'objectif du projet. Seule l'analyse des données du RNA-seq sera abordée ici, car cette technologie a été utilisée dans ce projet de thèse (Figure 7).

1.1.4.1. Le contrôle qualité des données de séquençage

L'évaluation de la qualité des lectures (ou *reads*) générées par le séquençage comprend plusieurs paramètres, comme la qualité de la séquence, le contenu en GC, la présence d'adaptateurs, la surreprésentation de certains k-mer (segment de lecture) ou des lectures dupliquées. Ainsi, les erreurs de séquençage, les artefacts de PCR ou les contaminations peuvent être repérés. Des paramètres différents doivent être appliqués pour chaque expérience et doivent être considérés en fonction du protocole et de l'organisme étudié. Cependant, des résultats homogènes au sein des échantillons d'une même expérience sont attendus. FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) est un outil largement utilisé pour les données obtenues sur des plateformes Illumina, NGSQC pour données générées sur tout type de plateforme (Dai et al., 2010). Le contrôle qualité est une étape à part entière des différentes analyses bio-informatiques des données de RNA-seq.

1.1.4.2. L'alignement des lectures sur le génome

Les lectures séquencées sont généralement alignées sur un génome ou un transcriptome de référence lorsque ceux-ci sont disponibles. Cette étape permet d'identifier la ou les régions génomiques correspondant à chaque lecture. Tout d'abord, l'alignement de lectures provenant de RNA-seq sur le génome a la particularité de pouvoir aboutir à des alignements contenant de larges écarts (appelés *gaps*), causés par les jonctions d'épissage. En effet, les lectures correspondent à des fragments de l'ARN mature épissé, les introns sont donc éliminés. L'alignement sur le génome permet d'une part de détecter de nouvelles jonctions d'épissage, mais également de caractériser de nouveaux gènes potentiels, grâce à l'alignement de lectures dans des régions qui ne sont pas connues pour être transcrites. Plusieurs limitations peuvent interférer sur cette étape clé de l'analyse. Les pseudogènes *processed* (voir « Le génome et l'épigénome ») correspondent à des ARN matures intégrés au génome. L'alignement de lectures correspondant à l'ARN transcrit à partir du gène

original peut être de meilleure qualité sur son pseudogène, ce qui peut induire des erreurs d'identification. Certaines lectures peuvent également être multialignées si elles appartiennent à des séquences de faible complexité, dans des régions répétées, ou à des régions partagées entre plusieurs gènes paralogues.

On peut distinguer les outils bio-informatiques en fonction de l'approche algorithmique qu'ils utilisent. La méthode *exon-first* débute l'alignement sans considérer les jonctions d'épissage éventuelles. Les lectures alignées en un seul bloc définissent les régions exoniques. Les lectures non alignées seront analysées une nouvelle fois en considérant la présence éventuelle d'un *gap* dans leur séquence, définissant ainsi les jonctions d'épissage. Cette approche est utilisée par Tophat2 (Kim et al., 2013), MapSplice (Wang et al., 2010) et SOAPsplice (Huang et al., 2011). L'approche *seed-and-extend* aligne une portion de la lecture en un seul bloc, puis tente « d'allonger » l'alignement en considérant les *gaps* éventuels. L'algorithme derrière cette méthode utilise une table de hachage pour indexer la séquence des lectures, ce qui optimise l'alignement au niveau du temps et des ressources computationnelles (pour revue, Li and Homer, 2010). Les outils les plus récents utilisent cette approche, comme STAR (Dobin et al., 2013), HISAT et HISAT2 (Kim et al., 2015).

Suite à l'alignement, plusieurs paramètres peuvent être considérés pour évaluer la qualité des données. D'une part, le pourcentage de lectures alignées est un témoin de la qualité du séquençage, ou de potentielles contaminations biologiques. Chez les mammifères, la qualité est considérée comme satisfaisante lorsque plus de 70 % des lectures sont alignées sur le génome (Conesa et al., 2016). D'autre part, la couverture uniforme des régions séquencées permet d'apprécier la qualité des ARN de départ. Par exemple, dans le cas du séquençage d'ARN polyadénylés, l'accumulation de lectures alignées dans la partie 3' du transcrit peut être le témoin d'ARN partiellement dégradé dans l'échantillon de départ (Conesa et al., 2016).

1.1.4.3. L'assemblage des transcrits

Encore aujourd'hui, l'assemblage des lectures pour reconstituer le transcrit d'origine constitue un défi bio-informatique. Plusieurs outils utilisent les résultats de l'alignement effectué en amont pour optimiser l'assemblage et éviter la reconstruction de transcrits chimériques (Phillippy, 2017).

Par exemple, Cufflinks (Trapnell et al., 2012) construit un graphe de chevauchement en fonction des lectures alignées sur le génome. En parcourant ce graphe et ainsi assembler les lectures en transcrits, l'outil minimise le nombre de fragments reconstruits expliquant les lectures du graphe. StringTie utilise un graphe dans lequel les nœuds représentent les exons ou fragments d'exons (alignés en un bloc sur le génome) et les arrêtes représentent les variations d'épissage (Pertea et al., 2015).

L'assemblage *de novo* permet la reconstruction de transcrits sur la seule base des lectures séquencées et est donc particulièrement indiqué pour les études effectuées chez des espèces sans génome ou transcriptome de référence. Cette stratégie peut également être utilisée dans les espèces modèles pour identifier de nouvelles isoformes ou de nouveaux gènes. Sans guide, les outils comme Trinity (Grabherr et al., 2011) ou SOAPdenovo (Luo et al., 2012 ; Xie et al., 2014) utilisent le chevauchement des lectures entre elles pour reconstruire le transcrit d'origine. Ils utilisent le graphe de DeBruijn, où chaque nœud représente un k-mer (segment d'une lecture) et les arrêtes représentent leur chevauchement. En théorie, la séquence représentée par un chemin continu au sein du graphe correspond au transcrit d'origine, mais les erreurs de séquençage, les portions répétées et les variants d'épissage complexifient grandement la structure du graphe, ayant pour conséquence des erreurs d'assemblage.

Pour tous les outils existants, les protocoles de séquençage pairés permettent de réduire le nombre de reconstructions chimériques. Les transcrits ainsi reconstruits peuvent ensuite être utilisés comme transcriptome de référence pour l'alignement des lectures, optimisant la quantification à venir.

1.1.4.4. La quantification des transcrits et la normalisation

La quantification est une étape fondamentale de l'analyse et de nombreux outils sont disponibles (Anders et al., 2015; Li and Dewey, 2011; Liao et al., 2014; Pollier et al., 2013; Trapnell et al., 2010). Le niveau d'expression de chaque gène est approché par l'estimation de son abondance (nombre de copies par rapport au nombre total de transcrits), elle-même estimée par rapport au nombre de lectures alignées. Plusieurs paramètres rentrent en compte tels que la taille effective du transcrit/gène, la longueur des lectures et la distribution des fragments. L'alignement et la quantification sont donc des étapes d'analyse intimement liées, qui peuvent être effectuées en même temps si le projet ne nécessite que la quantification des gènes. Les unités de mesure les plus communément utilisées sont FPKM (*fragments per kilobase of exon model per million mapped reads*), adapté au séquençage pairé, RPKM (*reads per kilobase of exon model per million reads*) et TPM (*transcripts per million*). Elles représentent différentes manières d'appréhender l'abondance d'un transcrit selon les paramètres énoncés précédemment et de normaliser son expression afin de permettre les comparaisons intra- et inter-échantillons.

La normalisation permet de corriger les biais techniques systématiques et rendre les données comparables entre les échantillons (pour revue et comparaison, Li et al., 2015, 2017b; Lin et al., 2016). Les méthodes statistiques descriptives sont communément utilisées, comme la normalisation par quantile, qui corrige l'expression des transcrits suivant l'hypothèse que les comptages des lectures ont la même distribution dans tous les échantillons (Bolstad et al., 2003). D'autres méthodes

évaluent des facteurs d'échelle (scale factors) en prenant en compte la taille des bibliothèques de chaque échantillon (nombre total de lectures alignées), comme la normalisation par TMM (*Trimmed mean of M-values*) (Robinson and Oshlack, 2010). Ces approches peuvent être directement intégrées aux outils d'analyse de l'expression différentielle. Par exemple, le calcul de facteurs d'échelle fait partie intégrante de DESeq, un package R dédié à l'analyse différentielle des données de RNA-seq (Anders and Huber, 2010). D'autres approches utilisent des méthodes des algorithmes de machine learning pour estimer et normaliser l'abondance des transcrits, tels que RSEM (RNA-Seq by Expectation-Maximization) (Li and Dewey, 2011) et Sailfish (Patro et al., 2014). Cette dernière méthode est innovante, car elle ne nécessite pas d'alignement préalable des lectures sur une référence.

Il est souvent nécessaire d'appliquer un seuil d'expression pour éliminer les transcrits les moins exprimés, pouvant être générés par des erreurs de séquençage, d'alignement ou d'assemblage. Il n'existe pas de consensus pour définir ce seuil. Suite à la quantification et à la normalisation de l'expression des gènes, il est nécessaire d'évaluer la qualité des échantillons à l'échelle d'un projet. La reproductibilité des résultats entre les réplicats techniques et biologiques est évaluée grâce à un test de corrélation global ou une analyse en composante principale (ACP), pour ainsi obtenir une représentation de la variabilité entre les échantillons (Conesa et al., 2016). Il n'existe pas de standard pour valider ou non la reproductibilité des réplicats car celle-ci dépend fortement du matériel biologique étudié. Par exemple, le travail sur organes humains implique toujours plus de variabilité inter-individuelle que des études sur une lignée murine (Hardwick et al., 2017).

1.1.4.5. L'analyse différentielle

Les études transcriptomiques ont pour but de comparer les transcriptomes entre des tissus, des individus, des réponses environnementales (produits chimiques, stress...) ou lors du développement, afin de décrypter les mécanismes transcriptionnels les caractérisant. L'analyse différentielle de l'expression est conditionnée par le protocole expérimental et l'objectif du projet, il n'y a pas de consensus clairs sur les protocoles d'analyses optimales. Cependant, trois étapes essentielles sont requises : la normalisation (voir « La quantification des transcrits et la normalisation »), l'estimation de la variabilité et les tests statistiques.

Suite à la normalisation, l'expression des transcrits peut être comparée entre les échantillons. L'une des méthodes de base pour mesurer la variation d'expression est le calcul de ratio (ou *fold-change*), qui représente le rapport de deux abondances. Comme pour la normalisation, plusieurs méthodes sont implémentées dans les outils utilisés, comme DESeq (Anders and Huber, 2010) et edgeR (Robinson et al., 2010) et sont répertoriées dans (Fang et al., 2012; Huang et al., 2015; Kvam et al., 2012; Wu and Wu, 2016).

Enfin, il est nécessaire d'évaluer la variabilité de l'expression d'un transcrit entre les réplicats, surtout lorsque leur nombre est limité (Law et al., 2014; Robles et al., 2012; Seyednasrollah et al., 2015). Des outils, tel que Limma (Ritchie et al., 2015) ou Cuffdiff (Trapnell et al., 2012), vont permettre d'évaluer la variabilité de l'expression d'un transcrit au sein des réplicats biologiques d'une condition expérimentale, mais également au sein de l'ensemble des échantillons et écarteront de la suite de l'analyse ceux démontrant une trop grande variabilité au sein des données.

Le choix de la méthode de classification des données (*clustering*) est fondamental pour l'interprétation biologique. Le clustering permet de réduire la complexité des données en décrivant des groupes de transcrits suivant une dynamique d'expression similaire entre les échantillons. Parmi les approches traditionnelles, les méthodes hiérarchiques définissent les groupes de gènes grâce à un calcul de distance de leurs profils d'expression, comme la distance euclidienne (revues dans Oyelade et al., 2016). Les méthodes de partitionnement, dont la méthode *k-means*, consiste en la fusion des groupes dont les centroïdes sont les plus proches. Il existe de nombreuses méthodes de classification (revues dans Oyelade et al., 2016), mais aucun algorithme de partitionnement n'est optimal pour toutes les structures de données (Pirim et al., 2012). Le choix de la stratégie doit prendre en compte la normalisation des données, leur distribution et le protocole expérimental (données statiques, étude temporelle).

1.1.4.6. La caractérisation de nouvelles isoformes et de nouveaux gènes

Comme évoqué précédemment (voir « Les longs ARN non codants »), le statut non codant de certains ARNinc a pu être remis en cause grâce à la découverte de gènes codant pour de courts peptides (Frith et al., 2006 ; Hanada et al., 2007 ; Kastenmayer et al., 2006 ; Oyama et al., 2004).

Plusieurs outils ont été mis en place pour prédire le potentiel codant d'une séquence nucléotidique et de nombreuses approches sont disponibles. Par exemple, Coding-Potential Assessment Tool (CPAT) (Wang et al., 2013), Coding Potential Calculator (CPC) (Kong et al., 2007), Coding-Non-Coding Index (CNCI) (Sun et al., 2013) et COME (Hu et al., 2017) utilisent les caractéristiques intrinsèques de la séquence nucléotidique pour modéliser son potentiel codant. D'autres tentent de prédire la présence de domaines protéiques, comme PfamScan (Mistry et al., 2007) ou HMMER (Finn et al., 2011), ou étudient la conservation des séquences potentiellement codantes entre les espèces, comme phyloCSF (Lin et al., 2011).

En complément de l'approche prédictive, il est possible d'intégrer les données de RNA-seq à une analyse protéomique. La protéomique informée par la transcriptomique (*Proteomics informed by Transcriptomics, PIT*) est une méthode qui permet d'associer des peptides détectés par spectrométrie de masse avec les transcrits séquencés (Evans et al., 2012). Brièvement, les transcrits reconstruits sont traduits en séquences peptidiques potentiels dans les trois cadres de lectures

potentiels et sont ensuite intégrés aux protéomes de référence, Uniprot (Pundir et al., 2015) et Ensembl (Yates et al., 2016). Les données de spectrométrie de masse vont ensuite être analysées et les peptides identifiés grâce à cette nouvelle référence enrichie. Ainsi, certains peptides détectés expérimentalement peuvent être identifiés comme des peptides codés potentiellement par des nouveaux transcrits assemblés.

Ainsi, la conjugaison de ces deux approches permet d'identifier de potentiels nouveaux gènes codants, soit précédemment annotés comme non codants, soit de nouveaux gènes non annotés (Chocu et al., 2014).

1.1.4.7. L'analyse fonctionnelle

L'analyse fonctionnelle des groupes de gènes co-exprimés va permettre d'associer une dynamique transcriptionnelle avec un processus biologique particulier. Chaque gène connu est annoté avec les processus biologiques, les fonctions moléculaires ou les voies moléculaires dans lesquels ses produits sont impliqués. Ces informations sont disponibles dans des bases de données telles que la *Gene Ontology* (GO) (Gene Ontology Consortium, 2015) ou la *Kyoto encyclopedia of genes and genomes* (KEGG pathways) (Kanehisa et al., 2016). Globalement, des méthodes d'enrichissement évaluent la surreprésentation (ou sous-représentation) des gènes associés à une même annotation dans un groupe par rapport à l'ensemble des gènes caractérisés dans l'étude transcriptomique. Parmi les nombreux outils disponibles (revus dans Khatri et al., 2012), DAVID (Huang et al., 2007) est le plus utilisé aujourd'hui, car il intègre 14 bases d'annotations différentes (Sherman et al., 2007), mais qui ne sont pas mises à jour. L'association des partitions de gènes co-exprimés à des processus biologiques permet d'inférer ces fonctions à des gènes non annotés.

1.1.4.8. L'analyse de réseaux

L'analyse différentielle et la classification des transcrits ne permettent pas de modéliser la dépendance au sein du transcriptome, comme l'effet des facteurs de transcription sur l'expression de leurs cibles. La représentation par graphe et l'analyse de réseau permettent de modéliser de telles interactions, les nœuds représentant généralement les transcrits ou les gènes.

L'information portée par les arrêtes peut varier. D'une part, elles peuvent représenter une corrélation (positive ou négative) significative entre les profils d'expression de deux gènes (Butte et al., 2000; De la Fuente et al., 2004). La méthode Weighted Correlation Network Analysis (WGCNA) est basée sur le feuillage de la matrice de covariance empirique (Zhang and Horvath, 2005). Des méthodes basées sur le calcul de l'information mutuelle ont également été proposées (Basso et al., 2005; Margolin et al., 2006). Cependant, une corrélation très élevée entre le niveau d'expression de deux gènes n'implique pas un lien direct entre les deux gènes (Ballouz et al., 2015). Il est également

possible d'intégrer des données d'interactions contenues dans les bases de données pour reconstruire le graphe, en prenant en compte les régulations entre les facteurs de transcription et leurs cibles connues, dans TRRUST (Han et al., 2015a) ou the Transcription Factor Encyclopedia (Yusuf et al., 2012), ou des données d'interaction contenues dans MIntAct (Orchard et al., 2014). L'analyse topologique du graphe permet notamment de détecter des modules de gènes très connectés ou des nœuds centraux, permettant ainsi de mettre en évidence nouveaux gènes d'intérêt (Li et al., 2016; Serin et al., 2016).

1.1.4.9. Intégration de données haut-débit : la multi-omique

L'analyse intégrative des données de transcriptomique avec d'autres données -omiques permet d'apprécier la régulation de l'expression des gènes en fonction des autres aspects moléculaires tels que la génomique ou l'épigénomique.

1.1.4.9.1. RNA-seq et génomique

L'intégration de séquençage d'ADN avec des données de RNA-seq peut permettre la découverte de polymorphismes mononucléotidiques (single nucleotide polymorphism, SNP) ou l'expression de loci de caractère quantitatif (expression quantitative trait loci, eQTL). L'étude des eQTL nécessite les profils transcriptomiques et génotypiques d'un tissu prélevé chez un grand nombre d'individus. La corrélation entre le niveau d'expression d'un gène et son génotype permet d'associer des variations génétiques à des caractères complexes comme la taille (Giambartolomei et al., 2014), la susceptibilité à certaines pathologies (Moffatt et al., 2007), ou même les régulations induites par l'architecture du génome sur son expression (Gaffney, 2013; Gilad et al., 2008). Il a également été démontré que les variations génétiques induisent des modifications dans l'expression des gènes (Battle et al., 2014; Lappalainen et al., 2013; Montgomery et al., 2010; Pickrell et al., 2010).

1.1.4.9.2. RNA-seq et épigénomique

L'intégration de données d'épigénétique avec la transcriptomique permet d'apprécier les conséquences de la méthylation de l'ADN ou de l'état de la chromatine sur l'expression des gènes.

D'une part, il s'agit d'évaluer la corrélation entre les profils de méthylation de régions génomiques avec la dynamique d'expression des gènes associés (Kim et al., 2011; Li et al., 2013a; Xie et al., 2011). Différentes approches permettent de définir l'ensemble des gènes dont la méthylation et l'expression ont des profils comparables (Jiao et al., 2014). D'autre part, la technologie de ChIP-seq (chromatin immunoprecipitation sequencing) permet de séquencer spécifiquement les régions génomiques sur lesquelles est fixée une protéine d'intérêt. Cette

technologie peut être utilisée pour cibler les marques d’histones de la chromatine. Ainsi, il est possible d’associer l’état de la chromatine et l’expression des gènes associés à ces régions (Kundaje et al., 2015; Madrigal and Krajewski, 2015) (revue dans (Angelini and Costa, 2014).

1.1.4.9.3. RNA-seq et smallRNA-seq

Comme énoncé précédemment (voir « Les petits ARN non codants »), les ARNnc remplissent de nombreuses fonctions de régulation de l’ARN, parfois en interagissant directement avec les transcrits qu’ils ciblent. De plus, les ARNlnc peuvent être des précurseurs pour les miARN. De fait, il est intéressant d’utiliser une approche intégrative des données de RNA-seq et de smallRNA-seq pour décrypter les mécanismes de régulation en jeu. Plusieurs outils sont disponibles pour prédire les cibles des miARN, comme mirWalk (Dweep et al., 2011) et miRBase (Kozomara and Griffiths-Jones, 2014). Ces prédictions sont raffinées par d’autres outils intégrant les données de smallRNA-seq et de RNA-seq en évaluant l’anti-corrélation des profils d’expression des miARN et leurs cibles prédites (Nam et al., 2009; Sales et al., 2010; Wu and Watson, 2009).

1.1.5. Les bases de données et référentiels en transcriptomique

1.1.5.1. Les banques de données généralistes

1.1.5.1.1. Les banques de séquences nucléotidiques

Plusieurs centres et instituts ont mis en place des bases de données permettant d’avoir accès à de nombreuses informations relatives aux gènes.

Les bases de données maintenues par le *National Cancer for Biology Information* (NCBI) sont parmi les plus utilisées de nos jours. EntrezGene (Brown et al., 2015) répertorie environ 17 millions de gènes (Mars 2017) et accumule de nombreuses informations sur leur séquence (RefSeq, Pruitt et al., 2014) et leurs annotations, telles que leur nom et symbole, leur structure intronique et exonique, les isoformes de transcrits... Le lien avec d’autres bases de données permet d’agréger sous un identifiant unique, l’EntrezGene ID, les informations relatives à leur homologie (HomoloGene, <https://www.ncbi.nlm.nih.gov/homologene/>) (NCBI Resource Coordinators, 2016), leur annotation avec des processus biologiques particuliers (Gene Ontology, The Gene Ontology Consortium, 2017), leur bibliographie associée (PUBMED, <https://www.ncbi.nlm.nih.gov/pubmed/>) et bien d’autres données. Avec la même volonté d’agréger, d’uniformiser et d’intégrer les données génomiques issues de différentes sources, l’*European Bioinformatics Institute* et le *Wellcome Trust Sanger Institute* ont mis en place une base de données centrée sur les gènes, Ensembl (Zerbino et al., 2018). La base de données de l’UCSC (*University of California Santa Cruz*) (<https://genome.ucsc.edu/index.html>) met à disposition des nombreuses ressources génomiques et outils bio-informatiques. Parmi ces

ressources, les séquences et annotations du génome de plus d'une centaine d'espèces et de génomes sont disponibles. L'outil de visualisation des génomes de l'UCSC est l'un des plus utilisés aujourd'hui, car il permet de visualiser simplement les données de génomique (structure, variants...), de transcriptomiques (expression...) et d'épigénomique (méthylation...) d'une région donnée.

Chaque base de données a développé sa propre procédure d'identification et d'annotation, ce qui engendre des différences significatives dans leur contenu. Une étude comparative des résultats de l'analyse de données de RNA-seq a mis en évidence l'importance de l'annotation du génome ou transcriptome de référence lors de l'alignement ou de la quantification des transcrits (Zhao and Zhang, 2015).

1.1.5.1.2. Les banques de séquences protéiques

Le consortium Uniprot a mis en place une des bases de données contenant les séquences et annotations des protéines (Bateman et al., 2017). D'une part, on retrouve une banque de données non redondante, agrégée grâce à une curation manuelle révisée, SwissProt et d'autre part, un ensemble de données issues d'une curation automatique non révisée, TrEMBL. La *Protein Data Bank* (PDB) contient des données sur la structure tridimensionnelle des protéines, issues d'analyse par cristallographie (<http://www.rcsb.org/pdb/>, Berman et al., 2000).

1.1.5.1.3. Les banques de données génomiques

Avec la démocratisation des technologies haut-débit de génomique, épigénomique, transcriptomique et protéomique, la communauté scientifique a mis en place des espaces de dépôts permettant aux scientifiques de partager leurs données. Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) (Barrett et al., 2012), ArrayExpress (Kolesnikov et al., 2015) et Sequence Read Archive (SRA) (Leinonen et al., 2011) sont des espaces de dépôt dédiés aux données brutes issues de la génomique (principalement des données de puces à ADN et de séquençage).

1.1.5.2. Les banques de données dédiées à la reproduction

Des bases de données dédiées au domaine de la reproduction ont également été mises en place. Certaines d'entre elles ont agrégé les connaissances accumulées au fil des études grâce à la curation manuelle ou des méthodes d'analyse de textes d'articles publiés. En analysant 43 000 articles publiés avant mars 2014, **FollicleOnline** décrit 580 gènes impliqués dans la folliculogénèse de nombreuses espèces, ainsi que les données expérimentales associées (Hua et al., 2015). Utilisant une approche similaire, **SpermatogenesisOnline** met en lumière 1666 gènes participant à la spermatogénèse (Zhang et al., 2013). L'outil intègre également un modèle de prédiction basé sur les résultats de la curation manuelle, des données de transcriptome de testicule adulte de souris et des

données d'interactions protéine-protéine et ainsi prédit l'implication de 762 gènes candidats dans la spermatogenèse.

Dès l'émergence des technologies telle que le SAGE, certains outils ont été mis en place pour valoriser et partager les résultats. **GonadSAGE** est une banque de données de SAGE comparant six temps de développement de gonades chez la souris. Elles peuvent être visualisées dans le navigateur de génome de l'UCSC (Lee et al., 2010). **GermSAGE** rend disponible des données de SAGE obtenues dans des cellules germinales souris : spermatogonie A, spermatoctes pachytènes et spermatides ronds (Lee et al., 2009).

En ce qui concerne les données de transcriptomique issues des puces à ADN, the GenitoUrinary Development Molecular Anatomy Project (**GUDMAP**) est un consortium international qui a mis en place l'agrégation de données variées de transcriptomique dans tout l'appareil urogénital de souris (Harding et al., 2011). Il regroupe notamment des données de transcriptomique d'organes, de cellules micro-disséquées ou triées par FACS de tout le système urogénital en développement, y compris testicule et ovaire (Harding et al., 2011). **L'Ovarian Kaleidoscope database (OKdb)** combine les méthodes d'analyse de texte et les résultats de puces de l'ovaire et de ses cellules à différents stades de développement (Hsueh and Rauch, 2012). En intégrant les annotations disponibles, l'accès est facilité aux diverses informations des gènes d'intérêt, sur les processus biologiques ou les voies dans lesquels il est impliqué, aux données d'expression des gènes à travers les diverses études transcriptomiques intégrées et les phénotypes auxquels il peut être associé. Ce type de ressource permet d'aiguiller les chercheurs dans leurs recherches. En ce qui concerne la lignée germinale, **GermOnline** (Lardenois et al., 2010) est un outil de génomique comparative, car il intègre des données de différentes espèces, se concentrant sur le développement de la lignée germinale, la méiose et la gamétogenèse.

Comme évoqué précédemment, le séquençage à ultra-haut-débit a permis d'élargir les classes d'ARNs caractérisés dans divers échantillons biologiques. **GermlncRNA** est une base de données spécifique des types cellulaires de la lignée germinale (Luk et al., 2015). En intégrant les résultats de plusieurs études de transcriptomiques (Gan et al., 2013; Lee et al., 2009, 2012; Soumillon et al., 2013; Sun et al., 2013a), il a été possible de créer un catalogue des ARNlnc recensés dans les spermatogonies A, les spermatoctes et spermatides. Enfin, **Spermbase** est une ressource dédiée au transcriptome du sperme humain, rat, souris et lapin (Schuster et al., 2016).

Pour finir, le Reprogenomic Viewer (**RGV**) est une base de données qui permet de visualiser et de comparer de nombreuses études de séquençage haut-débit - en génomique, transcriptomique et épigénomique - dans le domaine de la reproduction (Darde et al., 2015). Il met à disposition de la communauté un navigateur de génome qui permet d'explorer les différents jeux de données intégrés et permet des comparaisons interspèces des résultats.

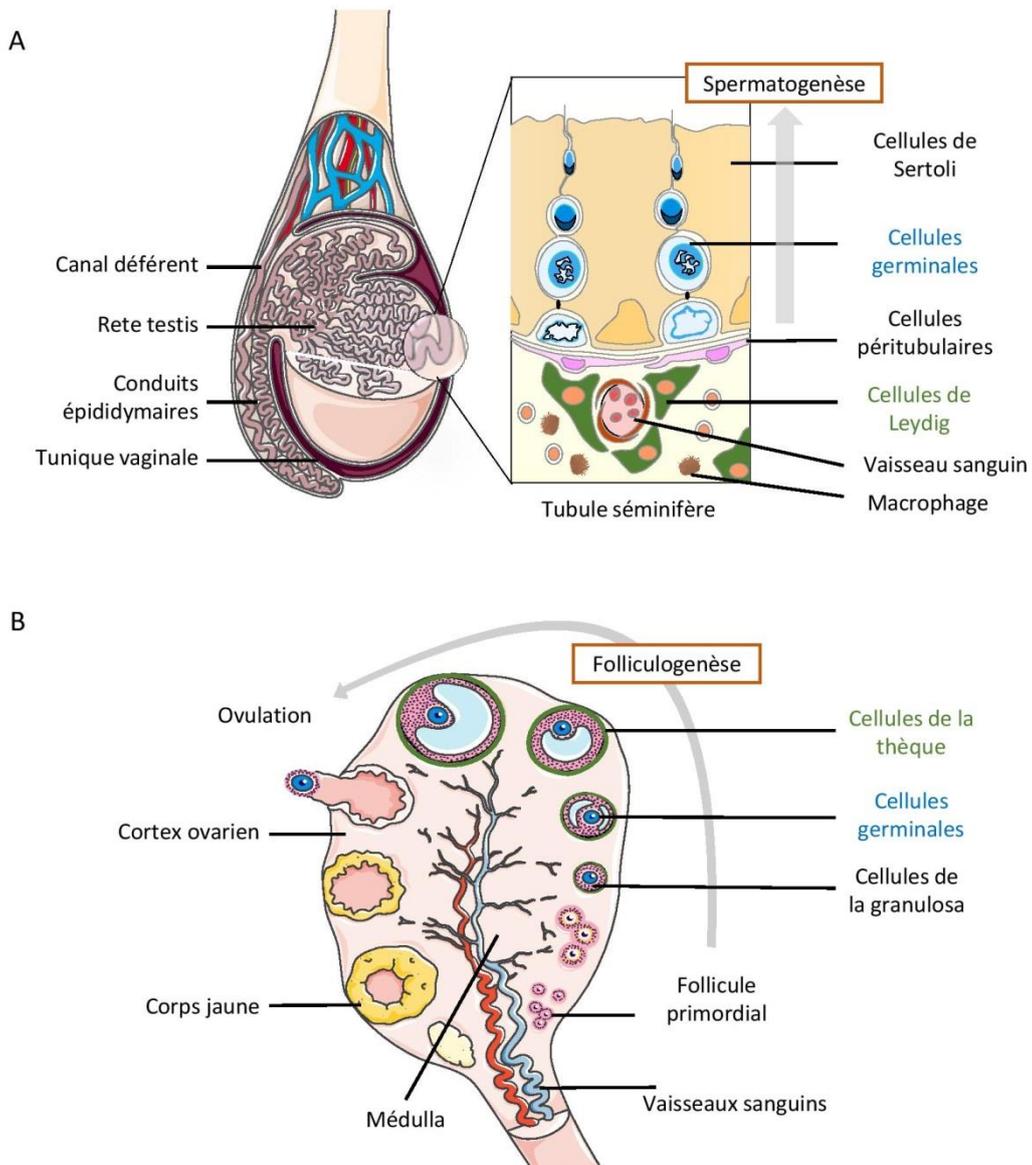


Figure 8. Anatomie du testicule (A) et de l'ovaire adulte (B).

1.2. La différenciation gonadique chez les mammifères

1.2.1. Les gonades adultes et fonctions de reproduction

Le testicule comme l’ovaire assument des fonctions analogues de production de gamètes et d’hormones. Cependant, les stratégies mises en place par les deux sexes pour assurer la reproduction de l’espèce sont fondamentalement différentes, voire opposées, mais toujours complémentaires.

Les populations cellulaires formant les deux gonades sont spécifiques de chacune, mais présentent des analogies de fonction qui permet de les catégoriser. La lignée germinale correspond aux cellules en différenciation aboutissant à la production de gamètes, spermatozoïdes ou ovocytes. Les cellules supportrices sont les cellules qui permettent le bon déroulement de la différenciation des gamètes grâce à leur sécrétion de facteurs paracrines, les cellules de Sertoli chez le mâle, ou les cellules de la granulosa chez la femelle. Enfin, les cellules stéroïdogènes vont assurer la production d’hormones au sein de la gonade, les cellules de Leydig dans le testicule ou les cellules de la thèque et de la granulosa dans l’ovaire.

1.2.1.1. Structures des gonades adultes

Le testicule adulte est composé de deux compartiments : les tubules séminifères et l’espace interstitiel. Les tubules séminifères s’enchevêtrent de façon compacte et se regroupent au niveau du rete testis. Un ensemble de nombreux tubules forme un lobule testiculaire, que l’on retrouve au nombre de 250 à 300 dans le testicule adulte (de Kretser et al., 2018). Les tubules séminifères sont constitués de trois types cellulaires : les cellules de Sertoli, les cellules germinales, interagissant avec les cellules de Sertoli et se différenciant et les cellules péri-tubulaires, délimitant les tubules séminifères de l’espace interstitiel (Figure 8A).

Les coupes histologiques d’ovaire adulte montrent également deux compartiments tissulaires. La médulla constitue la zone centrale de l’ovaire, constituée d’un tissu conjonctif lâche, très vascularisé et innervé. En périphérie de l’organe se trouve le cortex, constitué d’un tissu conjonctif riche en fibroblastes, le stroma ovarien, dans lequel sont inclus les follicules de la réserve ovarienne et les follicules en croissance. L’enveloppe ovarienne est formée par l’albuginée et un épithélium de surface (Bahr, 2018) (Figure 8B).

La première différence entre le testicule et l’ovaire est que la méiose des cellules germinales femelles commence pendant la vie fœtale, alors que celle-ci ne débute qu’à la puberté chez l’homme. Des plus, l’ovaire subira de constantes transformations au cours de la vie de la femme adulte. En effet, la folliculogénèse désigne le recrutement d’un follicule primordial et son développement jusqu’à l’ovulation ou à l’involution. Un ovaire est donc constitué de 90 à 95 % follicules

primordiaux quiescents et 5 à 10 % de follicules activés (Baerwald et al., 2012). Ce processus de maturation dure environ une année et 99 % des follicules activés subiront une atrophie au cours de leur développement. Au cours de la vie d'une femme, seulement 400 follicules aboutiront à l'ovulation. L'ovaire est donc un organe dynamique, qui subira des vagues successives d'activation au cours de sa vie.

1.2.1.2. La fonction exocrine des gonades adultes

Le testicule et l'ovaire sont fondamentalement différents au niveau de la stratégie de production de gamètes, le premier produisant une grande quantité de spermatozoïdes tout au long de la vie et le deuxième établissant un stock limité au cours de son développement, qui s'épuisera au fil de la vie de la femme.

La spermatogenèse est définie comme l'ensemble des transformations aboutissant à la formation d'un spermatozoïde haploïde mature à partir d'une spermatogonie et a lieu dans les tubules séminifères. Chez l'Homme, la puberté marque le début de la production de spermatozoïdes, qui augmente jusqu'à un pic d'environ 250 millions/jour à 20 ans, puis diminue progressivement (Sharpe, 2012). À 70 ans, le nombre moyen de spermatozoïdes produits par jour est d'environ 50 millions, mais les variabilités interindividuelles sont très importantes, des cas d'azoospermie étant détectés à tout âge (~50 % des hommes de 80 ans). Cette production intensive de spermatozoïdes est possible grâce au soutien de cellules de Sertoli, qui fournissent les facteurs biochimiques indispensables à chaque étape de différenciation et de maturation des cellules germinales. De plus, les spermatogonies ont la capacité de s'engager vers deux voies de développement : l'une vers la spermatogenèse, donc la production de gamètes, l'autre vers leur propre renouvellement. Aujourd'hui, il n'existe pas de consensus au sein de la communauté scientifique pour décrire les mécanismes permettant une capacité aussi robuste de persistance à travers la vie de l'homme (Boitani et al., 2016; Lord and Oatley, 2017).

La stratégie ovarienne est différente de celle du testicule. L'ovogenèse débute pendant la vie fœtale. À noter qu'il faut distinguer l'ovogenèse, correspondant aux transformations qui permettront la production un ovocyte fécondable à partir des précurseurs germinaux, de la folliculogénèse, qui désignent la maturation du follicule primordial jusqu'à l'excrétion de l'ovocyte et à la formation d'un corps jaune. En effet, comme décrit dans le chapitre dédié au « Développement ovarien », les ovocytes entrent en phase de méiose à partir de la 13^{ème} semaine de développement (SD, équivalent au nombre de semaines de gestation), de manière asynchrone (Carninci et al., 2005; Motta, 1997). Ils subissent alors un arrêt méiotique à partir de la 17^{ème} SD et sont individualisés dans un follicule primordial. Le cycle méiotique ne reprendra que lors de l'ovulation. À la naissance, l'ovaire dispose d'un stock d'environ 1 million de follicules primordiaux, susceptibles d'être activés et de conduire à

l'ovulation d'un ovocyte. Ce stock sera réduit à 300 000 follicules au début de la puberté. À environ 37 ans, la femme dispose toujours de 25 000 follicules primordiaux, mais la vitesse de dégénérescence s'accélère jusqu'à l'épuisement du stock, ayant pour conséquence la mise en place de la ménopause (Hansen et al., 2008).

1.2.1.3. La fonction endocrine des gonades adultes

1.2.1.3.1. La stéroïdogénèse

Dans le testicule, le taux de testostérone est directement lié au nombre des cellules de Sertoli et à l'activité des cellules de Leydig. L'étude de l'évolution du taux plasmatique au cours de la vie a pu mettre en évidence trois vagues de différenciation de cellules Leydig (Griswold and Behringer, 2009; Wainwright and Wilhelm, 2010). La première vague a lieu pendant la vie fœtale, produisant des cellules de Leydig fœtales (voir chapitre « Cellules de Leydig »). Ensuite, 2 à 3 mois suivant la naissance, les cellules de Leydig se différencient pendant la « mini-puberté » (Prince 2001). Cette vague semble être induite par la réactivation de l'axe hypothalamo-hypophysio-gonadique, qui conduit à la production de LH circulant, associée à une augmentation du niveau de testostérone (Mann and Fraser 1996). Le nombre des cellules de Leydig immatures va décroître progressivement jusqu'à la troisième vague de différenciation à la puberté (Svechnikov et al., 2010). L'origine et le devenir des différentes populations de cellules de Leydig sont encore débattus.

La stéroïdogénèse du testicule permet la synthèse de testostérone et de dihydrotestostérone (DHT) à partir de différentes sources de cholestérol par les cellules de Leydig et est décrite dans le chapitre « Synthèse des stéroïdes ». Une quantité limitée de testostérone peut être convertie en œstradiol par l'aromatase dans les cellules de Sertoli, mais le testicule n'est qu'une source mineure d'œstrogène circulant (Li et al., 2018). Au sein de la gonade, la testostérone peut se fixer aux récepteurs aux androgènes (AR). Avant la fixation de l'hormone, son récepteur est localisé dans le cytosol, lié à un complexe protéique chaperon (Hsp90, Hsp70 et FKBP51). La fixation de l'hormone entraîne des modifications de la conformation de son récepteur, sa translocation vers le noyau, et son interaction avec la machinerie transcriptionnelle (McEwan and Smith, 2018). Elle agit sur les cellules de Sertoli, péricubulaires et sur les cellules de Leydig et permet ainsi réguler de la spermatogénèse (Mainwaring et al., 1988). Les œstrogènes (l'œstradiol notamment) agissent via les récepteurs aux œstrogènes exprimés au niveau des cellules de Leydig (ER α), de Sertoli (ER β) ainsi que dans les cellules germinales (Pelletier et al., 2000). Les œstrogènes modulent les effets des gonadotrophines et de la testostérone sur les fonctions testiculaires. Il a été démontré l'importance de ces hormones sur la spermatogénèse et la fertilité (O'Donnell et al., 1996). Pendant la vie de l'homme, la testostérone est nécessaire à la spermatogénèse, mais aussi à l'acquisition des caractères sexuels secondaires. Cette hormone peut également contrôler la fonction érectile et stimuler la libido. Hors

du spectre de la reproduction, elle intervient également dans l'augmentation de la masse musculaire, la production de sébum par la peau et dans la croissance du squelette (Snyder et al., 2016; Wu et al., 2010).

Dans l'ovaire, les cellules de la thèque sont localisées dans le tissu interstitiel, séparées des cellules de la granulosa par une membrane basale. Elles n'apparaissent que lors de la formation des follicules secondaires, dans l'ovaire adulte (Orisaka et al., 2009). Elles semblent partager une origine commune avec les cellules de la granulosa, des cellules précurseurs de l'épithélium cœlomique (Liu et al., 2015). Ces cellules stéroïdogéniques produisent des androgènes, comme l'androstènedione, qui est ensuite converti en œstrogènes, dont l'œstradiol, via CYP19 exprimé dans les cellules de la granulosa (Richards and Pangas, 2010). Après l'ovulation, le corps jaune se forme et produit également de l'œstradiol, mais aussi de la progestérone. Ces hormones étant produites lors du cycle ovarien, leur niveau sérique est également cyclique. Au sein de la gonade, l'œstradiol stimule la prolifération et l'activité anti-apoptotique des cellules de la granulosa (Couse et al., 2005), mais participe également à l'expression des composants de la matrice extracellulaire et de la membrane basale (Zalewski et al., 2012). Il constitue donc un régulateur paracrine indispensable à la croissance folliculaire (Goldenberg et al., 1972). Des récepteurs de la progestérone sont exprimés par les cellules de la granulosa des follicules préovulatoires en réponse au pic de LH. Après invalidation de ces récepteurs par KO, les souris sont stériles, elles présentent un défaut d'ovulation (pour revue, Akison & Robker, 2012). Les hormones ovariennes participent également au cycle utérin. La progestérone produite par l'ovaire permet le développement de la muqueuse utérine lors des cycles et donc participe activement à la gestation (Groothuis et al., 2007).

1.2.1.3.2. Les hormones polypeptidiques

1.2.1.3.2.1. L'activine et l'inhibine

Les activines et inhibines font partie de la famille des *transforming growth factor β* (TGF β). L'activine est produite par de nombreuses cellules, comme les macrophages et les cellules dendritiques, et remplit donc divers rôles, comme celui de médiateur de la fonction de la lignée cellulaire myéloïde (monocyte et macrophage).

Les activines et inhibines produits par le testicule et l'ovaire ont une action contraire au niveau de l'axe hypothalamo-hypophysaire. L'activine se lie à ses récepteurs présents sur les cellules de l'hypophyse et stimule la production de FSH. L'inhibine exerce une action antagoniste sur ces récepteurs, inhibant de fait la production de FSH (Bilezikjian et al., 2004). Au-delà de leurs fonctions au sein de l'axe hypothalamo-hypophysaire, ces deux hormones sont également impliquées dans une régulation paracrine de l'activité des gonades.

Dans le testicule adulte, l'activine est produite par les cellules de Sertoli, et joue un rôle dans la régulation de la maintenance de la barrière hémato-testiculaire, et dans la différenciation des cellules germinales (Loveland, 2018). L'inhibine est également produite par les cellules de Sertoli, et son rôle majeur est son inhibition de la production de FSH par l'hypophyse (Robertson, 2018).

Dans l'ovaire humain, l'activine est produite par les cellules de la granulosa, et la production des deux inhibines est dépendante du cycle ovarien : l'inhibine B est principalement produite par les cellules de la granulosa des follicules en croissance, et l'inhibine A est produite par le corps jaune suite à l'ovulation (Bernard et al., 2018). L'inhibine stimule la production d'androgènes par les cellules de la thèque, alors que l'activine l'inhibe. Chez la souris, l'action de l'activine au sein des cellules de la granulosa est dépendante de l'âge de l'individu. D'une part, dans les follicules de souris immatures, l'activine semble stimuler l'expression des récepteurs à la FSH, ce qui augmente la prolifération des granulosa et la production d'œstrogène via une augmentation de l'action de la FSH. Chez les souris plus âgées, l'activine bloque la croissance folliculaire induite par la FSH. Ces modes de régulation âge-dépendant ne sont pas encore décryptés (Bernard et al., 2018).

1.2.1.3.2.2. L'INSL3 et son action dans le testicule adulte

L'INSL3 est une hormone peptidique de la famille des *insulin relaxin*, et agit grâce à sa fixation sur son récepteur (RXFP2) (Ivell et al., 2017). Elle est produite par les cellules de Leydig adultes, mais bien que son rôle dans la première phase de descente testiculaire soit bien défini (voir « Synthèse de l'insulin-like factor 3 »), son implication à l'âge adulte est moins bien définie. Elle semble être impliquée dans la morphogenèse des os, dans la régulation de la stéroïdogenèse et comme un facteur de survie des cellules du sperme (Ivell et al., 2017).

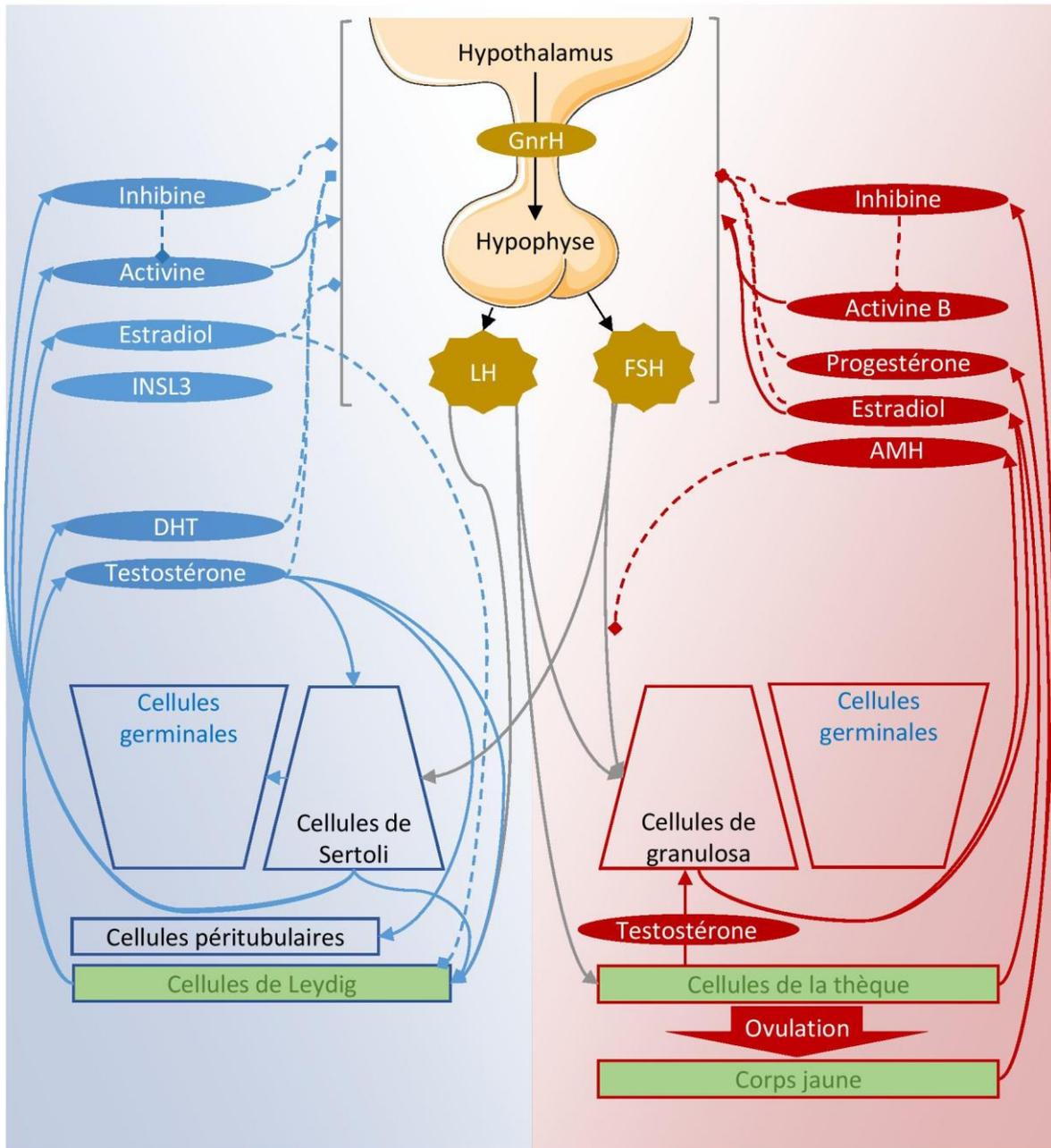


Figure 9. Fonction et régulation endocrine des gonades adultes

Les flèches pleines représentent une régulation positive. Les flèches en pointillés représentent une régulation négative.

1.2.1.3.2.1. L'AMH et son action dans l'ovaire adulte

L'AMH fait partie de la famille des TGF β . Cette hormone est principalement associée à la masculinisation du tractus génital mâle pendant la vie fœtale (voir « Masculinisation du tractus génital : l'AMH »), mais est également impliquée dans la régulation du cycle ovarien (revue dans Mossa and Ireland, 2018). Elle est produite par les cellules de la granulosa des follicules en développement, en fonction du cycle ovarien. En effet, sa production est régulée par de nombreux facteurs encore mal définis.

D'une part, elle serait régulée par les BMPs (BMP4, BMP6, BMP15) produits par les ovocytes et les cellules de la thèque. Sa production serait également dépendante de la FSH, mais ce processus est spécifique des espèces (Monniaux et al., 2012; Scheetz et al., 2012). Elle commence à être exprimée par les follicules primaires, et atteindra son pic de production dans les follicules préantraux. Ensuite, sa production diminue dans le follicule en phase préovulatoire.

L'AMH participe à la régulation du développement folliculaire. Chez des souris mutantes, l'absence de l'hormone entraîne un recrutement accéléré des follicules primordiaux, ce qui aboutit à un épuisement précoce de la réserve ovarienne. Les cultures *in vitro* d'ovaires humains ont également démontré que l'AMH empêche l'activation du développement folliculaire et inhibe également les effets de la FSH sur les follicules en développement (revu dans Dewailly et al., 2014). Globalement, cette hormone a le rôle d'inhibiteur de l'activation des follicules primordiaux, afin d'éviter un appauvrissement prématuré de la réserve ovarienne.

1.2.1.3.3. L'axe hypothalamo-hypophysaire

L'axe hypothalamo-hypophysaire contrôle l'activité endocrine du testicule et de l'ovaire grâce à deux gonadotrophines qui cibleront spécifiquement certaines cellules somatiques au sein de chaque gonade (Kaprra and Huhtaniemi, 2017) (Figure 9). Tout d'abord, La GnRH est une neurohormone synthétisée au niveau de l'hypothalamus, sécrétée de manière pulsatile. Elle est libérée par l'extrémité axonale des neurones hypothalamiques et cible les cellules gonadotropes de l'hypophyse. Deux hormones glycoprotéiques dimériques sont alors libérées dans la circulation générale : la LH et la FSH. Ces deux hormones glycoprotéiques possèdent une sous-unité α commune et une sous-unité β spécifique (Kaprra and Huhtaniemi, 2017).

Le testicule produit de la testostérone, de la dihydrotestostérone (DHT) et de l'œstradiol. La LH active les récepteurs LHCGR (*Luteinizing hormone/choriogonadotropin receptor*) des cellules de Leydig et stimule la production d'androgènes et la spermatogenèse chez l'adulte (Kaprra and Huhtaniemi, 2017). Les récepteurs FHSR ne sont présents que sur les cellules de Sertoli, la FSH influe donc sur la fonction endocrine via les facteurs paracrines des cellules supportrices régulant

l'activité des Leydig (IGF-1 ou le TGF- β) (Tostain et al. 2004). Les hormones produites par le testicule ont la capacité de ralentir les pulsations de GnRH hypothalamique, inhibant ainsi la production de LH et FSH. La testostérone inhibe préférentiellement la LH, tandis que les œstrogènes inhibent la LH et la FSH.

Dans l'ovaire, la LH se lie à son récepteur LHCGR localisé sur la membrane plasmatique des cellules de la granulosa et de la thèque et stimule la production d'œstrogènes, l'ovulation et le développement du jaune. La FSH participe à la multiplication et à la différenciation des cellules de la granulosa lors de la croissance folliculaire. Elle est également nécessaire à la survie et au développement des follicules (Kumar et al., 1997; Matsuda et al., 2012). Les œstrogènes sont capables d'induire un rétrocontrôle positif ou négatif au niveau hypothalamique et hypophysaire en fonction de la phase du cycle ovarien (Radovick et al., 2012; Skynner et al., 1999). La progestérone quant à elle ne peut induire qu'un rétrocontrôle négatif sur la production de GnRH (pour revue, d'Anglemont de Tassigny and Colledge, 2010).

Enfin, conjointement à leurs actions paracrines au sein des gonades adultes, l'activine et l'inhibine sont impliquées dans la régulation de l'axe hypothalamo-hypophysaire (Bilezikjian et al., 2004). L'activine B est produite par les cellules de l'hypophyse, qui exerce une stimulation autocrine et paracrine de la production de FSH. Cependant, l'inhibine produite par les gonades adultes régule son action, en agissant comme antagoniste sur ses récepteurs. Cela permet donc à l'inhibine de participer à régulation de la production de FSH par l'hypophyse.

1.2.2. La gonade indifférenciée

La plupart des connaissances actuelles sur les processus de différenciation et de développement des gonades chez les mammifères proviennent d'études effectuées chez la souris et dans une moindre mesure, chez le rat. Étant donné les différences chronologiques et dynamiques du développement entre l'Homme et les espèces murines, l'extrapolation des mécanismes impliqués chez les rongeurs vers les humains n'est pas automatique (Maheshwari and Fowler 2008). Chez l'Homme, la détection de mutations responsables de pathologies apparentées à des anomalies du développement sexuel permet de mettre en lumière des gènes importants. Cependant, les études fonctionnelles pour déterminer les mécanismes précis dans lesquels ils sont impliqués sont difficiles à mettre en place, étant donné les verrous technologiques, mais surtout éthiques, imposés par le travail sur le fœtus humain. Actuellement, en extrapolant les résultats obtenus chez les organismes modèles, recoupés avec les gènes responsables d'anomalies du développement sexuel (ADS) humains et complétés par les quelques études menées sur testicules et ovaires fœtaux humains, il est dorénavant possible d'appréhender plus précisément les mécanismes en jeu chez l'Homme (revue dans Bashamboo et al., 2017).

1.2.2.1. La crête génitale

La crête génitale trouve son origine dans la surface ventro-medial du mésonéphros (structure embryonnaire des reins), au cours de la 4^{ème} semaine de développement (SD), suite à l'épaississement et à la prolifération de l'épithélium coelomique (Tanaka and Nishinakamura, 2014). Chez l'homme, les gènes contrôlant le développement de la crête génitale ne sont pas encore bien connus. Chez la souris, *Gata4* est exprimé à partir de 10,0 jours post-coitum (jpc) dans la moitié antérieure de la crête génitale, puis transite vers l'épithélium coelomique (Hu et al., 2013). Son expression est suivie par l'épaississement de l'épithélium coelomique, processus capital à la formation de la crête génitale. Son rôle a été confirmé chez des souris mutantes ne possédant pas *Gata4*, où l'épithélium coelomique demeurait monocouche et la crête génitale ne pouvant se développer (Hu et al., 2013). Les gènes *Wnt4* et *Rspo1*, communément considérés comme des acteurs clés de la différenciation ovarienne, sont également exprimés dans la région de l'épithélium coelomique de la gonade bipotentielle. Chez la souris, la délétion de ces deux gènes altère l'épaississement de l'épithélium, indiquant leur implication dans ce processus (Chassot et al., 2012). Ensuite, *WT1* est un gène codant pour 24 isoformes de facteur de transcription à doigts de zinc, dont l'isoforme WT1-KTS est un facteur clé pour le développement de la crête génitale murine (Hammes et al., 2001; Kreidberg et al., 1993). Sa délétion entraîne une mort cellulaire accrue, qui aboutira à un empêchement de la crête génitale.

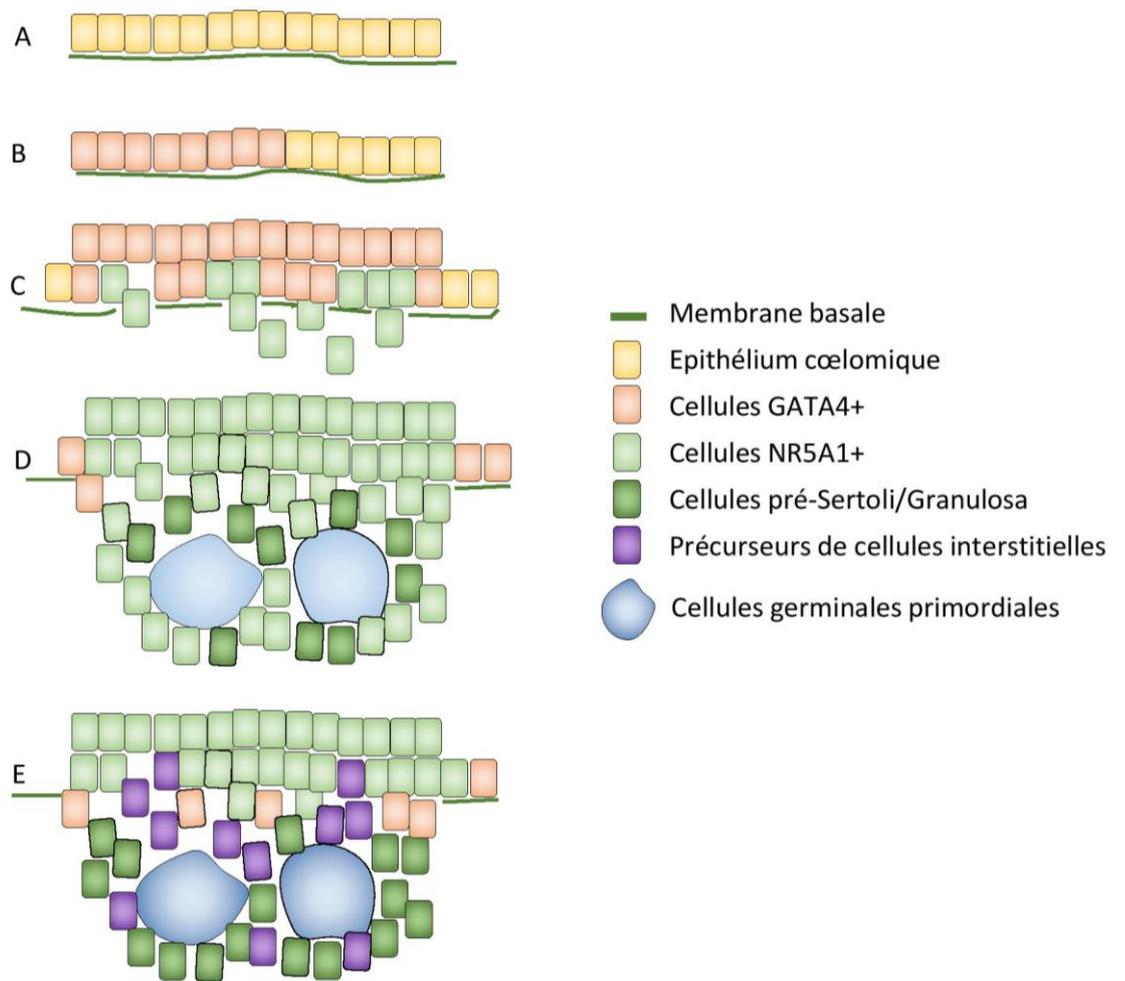


Figure 10. Schématisation de la formation de la crête génitale et mécanismes moléculaires en jeu.

(A) L'épithélium cœlomique repose sur une membrane basale. (B) Certaines cellules expriment GATA4, puis (C) les cellules exprimant NR5A1 (SF1) s'ingressent pour former la crête génitale. (D) Les progéniteurs des cellules somatiques continuent à proliférer et les CGP colonisent la future gonade. (E) La migration et prolifération des cellules de l'épithélium cœlomique va fournir les précurseurs des différentes populations cellulaires somatiques de la gonade.

Avec le cofacteur *CITED2* (Wilhelm, 2002), il induit également l'expression du gène *NR5A1* (nuclear receptor subfamily 5, codant pour SF1, steroidogenic factor 1), contribuant tous deux à la maintenance de l'épithélium (Luo et al., 1994). Pendant cette phase du développement, *NR5A1* (*SF1*) est exprimé exclusivement dans la crête génitale et est dès lors considéré comme l'un de ses marqueurs. Son expression persiste après la différenciation de la gonade, dans les cellules pré-Sertoli et les cellules de Leydig fœtales jusqu'à 12.5 jpc (Ikeda, 1994) et dans l'ovaire jusqu'à 13.0 jpc. *NR5A1* constitue une clé de voute de la maintenance et du développement de la crête génitale. Dans le cas où son expression n'est pas maintenue par ses différents régulateurs, tels qu'*EMX2*, *CBX2*, *LHX9* ou *SIX1/SIX4*, la structure dégénère à partir de 12.5 jpc (Birk et al., 2000; Katoh-Fukui et al., 2012; Kusaka et al., 2010; Mazaud et al., 2002; Sproll et al., 2018; Wilhelm, 2002). L'expression de *NR5A1* est également activée par le récepteur à l'insuline (*INSR*) et le facteur de croissance insuline-like 1 (*IGF1R*), ils font donc partie des promoteurs du développement de la crête génitale, induisant la prolifération des cellules somatiques en son sein (Nef et al., 2003; Pitetti et al., 2013a). Un réseau complexe gouverne la formation et la maintenance de la crête génitale, comprenant la coopération de plusieurs facteurs.

Ces observations ont été faites chez la souris, mais certains cas de mutation sur le gène *WT1* chez des patients 46, XY atteints de dysgénésie gonadique ont confirmés le rôle central de ce gène dans la formation de la gonade (Moorthy et al., 1987). De même, des mutations sur le gène *NR5A1* induisent un large spectre de phénotypes chez les patients, certains présentant une dysgénésie complète des gonades et des glandes surrénales (revu dans Domenice et al., 2016).

Suite au développement des cellules dans la crête génitale, l'ingression de l'épithélium cœlomique est permise grâce à une transition épithélio-méchymateuse des cellules, qui aboutira à la formation de la gonade (revue dans Yoshino and Saito, 2018). L'ingression des cellules est dirigée par la polarisation de celles-ci, gouvernée par le système Notch. Les cellules progénitrices proliférant, elles présentent une forte activité Notch et se maintiennent dans l'épithélium. Les cellules filles produites expriment quant à elles l'antagoniste de Notch, *NUMB*, vont s'ingresser dans la gonade dans le but de devenir des progénitrices intermédiaires qui se différencieront en cellules somatiques (Lin et al., 2017).

1.2.2.2. Origine et migration des cellules germinales

Les cellules germinales primordiales (CGP) apparaissent pendant la 2^{ème} et la 3^{ème} SD dans l'épiblaste (De Felici and Farini, 2012; Leitch et al., 2013). Elles subissent une première étape de migration extra embryonnaire vers la paroi de la vésicule vitelline où elles peuvent être observées au début de la 4^{ème} SD (Witchi, 1948).

Il est possible d'étudier les CGP lors de cette migration, mais aucun protocole ne permet de les caractériser pendant leur différenciation, étant donné les verrous techniques évidents imposés par l'étude du développement fœtal aussi tôt dans la grossesse (revu dans Tang et al., 2016a). Pendant longtemps, le modèle proposé pour les mécanismes d'induction des CGP chez l'homme s'inspirait des études menées chez la souris, mais les différences interspèces existent déjà lors du développement précoce de l'embryon. En effet, les embryons de rongeurs ont une forme cylindrique, alors que d'autres mammifères, dont l'Homme, se développent en un disque plat. Ces configurations spatiales différentes pourraient avoir un impact sur les gradients de signalisations et donc en aval, sur les voies moléculaires de la différenciation des CGP (Irie et al., 2014; Tang et al., 2015). Des études transcriptomiques récentes ont permis de comparer les programmes transcriptionnels gouvernant la spécification des CGP entre la souris et l'homme (Kobayashi et al., 2017). Ainsi, la partie du programme conservée entre les deux espèces a peut être défini, mais des différences significatives ont également été décrites. Notamment, *Sox2* est indispensable à la spécification des CGP murines, mais n'est pas exprimé par leurs équivalents humains. Cependant, *SOX17* est un facteur de transcription spécifiquement exprimé chez l'Homme. De même, *PRDM14* est requis pour la spécification souris, mais n'est pas observé chez l'homme.

Après leur différenciation, les CGP colonisent la crête génitale au cours de la 5^{ème} SD (Witchi, 1948). Lors de cette migration, elles entrent dans une phase de prolifération rapide grâce à une forte activité mitotique (Bendel-Stenzel et al., 1998; Francavilla et al., 1990; Rouiller-Fabre et al., 2009; Wylie, 1999). Plusieurs mécanismes chimiotactiques permettent leur adressage à la crête génitale. D'une part, les CGP expriment le récepteur KIT/c-KIT et les cellules somatiques expriment son ligand, KITLG/SCF/STEEL (Matsui et al., 1990; Richardson and Lehmann, 2010). Le ligand KITLG est également exprimé par les fibres nerveuses du mésentère dorsal (Høyer et al., 2005), qui pourraient donc participer à leur migration (Mamsen et al., 2012; Møllgård et al., 2010). D'autre part, les cellules germinales expriment le récepteur CXCR4 (Ara et al., 2003; Molyneaux et al., 2003), guidant leur migration vers la crête génitale, qui secrète le ligand CXCL12 (Richardson and Lehmann, 2010). En l'absence de cette chimiokine, les cellules germinales primordiales n'achèvent pas leur migration jusqu'aux crêtes génitales. De plus, leur migration est favorisée par le développement et l'expansion concomitants de l'embryon (Freeman, 2003; Hara et al., 2009).

Lorsque leur migration est complète au cours de la 6^{ème} SD, les CPG sont désignées sous le nom de gonocytes (XY) ou ovogonie (XX). Dans le cas des gonocytes, les cellules conserveront une forte activité mitotique jusqu'à la 10^{ème} SD, puis entreront en quiescence mitotique de manière asynchrone. Dans l'ovaire, les ovogonies proliféreront également jusqu'à la 10^{ème} SD, puis entreront en méiose (voir « Cellules germinales »).

De manière intéressante, il a été démontré chez la souris que la présence d'une crête génitale n'est pas requise pour induire la migration des CGP (Hu et al., 2013). La migration des CGP n'est pas non plus requise pour la formation de la crête génitale (McLaren, 1991), ou plus tard, la différenciation des différentes populations cellulaires somatiques du testicule ou de l'ovaire (Handel and Eppig, 1979; Maatouk et al., 2012; McCoshen, 1982). Ainsi, la gonade indifférenciée provient de deux processus biologiques complémentaires mais apparemment indépendants.

Les cellules germinales et leurs précurseurs sont les cellules essentielles à la reproduction. Dès lors, une question essentielle reste en suspend : pourquoi la survie de l'espèce dépend de mécanismes aussi complexes que la spécification à l'extérieur de la gonade de quelques dizaines de cellules, et de leur migration à travers un organisme en développement, rempli de signaux de toutes sortes travaillant à la différenciation des autres organes ? Comment ces mécanismes, apparemment susceptibles à des erreurs de toute sorte (mauvaise différenciation, dégénérescence, mauvais adressage), ont acquis une robustesse telle que le règne animal persiste ?

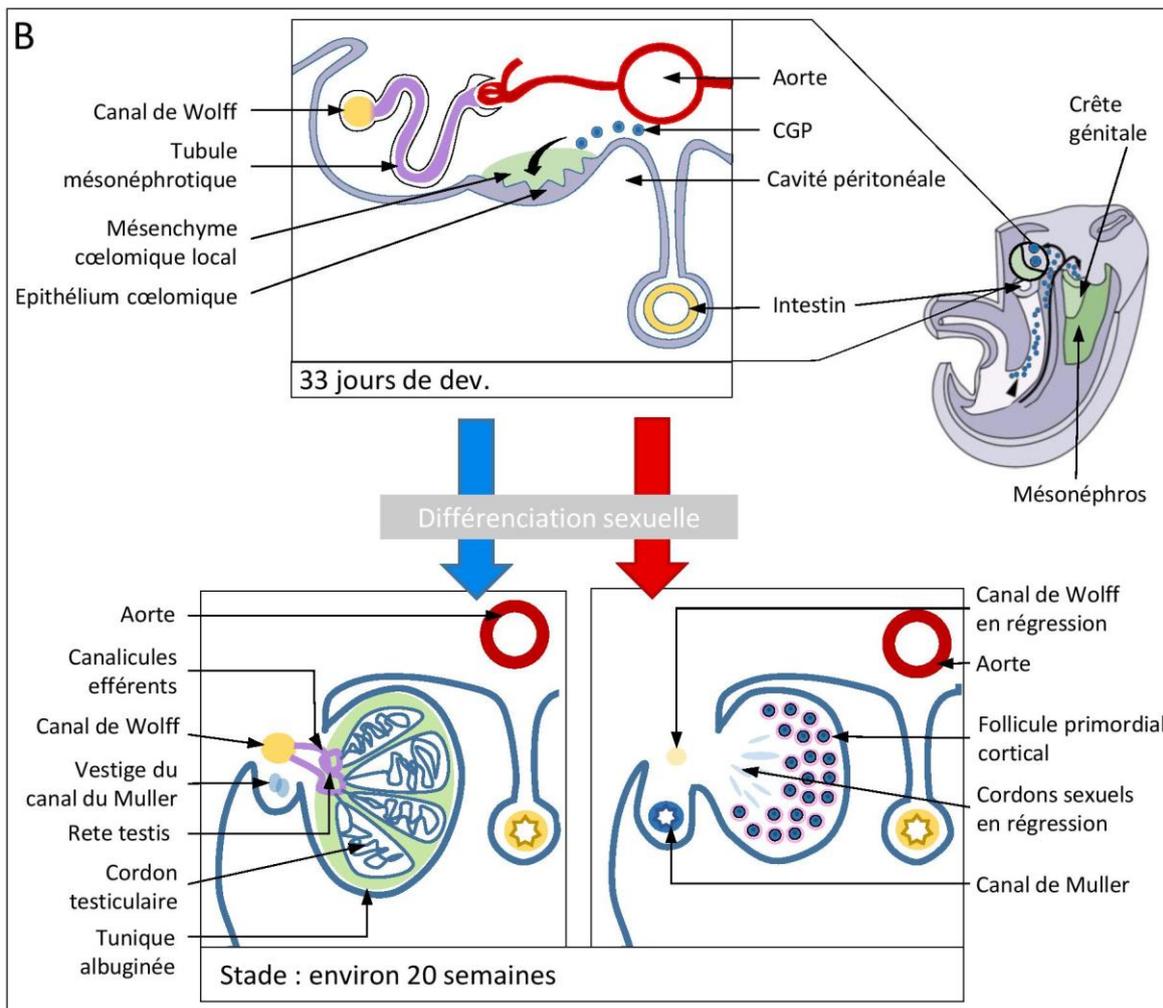
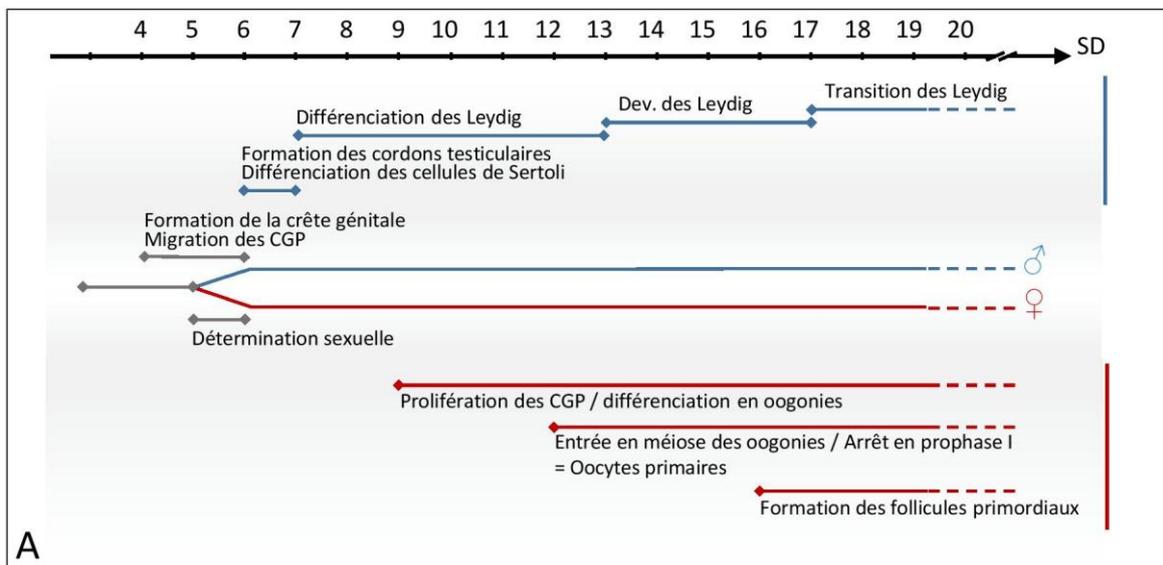


Figure 11. Différenciation de la gonade bipotentielle en testicule ou ovaire.

(A) Chronologie de la différenciation des populations cellulaires. (B) Schématisation de la formation de la gonade bipotentielle et de son devenir suite à la différenciation sexuelle.

1.2.3. La différenciation sexuelle

La différenciation sexuelle des gonades est le point de départ de la différenciation et du développement du testicule et de l'ovaire. Grâce à une fine régulation de l'expression de certains gènes débutant au cours de la 6^{ème} semaine de développement (SD), la gonade bipotentielle pourra s'engager vers une trajectoire mâle ou femelle. De multiples phases de différenciation et de développement cellulaires spécifiques de chaque gonade s'ensuivront, concomitante avec la formation des différentes structures tissulaires (Figure 11).

1.2.3.1. Les cellules somatiques : origine de la différenciation

L'épithélium cœlomique donnera naissance à deux types de cellules somatiques : les cellules supportrices et les cellules stéroïdogènes, dans le testicule et dans l'ovaire. Les cellules supportrices seront les premières à se différencier vers la voie male, en cellules de Sertoli, ou vers la voie femelle, en précurseurs de la granulosa. Elles gouvernent ainsi la détermination sexuelle, plus tard la différenciation des cellules stéroïdogènes, puis à l'âge adulte, soutiendront la gamétogenèse (voir « Cellules de Sertoli », »Cellules de pré-granulosa » et « La fonction exocrine des gonades adultes »).

1.2.3.1.1. Trajectoire mâle

- *SRY : Régulation et mode d'action*

Historique de la découverte

La découverte des chromosomes sexuels a eu lieu en 1921, mais leur implication dans la détermination sexuelle a été définie dans les années 60. Grâce à l'observation des caryotypes anormaux de patients présentant des anomalies du développement sexuel, un syndrome de Turner (Femme XO) et le syndrome de Klinefelter (homme XXY) (Jacobs and Strong, 1959), les chercheurs ont observé que la différenciation de la gonade en testicule était dépendante de la présence ou non du chromosome Y, indépendamment du nombre de chromosomes X. Cependant, l'identification du facteur déterminant du testicule (*TDF*, testis determining factor) a longtemps constitué un défi scientifique. L'identification d'une région de 35kb transloquée du chromosome Y chez des patients XX présentant des caractéristiques mâles a permis d'orienter les recherches (Palmer et al., 1989). Ce n'est que dans les années 1990 que les chercheurs ont su identifier le gène **SRY** (sex determining region Y) (Gubbay et al., 1990; Sinclair et al., 1990). Aujourd'hui, plus de 50 mutations du gène *SRY* ont été associées à des cas d'inversion du sexe chez l'homme (revue dans Gunes et al., 2016).

Le gène *SRY* fait partie de la famille des SOX qui encodent des facteurs de transcription contenant des domaines HMG (High mobility group). Ce domaine lui permet de lier spécifiquement l'ADN grâce au motif de la boîte HMG et de lui appliquer une courbure, permettant ainsi l'interaction des différents facteurs de transcription, donc de réguler l'expression de ses gènes cibles (Harley and Goodfellow, 1994; She and Yang, 2016).

Conservation et profil d'expression spatio-temporel chez les mammifères

La fonction de *SRY* est conservée chez les mammifères. En effet, les mutations sur le gène induisent des dysgénésies gonadiques pures chez l'homme (Jäger et al., 1992) et des phénotypes d'inversion du sexe chez la souris. De plus, l'ajout de la fonction *Sry* à des souris de caryotype XX induit le développement d'un testicule (Koopman et al., 1991), confirmant son rôle clé dans le développement mâle. Cependant, bien que sa fonction soit conservée, le gène *SRY* et ses homologues exhibent des différences interspèces, tant au niveau de leurs séquences qu'au niveau de leur expression au cours du développement gonadique.

Tout d'abord, des études phylogénétiques ont démontré que le gène *SRY* résulterait de l'évolution différentielle sur le chromosome Y du gène *SOX3* qui serait son homologue sur le chromosome X, divergeant il y a environ 166-148 millions d'années (Foster and Graves, 1994; Graves, 2002; Wallis et al., 2008). Ensuite, l'expression de *Sry* pendant le développement de la souris a été observée dans les cellules précurseurs des Sertoli, pendant seulement quelques heures dans chaque cellule, entre 10.5 et 12.5 jpc (Hacker et al., 1995; Koopman et al., 1990). Chez l'homme, l'expression de *SRY* est initiée entre 41 et 44 jours post ovulation et le gène est toujours faiblement exprimé pendant la vie fœtale (Hanley et al., 2000).

Maintien et amplification de l'expression de *Sry*

Certains gènes impliqués dans le développement de la crête génitale interviennent dans l'induction de *SRY*, comme *WT1* (*WT1-KTS* impliqué dans la formation de la crête, l'isoforme *WT1+KTS* dans la différenciation masculine) (Hammes et al., 2001; Hersmus et al., 2012; Hossain and Saunders, 2001; Shimamura et al., 1997) et *NR5A1* (*SF1*) (Achermann et al., 1999; de Santa Barbara et al., 2001; Sekido and Lovell-Badge, 2013) (Figure 12). Chez la souris, l'activation de *Map3k4* par *GADD45G* induit la phosphorylation de *GATA4*, permettant l'interaction de ce dernier avec *FOG2*, et réguleront l'expression de *Sry* (Gierl et al., 2012b; Lourenco et al., 2011; Miyamoto et al., 2008; Tevosian et al., 2002; Warr et al., 2012b). Cependant, aucune mutation de *MAP3K4* ou de *GADD45G* n'a été associée à des ADS chez l'homme, et aucun site de fixation de *GATA4* fonctionnel n'est présent dans le promoteur de *SRY*. Cependant, la voie des MAP kinases semble être impliquée car des cas de perturbation de la détermination testiculaire ont été observés dans le cas de

mutations sur *MAP3K1* (Pearlman et al., 2010). CBX2 est également un potentiel régulateur indirect de *SRY* au regard des conséquences phénotypiques d'une mutation dans sa séquence (Femme XY) (Biason-Lauber et al., 2009). En effet, ces résultats suggèrent que l'isoforme 1 de CBX2 est impliquée dans la régulation de *SRY* via l'induction de l'expression de *NR5A1(SF1)* ou de *SOX9* et l'inhibition de facteurs de la voie femelle, comme *FOXL2* (Biason-Lauber et al., 2009; Eid et al., 2015). Etant donné que l'isoforme 2 de CBX2 est impliquée dans la maintenance de la crête génitale via la régulation potentielle d'*EMX2*, le gène serait donc impliqué à plusieurs niveaux dans le développement des gonades, en fonction de l'isoforme transcrite, comme le cas de WT1 (isoforme – KTS et +KTS) (Sproll et al., 2018).

De plus, l'expression de *SRY* est également la cible de régulations épigénétiques via la méthylation de l'ADN, notamment grâce à la demethylase KDM3A (JMJD1A) chez la souris (Kuroki et al., 2013), de la chromatine (Tachibana, 2015) et des modifications des histones, contrôlées par CBP (CREBBP, CREB-binding protein) et p300 (EP300), responsable de cas de ADS (Carré et al., 2017; Eggers et al., 2016).

Les régulateurs de *SRY* ont été étudiés dans plusieurs espèces et bien que certains mécanismes apparaissent conservés entre l'homme et la souris, de nombreuses spécificités ont également été mises en évidence (revues dans Larney et al., 2014). Encore aujourd'hui, le réseau transcriptionnel qui contrôle l'expression de *SRY* n'est pas encore élucidé.

- *SOX9 : Antagoniste de la différenciation ovarienne*

Le gène cible le plus connu de *SRY* est *SOX9* (Sry-box 9), qui code également pour un facteur de transcription possédant un domaine HMG. Il est nécessaire et suffisant pour induire la différenciation de la gonade vers le testicule, car sa déficience dans une gonade XY peut induire une inversion de sexe chez l'homme et la souris et son expression dans une gonade XX peut induire sa différenciation vers la voie mâle (Barrionuevo et al., 2006; Bishop et al., 2000; Chaboissier, 2004; Foster et al., 1994; Huang et al., 1999; Vidal et al., 2001). De plus, de nombreuses mutations de *SOX9* ont été détectées chez des patients présentant des dysgénésies complètes ou des 46,XY ADS (Cameron et al., 1996; Foster et al., 1994; Ninomiya et al., 2000; Pop et al., 2005; Wagner et al., 1994).

SOX9 est une cible directe de *SRY*, mais requiert également l'action synergique d'autres facteurs déjà décrits, tels que GATA4, WT1 et NR5A1 (SF1) (Figure 12). Une expérience de CHIP (Chromatin immunoprecipitation) sur des gonades murines a identifié une région de 15 kb dans la région promotrice de *SOX9* liée par *SRY*, appelée TES (Sekido and Lovell-Badge, 2008). Le cœur de cette région, appelé TESCO, semble également être la cible d'autres facteurs tels que NR5A1 et *SOX9* lui-même, pour amplifier son expression et ainsi verrouiller la différenciation des cellules de

Sertoli. Récemment, une expérience de Crispr-Cas9 (technique d'édition de génome) chez la souris a démontré que la délétion de cette région n'entraînait pas une inversion du sexe chez la souris, mais diminuait significativement l'expression de *Sox9* (Gonen et al., 2017). Cela indique donc que même chez le modèle murin, largement étudié depuis plusieurs années, d'autres régulateurs encore inconnus permettent la maintenance de l'expression de *Sox9* au sein des cellules de Sertoli. En juin 2018, des avancées significatives ont été faites. Chez la souris, une étude a démontré l'existence d'une région promotrice de *Sox9* chez la souris, appelée Enh13, dont la délétion induit une inversion du sexe complète, en présence ou non de la région TESCO (Gonen et al., 2018). Cette nouvelle découverte remet en cause TESCO en tant que région principale dans la régulation de *Sox9*.

Récemment, une étude utilisant la technologie de CHIP-seq pour cartographier les gènes potentiellement régulés par SOX9 a été effectuée sur le testicule fœtal bovin et murin (Rahmoun et al., 2017). Cela a pu mettre en évidence l'interaction de SOX9 avec des milliers de gènes cibles, y compris les gènes déjà connus pour être impliqués dans le développement testiculaire. Aucune étude fonctionnelle n'a pu être encore effectuée pour valider la régulation transcriptionnelle potentiellement induite par SOX9 sur ces nouvelles cibles candidates.

Il serait également important de comprendre les relations entre *SOX9* et les autres gènes de la famille des SOX impliqués dans le développement gonadique, tels que *SOX8*, *SOX10* et *SOX3*. Par exemple, *SOX3* est impliqué dans plusieurs cas de ADS (Grinson et al., 2016; Haines et al., 2015; Moalem et al., 2012; Sutton et al., 2011; Vetro et al., 2015), mais les mécanismes moléculaires dans lesquels il est impliqué ne sont pas encore connus. Cependant, il a été démontré que ces orthologues peuvent avoir une activité compensatoire chez la souris (Bergstrom et al., 2000).

- *Amplification autocrine/paracrine du signal Sox9 dans les cellules de Sertoli en différenciation*

La place de *Fgf9* dans le réseau transcriptionnel gouvernant la détermination sexuelle a pu être principalement étudiée chez des souris mutantes. *Fgf9* est régulé par SOX9 et participe au rétrocontrôle positif de celui-ci, ainsi tous deux forment une boucle de régulation positive (Hiramatsu et al., 2010). Il exerce également un recrutement paracrine grâce à l'expression de son récepteur *Fgfr2* (Kim et al., 2006) (Figure 12). Chez l'homme, un cas de patient présentant une mutation missens sur *FGFR2* associée à une dysgénésie testiculaire complète a été rapportée (Bagheri-Fam et al., 2015). *Fgf9* est également l'antagoniste du système de gène pro-ovarien WNT4/RSPO1/ β -catenin (Jameson et al., 2012a), mais les mécanismes moléculaires précis n'ont pas encore été décrits.

La prostaglandine D2 (PGD2) semble avoir une action similaire et indépendante de FGF9, amplifiant l'expression de *Sox9* et en participant au recrutement de cellules somatiques vers la différenciation en cellules de Sertoli (Adams and McLaren, 2002; Moniot et al., 2009; Wilhelm et

al., 2007). Ce système apparaît comme un réseau de régulation de soutien, afin de verrouiller la différenciation des cellules somatiques vers une destinée mâle et donc assurer le futur de la gonade vers le testicule (Wilhelm et al., 2005).

À cause de leur implication dans la prolifération des cellules progénitrices somatiques de la gonade, les facteurs de croissance de type insuline participent à l'établissement des cellules de Sertoli dans le testicule. Chez la souris, l'inactivation du récepteur à l'IGF1, *Igf1r*, entraîne une réduction de la prolifération des cellules de Sertoli, de la taille testiculaire et de la production séminale (Pitetti et al., 2013a). De manière similaire, le facteur de croissance *Nrg1* semble impliqué dans la prolifération des cellules progénitrices des cellules de Sertoli, et son inactivation entraîne une hypoplasie testiculaire chez la souris (Gregoire et al., 2018).

- *Les gènes impliqués dans les ADS : fonction encore mal comprise dans la différenciation sexuelle*

Chez la souris adulte, l'inhibition de *Dmrt1* dans le testicule adulte induit l'expression de *Foxl2* ainsi que la reprogrammation des cellules de Sertoli en cellules similaires à celles de la granulosa. Des cellules de la thèque apparaissent et produisent des œstrogènes et les cellules germinales subissent une féminisation (Matson et al., 2011). Chez l'homme, la perte de *DMRT1* est associée à un large spectre de dysgénésie gonadique 46, XY (Eggers and Sinclair, 2012), suggérant son rôle clé dans la différenciation testiculaire.

- « *Phylosophie* »

Il est pertinent de s'interroger sur la légitimité d'un système aussi complexe d'amplification de signal paracrine. Il semblerait que le système XY mis en place chez le mammifère ait évolué en considération de la fragilité de *SRY*. En effet, chez la souris, *Sry* est exprimé à un niveau proche de son seuil d'inactivité, pendant une période très transitoire (Kashimada and Koopman, 2010). Chez l'homme, son niveau d'expression est très faible par rapport au niveau de *Sox9*.

De plus, la plus grande faiblesse de *SRY* est sa localisation sur le chromosome Y, car il ne peut corriger ses mutations en utilisant la machinerie classique mise en place lors de la méiose. Ainsi, il a été observé que certaines dysgénésies gonadiques de patients 46,XY sont dues à des mutations sur *SRY*. On peut donc penser que l'évolution va conduire à l'incapacitation progressive de *SRY* (Kashimada and Koopman, 2010; Polanco and Koopman, 2007), conjointement à la disparition du chromosome Y (Graves, 2006). Donc, si la différenciation de chaque cellule dépendait du niveau d'expression seule de *SRY*, on pourrait assister à une différenciation hétérogène des cellules somatiques et donc à un mélange de cellules Sertoli et de la granulosa. Ce système, dépendant d'un type cellulaire plutôt que l'expression d'un seul gène, apparaît comme plus robuste, se propageant grâce à l'amplification autocrine et paracrine du signal de *SOX9* (Sekido and Lovell-Badge, 2008), gène subissant moins de pression évolutive.

En 2006, Jennifer Marshall Graves a partagé dans *Nature* sa prédiction selon laquelle le chromosome Y pourrait disparaître dans les 5 millions d'années à venir (Graves, 2006). En effet, celui-ci a perdu plus de 95 % de ses gènes durant les 190 millions d'années passées, portant aujourd'hui environ 50 gènes. Cependant, cette disparition ne serait pas synonyme de disparition des mâles. Si l'on considère les différentes stratégies de différenciation sexuelle mises en place au sein du règne animal, ces systèmes semblent évoluer très rapidement. D'une part, il est possible que le gène soit dupliqué ou déplacé vers un autre chromosome, comme *DMRT1* semble l'avoir fait chez le poisson medaka (Kondo et al., 2004). D'autre part, certaines mutations sur des gènes intervenant en aval de *SRY* dans la cascade de la différenciation sexuelle peuvent induire une inversion du sexe. Par exemple, une augmentation du dosage de *SOX9* peut induire le développement de testicule chez un individu XX (Qin and Bishop, 2005). L'évolution du génome pourrait permettre que l'un de ces gènes prenne la place du gène régulateur de la différenciation sexuelle.

Cependant, cette théorie selon laquelle le chromosome Y sera amené à disparaître a été contestée par plusieurs travaux, notamment grâce à l'étude comparative de sa séquence et des gènes qu'il porte dans plusieurs espèces. Notamment, les chromosomes Y de l'homme et du macaque rhésus, deux espèces ayant divergé il y a environ 30 millions d'années, sont remarquablement similaires, y compris au niveau de la conservation des gènes qu'ils portent (Hughes et al., 2010). Une récente étude ayant caractérisé de manière approfondie la séquence du chromosome Y de 62 patients (comprenant des pères et leurs fils) a également mis en évidence que le processus de conversion intra-chromosomique permettrait de protéger les gènes de la dégénérescence (Skov et al., 2017). Cela contredit donc l'hypothèse de la dégénérescence inexorable du chromosome Y.

L'étude des systèmes du déterminisme sexuel au sein du règne animal a mis en évidence pléthore de modèles évolutifs, et les données actuelles permettent d'émettre des hypothèses sur l'évolution passée du chromosome Y humain, mais ne permettent pas de prédire son avenir (revu dans Bachtrog, 2013; Gu and Walters, 2017; Mank et al., 2014; Wright et al., 2016).

1.2.3.1.2. Trajectoire femelle

La détermination sexuelle menant vers la différenciation femelle a également lieu dans les futures cellules supportrices, les cellules de la granulosa. Comme pour le testicule, elles requièrent l'expression des gènes comme *GATA4/FOG2*, *WT1*, *NR5A1 (SF1)* (initialement exprimés par la gonade indifférenciée) et de l'insuline pour maintenir leur survie et leur expansion (Chen et al., 2017a; Combes et al., 2010; Manuylov et al., 2008; Pitetti et al., 2013b). De manière similaire à leur influence sur l'expression de *SRY* et de *SOX9* pendant la différenciation testiculaire, cet ensemble de gènes facilite l'expression de la voie de signalisation ovarienne WNT4/RSPO1/ β -catenin. Avant la différenciation sexuelle, l'expression de ces trois gènes n'est pas différentielle entre une gonade XX et une gonade XY et plus globalement, les cellules somatiques femelles subiront moins de modifications transcriptionnelles que les cellules mâles (Jameson et al., 2012b; Munger et al., 2013; Nef et al., 2005).

- *La voie RSPO1/WNT4/ β -Catenin*

Chez la souris, *Wnt4* et *Rspo1* sont exprimés par les cellules précurseurs somatiques et exercent une régulation autocrine, par l'amplification de leur propre expression (Chassot et al., 2008; Trautmann et al., 2008; Yao et al., 2004a) et une régulation paracrine, via les récepteurs Frizzled, LRP5 et LRP6 pour WNT4 et LGR4/5 pour RSPO1. Leur fixation à leurs récepteurs déclenche une cascade d'activation qui va permettre la stabilisation et la translocation nucléaire de la β -catenin (*CTNNB1*) cytoplasmique (Chassot et al., 2008) (Figure 12).

Chez l'homme, *WNT4* n'est pas différentiellement exprimé pendant le développement de l'ovaire et du testicule (Mamsen et al., 2017), cependant le gène demeure important pour la différenciation ovarienne. En effet, des mutations non fonctionnelles de *WNT4* ont déjà été associées à des cas de patients 46,XX présentant un phénotype virilisé et un défaut de structures Mulleriennes (Biaison-Lauber et al., 2004; Philibert et al., 2011). *RSPO1* est spécifiquement exprimé par l'ovaire (Mamsen et al., 2017; Tomaselli et al., 2011) et sa perte est associée à une inversion complète du sexe chez des individus 46,XX (Parma et al., 2006; Tomaselli et al., 2008). Dans un cas opposé, le gain des gènes *WNT4* et *RSPO1* chez un individu 46,XY entraîne une inversion du sexe, vers l'ovaire (Jordan et al., 2001).

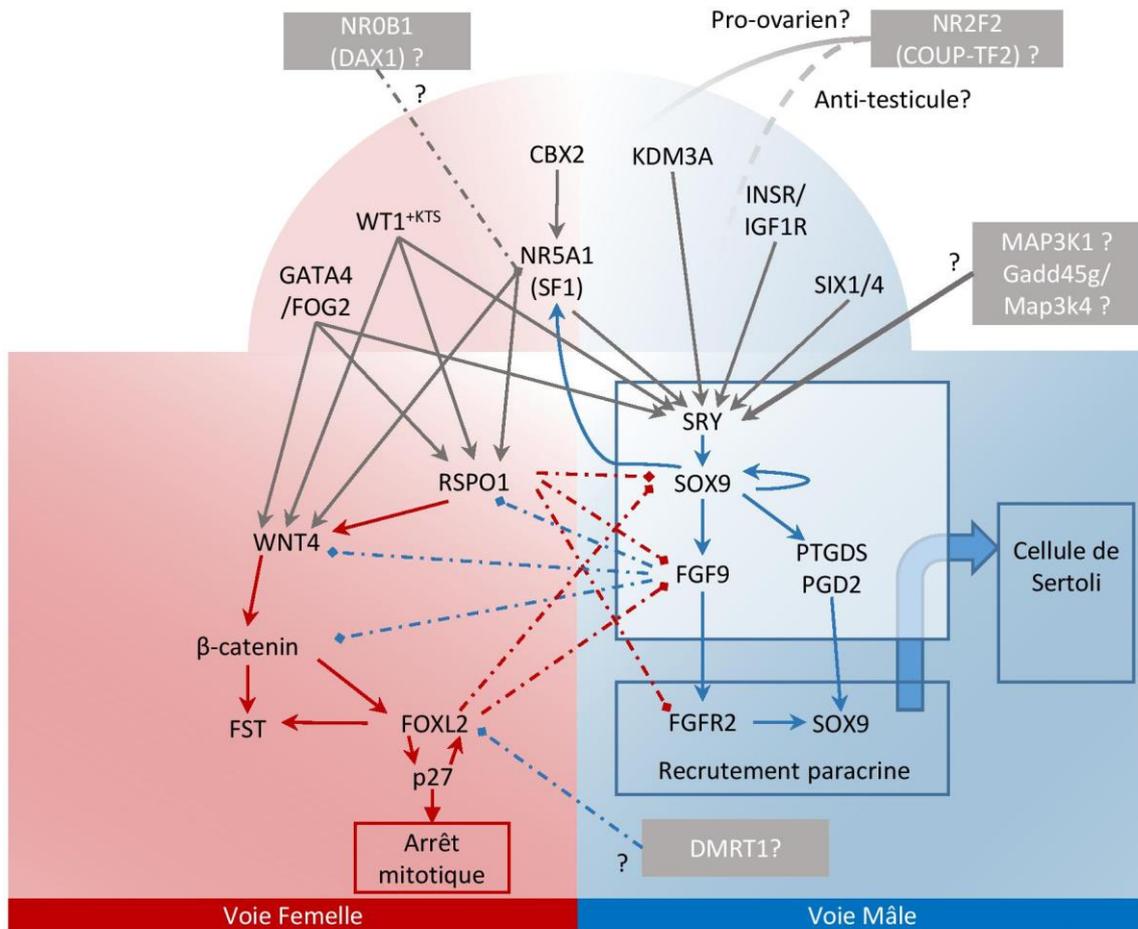


Figure 12. Régulation transcriptionnelle de la différenciation sexuelle.

Certains gènes initient et maintiennent l'expression des gènes déclenchant la cascade transcriptionnelle gouvernant la différenciation sexuelle (arêtes grises). La voie femelle (liens rouges) et la voie mâle (liens bleus) sont dirigées par un réseau complexe activation de la transcription (flèches) et d'inhibition (pointillés) antagonistes des gènes appartenant à la voie alternative de différenciation. Certains gènes ont été associés à des anomalies du développement sexuel humain, mais leurs implications moléculaires au sein de ce réseau restent à confirmer (boîtes grises).

- *FOXL2*

Suite à l'expression du système WNT4/RSPO1/ β -catenin, la voie de différenciation femelle est renforcée par l'expression du gène *FOXL2* (revue dans Pannetier et al., 2016). Chez l'Homme, une mutation dans le gène induit une dégénérescence ovarienne prématurée, mais pas d'inversion du sexe (Crisponi et al., 2001; Harris et al., 2002). Cependant, aucune mutation homozygote de *Foxl2* n'a encore été observée chez une patiente, il est donc difficile de définir si ce gène est indispensable au développement ovarien humain. Chez la chèvre, l'inactivation de ce gène entraîne une inversion du sexe complète (Boulangier et al., 2014). Des expériences de souris mutantes ont démontré que les voies de transcription initiées par FOXL2 et WNT4/RSPO1/ β -catenin des rôles complémentaires dans la différenciation ovarienne (Auguste et al., 2011; Ottolenghi et al., 2007).

- *Les gènes associés à des ADS*

Dax1 (Nr0b1) a été pendant un temps considéré comme un gène déterminant pour la différenciation ovarienne (Swain et al., 1998), car les souris transgéniques XY possédant plusieurs copies du gène développaient une inversion du sexe, du testicule vers l'ovaire (Swain et al., 1998). Cependant, ce rôle pro-ovarien a été invalidé, car la délétion du gène dans des souris mutantes n'a pas démontré d'effet sur la différenciation des précurseurs de granulosa (Meeks, 2003; Meeks et al., 2003a, 2003b; Yu et al., 1998). Ces données suggèrent donc un rôle dose-dépendant d'activité anti-testicule du gène, confirmé par la détection d'une duplication de *DAX1* dans un cas de patient 46,XY atteint d'inversion du sexe (Zanaria et al., 1994). Les résultats d'expériences menées chez la souris suggèrent que DAX1 serait un inhibiteur de l'activation de *Nr5a1* (Lalli and Sassone-Corsi, 2003).

Récemment, il a été démontré que *Nr2f2 (Coup-tf2)* promouvait la dégénérescence du tractus génital mâle dans l'ovaire souris (Zhao et al., 2017). Chez l'homme, l'inversion du sexe d'un patient 46,XX dont une mutation avait invalidé fonctionnellement le gène *NR2F2 (COUP-TF2)* a été rapportée (Bashamboo et al., 2018). Dès lors, il semble que ce gène joue un rôle pro-ovarien et anti-testicule lors de la différenciation ovarienne. Cependant, aucune fonction analogue n'est décrite chez la souris, les mécanismes moléculaires en jeu restent donc à caractériser.

1.2.3.1.3. L'antagonisme des voies mâles et femelles

Lorsque la détermination sexuelle est initiée, la voie de différenciation femelle induite par WNT4/RSPO1/ β -catenin est inhibée dans les gonades XY par les gènes *WT1, FGF9, SOX9 et ZNRF3* (Chang et al., 2008; Harris et al., 2018; Kim et al., 2006; Maatouk et al., 2008). De la même manière, WNT4/RSPO1/ β -catenin supprime l'expression de *SOX9* et de *FGF9* (Kim et al., 2006; Maatouk et al., 2008). Des expériences de CHIP ont démontré que FOXL2 avait la capacité de se fixer à la région

TESCO et ainsi d'inhiber continuellement l'expression de Sox9, même à l'âge adulte (Uhlenhaut et al., 2009).

En comparaison du nombre de mutations induisant des ADS chez des patients 46,XY, peu de cas d'ADS 46, XX sont connus. Malgré l'observation d'antagonisme mutuel entre des paires de gènes pro-ovaire et pro-testicule, il est certainement plus adapté de considérer les réseaux de régulation déterminant du testicule ou de l'ovaire comme des réseaux complexes opposés à différents niveaux. En effet, c'est dans la définition même d'un déterminant du sexe que son inactivation entraîne une inversion du sexe partielle ou complète, due à l'activation de sa voie transcriptionnelle alternative. Dans ce sens, tous les gènes pro-testicules sont également définis comme des gènes anti-ovaires, leur inhibition ou mutation induisant des phénotypes d'inversion du sexe « male-to-female ». Cependant, dans le cas des gènes pro-ovariens, comme *WNT4*, *RSPO1* ou *FOXL2*, l'inversion du sexe induite par leur inhibition apparaît partielle, suggérant une complexité plus importante dans les voies de différenciation femelle, au niveau moléculaire ou cellulaire. Il apparaît nécessaire d'effectuer de plus amples recherches permettant de décrypter les mécanismes antagonistes entre la différenciation masculine et féminine, y compris la découverte de nouveaux gènes déterminants du sexe.

De plus, ce réseau de régulation et de maintien de « l'identité gonadique » semble être requis même à l'âge adulte. Chez la souris, des expériences d'inhibition conditionnelle ont mis en évidence une reprogrammation des populations cellulaires somatiques et une réorganisation de la gonade. La délétion ubiquitaire de *Foxl2* dans l'ovaire adulte entraîne une reprogrammation des cellules de la granulosa, l'apparition de cellules apparentées à des cellules de Sertoli, exprimant des marqueurs tels que *Sox9* et *Dmrt1*, associés à des cellules de Leydig produisant de la testostérone et les structures folliculaires disparaissent en faveur de structures similaires à des tubules séminifères (Uhlenhaut et al., 2009). Dans le testicule adulte, l'inactivation de *Dmrt1* (Minkina et al., 2014) aboutit également à une inversion du sexe. Ces expériences révèlent donc une étonnante plasticité des gonades murines.

Des études récentes ont également démontré que, au-delà du programme transcriptionnel spécifique des voies mâles et femelles de la différenciation gonadique, le cycle cellulaire serait une composante à part entière du processus. En effet, parmi les premiers mécanismes moléculaires activés après l'expression de SRY, la machinerie de prolifération des cellules de Sertoli se met en place, en parallèle du recrutement paracrine des cellules somatiques adjacentes. Au contraire, l'engagement des cellules vers une différenciation femelle, en cellules de la granulosa, est associé avec une sortie du cycle cellulaire médiée par p27 (Maatouk et al., 2013; Mork et al., 2012).

Pendant longtemps, l'entrée dans un cycle cellulaire était considérée comme une conséquence directe de la différenciation cellulaire, mais des recherches ont démontré que des cycles cellulaires défectueux peuvent précéder et influencer les processus de différenciation (Lange & Calegari 2010; Singh & Dalton 2009). L'hypothèse est que les différentes phases du cycle cellulaire représentent des « fenêtres d'opportunité » pour la différenciation d'une cellule. Cela a été illustré par les expériences menées par Mummery et collaborateurs, démontrant que la différenciation des cellules de carcinome embryonnaires ne pouvait être induite par l'acide rétinolique seulement par une exposition pendant la phase G1 du cycle cellulaire, et non pendant la phase S (Mummery et al., 1987).

Il devrait donc être considéré que la capacité d'une cellule à se diriger vers une voie de différenciation donnée peut être dépendante de cette « fenêtre d'opportunité », qui permet l'interaction de la machinerie du cycle cellulaire et des facteurs développementaux qui affectent l'activité et la disponibilité des signaux, récepteurs de surfaces et facteurs de transcription impliqués dans ces processus de différenciation. La machinerie transcriptionnelle du cycle cellulaire devrait donc être considérée comme un facteur à part entière impliqué dans la différenciation sexuelle (revu dans Bunce and Capel, 2016).

1.2.3.2. Devenir des structures embryonnaires génitales

La stratégie de différenciation des gonades, où une structure bipotentielle s'engagera vers une destinée masculine ou féminine, est différente des autres stratégies différenciant les hommes et les femmes à l'âge adulte. En effet, les canaux de Muller et de Wolff sont des structures primaires complètement distinctes au début de la grossesse et l'un dégénérera en faveur de l'autre en fonction du sexe biologique, mais rempliront des fonctions analogues d'acheminement des gamètes. Le tissu précurseur dégénéré persistera à l'âge adulte en tant que vestige embryonnaire. De plus, les organes externes du tractus génital proviennent de structures primaires communes, comme les gonades, mais resteront similaires au niveau de la différenciation des différents types cellulaires, seulement structurées et façonnées de manière différente. Seule la différenciation du testicule et de l'ovaire sera détaillée dans ce manuscrit.

Voie Male	Structure embryonnaire	Voie Femelle
Testicule	Gonade indifférenciée	Ovaire
Tubules séminifères	Cortex	Follicules ovariens
Rete testis	Médullaire	Rete ovarii
Gubernaculum	Ligament inguinal	Ligament utéro-ovarien et ligament rond
Canalicules efférents, Paradidyme	Tubules mésonéphrotiques	Paroophore, Epoophore
Appendice de l'épididyme, Épididyme, Canal déférent, Canal éjaculateur, Vésicule séminale	Canal de Wolff	Canal de l'époophore, Canal de Gartner, Hydatide de Morgani
Appendice testiculaire	Canal de Müller	Hydatides, Utérus, Trompes utérines, Vagin
Urètre, Prostate, Glandes bulbo-urétrales	Sinus urogénital	Urètre, Vestibule du vagin, Glandes urétrales, para-urétrales et vestibulaires
Pénis	Tubercule génital	Clitoris
Corps caverneux, Corps spongieux	Plis urogénitaux	Petites lèvres
Scrotum	Bourrelets labo-scrotaux	Grandes lèvres

Tableau I. Homologies entre les structures embryonnaires génitales et leurs dérivés différenciés dans chaque sexe (vestiges embryonnaires)

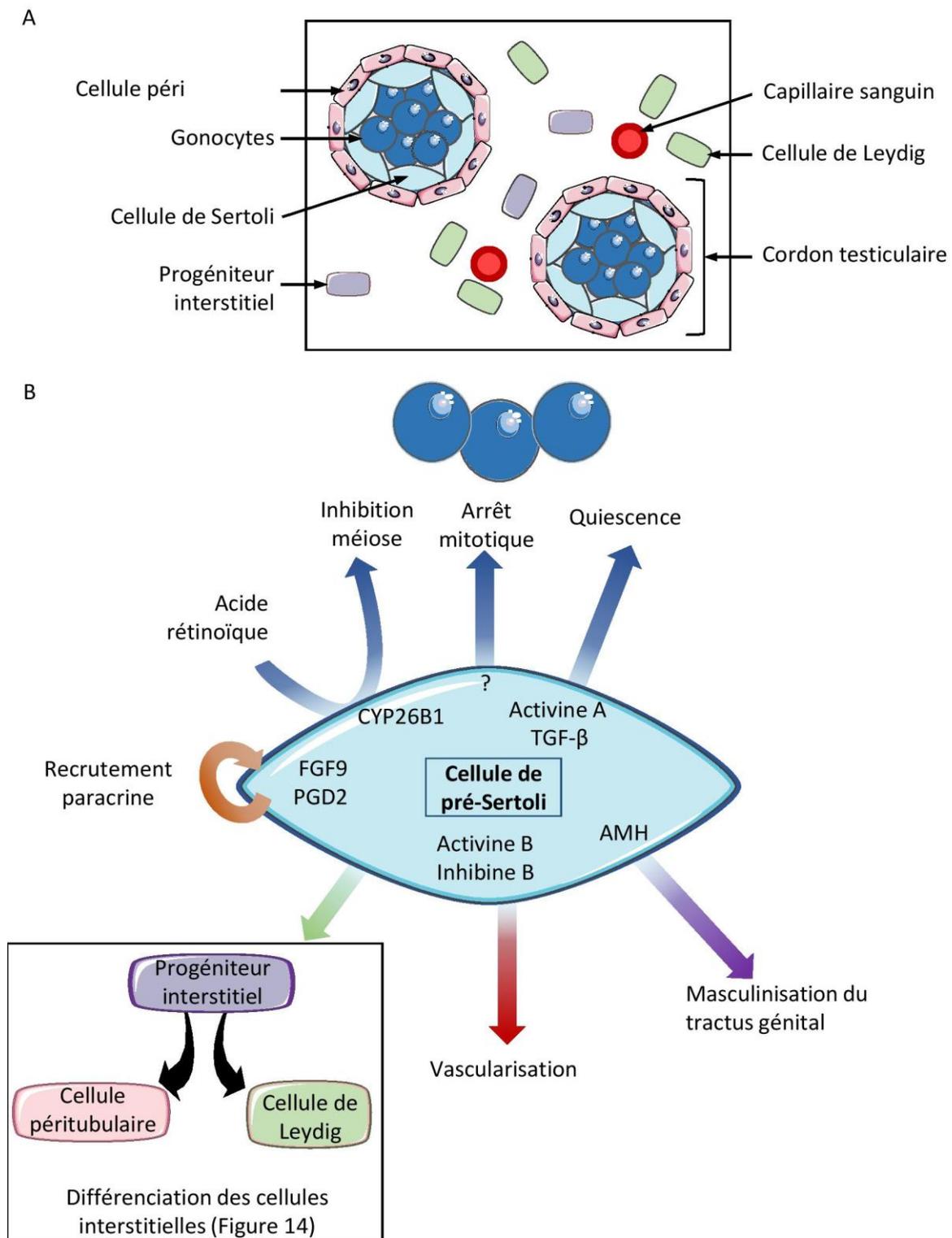


Figure 13. (A) Schéma des différents compartiments tissulaires du testicule fœtal : les cordons testiculaires et le tissu interstitiel. (B) Rôle central de la cellule de Sertoli dans la différenciation des autres populations cellulaires du testicule et dans le développement testiculaire.

1.2.4. Le développement gonadique

1.2.4.1. Développement testiculaire

Au début du développement des testicules, ceux-ci se trouvent en position abdominale, attachés au mésonéphros, correspondant aux reins fœtaux et contenant les structures qui se différencieront en tractus génital masculin (canaux de Wolff) ou féminin (canaux de Muller) suite à la différenciation des gonades vers l'une ou l'autre des voies.

1.2.4.1.1. *Les cordons sexuels : tubules séminifères en devenir*

La formation des cordons sexuels au cours de la 7^{ième} SD est l'évènement définissant la différenciation testiculaire. Les cellules de Sertoli vont alors former des ponts membranaires intercellulaires, s'agréger autour des groupes de cellules germinales et acquérir leur polarisation. Dans les gonades déficientes en cellules germinales, les cordons testiculaires se développent, indiquant que leur formation est dirigée par les interactions entre les cellules supportrices plutôt que par les cellules germinales (Escalante-Alcalde and Merchant-Larios, 1992; McCoshen, 1982, 1983; Merchant, 1975). Dès lors, si les cellules germinales ne sont pas requises, les signaux leur permettant d'être entourées par les cellules de Sertoli restent flous. De plus, les mécanismes moléculaires permettant aux cellules de Sertoli de se structurer en forme tubulaire et non en amas, sont encore aujourd'hui mystérieux.

Dans les cordons sexuels, les cellules germinales se retrouvent dans toute la lumière du tube (Figure 13A). Cependant, à la puberté, les cordons se transforment en tubules séminifères fonctionnels : les cellules de Sertoli cessent de se diviser, s'allongent en formant des jonctions serrées entre elles, et forment la barrière hémato-testiculaire. Les cellules germinales se retrouvent vers la membrane basale du tubule et les Sertoli commencent à sécréter le fluide tubulaire.

1.2.4.1.1.1. Cellules germinales

Les cellules germinales que l'on retrouve dans le testicule fœtal ont pour origine les cellules germinales primordiales ayant migré dans la crête génitale pendant le stade indifférencié de la gonade, désignées alors sous le nom de gonocytes (voir « Origine et migration des cellules germinales »).

Dans le testicule fœtal humain, trois populations distinctes de cellules germinales coexistent entre la 9^{ième} et la 30^{ième} semaine de développement (SD) : les gonocytes, les cellules intermédiaires et les spermatogonies fœtaux (Fukuda et al., 1975). Les gonocytes sont présents dès la différenciation gonadique, à partir de la 9^{ième} SD. Puis à partir de la 10^{ième} SD, les cellules intermédiaires apparaissent et se développent jusqu'à devenir les cellules germinales dominantes dans le testicule à la 15^{ième} SD.

C'est à ce moment que les spermatogonies fœtales apparaissent et se divisent jusqu'à la 22^{ème} SD, où elles deviennent à leur tour les cellules majoritaires, alors que les gonocytes disparaissent.

Au niveau moléculaire, ces trois populations cellulaires ont pu être distinguées grâce à leurs transcriptomes propres. En effet, plusieurs études ont pu cartographier l'expression de marqueurs spécifiques à chaque stade de différenciation chez l'homme grâce à l'immuno-histochimie (Anderson et al., 2007; Gaskell et al., 2004; Reijo et al., 2000; Skakkebaek et al., 2016), ou au RNA-seq (Gkoutela et al., 2015). Ainsi, caractérisation des différentes populations germinales ont notamment permis de mettre en évidence les potentiels gènes induisant l'arrêt mitotique des cellules germinales, telles que *CDKN2B* et *CDK6*. *NANOS2* prévient l'entrée en méiose en inhibant l'expression de *STRA8* (voir « Oocytes – méiose ») (Feng et al., 2014; Suzuki and Saga, 2008).

1.2.4.1.1.2. Cellules de Sertoli

Les cellules de Sertoli ont un rôle essentiel dans le testicule mature : ce sont les cellules supportrices de la spermatogenèse, elles permettent la différenciation complète de la lignée germinale jusqu'à la production de spermatozoïdes dans la lumière des tubules séminifères (voir « La fonction exocrine des gonades adultes »). La différenciation et le développement des cellules de Sertoli pendant la vie fœtale sont des étapes clés pour la future santé reproductive de l'homme, car leur nombre final dans le testicule mature conditionne la qualité de la spermatogenèse (Petersen and Söder, 2006). Elles ont également un rôle central dans la différenciation et le développement du testicule fœtal et des autres populations cellulaires qui le composent, notamment les cellules stéroïdogènes (Figure 13B).

- *Différenciation des cellules de Sertoli*

Les cellules de Sertoli se différencient dès les premières semaines de la vie fœtale. Comme décrit précédemment, elles sont les actrices majeures de la différenciation du testicule, exprimant les premiers effecteurs de la différenciation gonadique mâle, *SRY* et *SOX9*. Les processus moléculaires décrits dans la section « Trajectoire mâle » font partie intégrante des processus de différenciation. En effet, la maintenance de la différenciation des Sertoli et leur recrutement paracrine sont des processus capitaux pour la différenciation testiculaire. Cette phase rapide d'expansion des cellules de Sertoli a lieu vers la 6^{ème} et 7^{ème} SD dans le testicule humain (Ostrer et al., 2007) et elles continueront leur prolifération jusqu'à la 14^{ème} SD. D'autres phases de développement de la population Sertolienne auront lieu pendant la vie néonatale et prépubertaire (Jørgensen and Rajpert-De Meyts, 2014).

- *Rôle pendant la vie fœtale*

Masculinisation du tractus génital : l'AMH

En plus des facteurs permettant le recrutement paracrine des cellules de Sertoli, celles-ci ont une activité endocrine, sécrétant l'AMH, hormone indispensable à la masculinisation du tractus génital, via l'induction de la régression des précurseurs des voies féminines (canaux de Muller) et le développement des voies masculines (canaux de Wolff). Elle appartient à la même superfamille que le TGF- β (Lane and Donahoe, 1998) et se fixe à un récepteur membranaire de type II, l'AMHR2, exprimé sur les cellules mésenchymateuses autour des canaux de Muller (Baarends et al., 1994).

L'expression de l'AMH est induite par la régulation transcriptionnelle de SOX9, WT1 et NR5A1 (SF1) (Arango et al., 1999; Nachtigal et al., 1998; Tremblay and Viger, 1999). Des cas de syndromes des canaux de Muller persistants ont été décrits dans le cas de mutation dans le gène de l'AMH ou de son récepteur, AMHR2 (Imbeaud et al., 1994, 1995; Knebelmann et al., 1991). De plus, les cellules de Leydig expriment l'AMHR2 (Racine et al., 1998) et l'AMH figure comme un inhibiteur de la transcription de *CYP17A1*, enzyme clé de la stéroïdogénèse (Teixeira et al., 1999). La délétion du récepteur chez des souris mutantes entraîne une hyperplasie des cellules stéroïdogènes (Behringer et al., 1994). Ces résultats suggèrent donc que l'AMH jouerait un rôle d'inhibiteur paracrine de l'activité stéroïdogène des cellules de Leydig fœtales. Des expériences *in vitro* ont également démontré sa contribution dans l'induction de la migration des cellules du mésonéphros dans le testicule en développement et donc dans la vascularisation testiculaire (Ross et al., 2003; Vigier et al., 1987).

Quiescence des cellules germinales

Les cellules de Sertoli interviennent également dans la différenciation et la survie de la lignée germinale. Il a été démontré que l'entrée en méiose des ovogonies, cellules germinales de l'ovaire, est contrôlée par l'acide rétinoïque (voir chapitre « Oocytes – méiose »). Dans le testicule, la cellule de Sertoli, voisine des cellules germinales, exprime *CYP26B1*, qui catabolise l'acide rétinoïque et inhibe leur entrée en méiose (Bowles et al., 2006; Koubova et al., 2006). Elles expriment également *FGF9*, qui réduit la réceptivité des cellules germinales à l'acide rétinoïque (Bowles et al., 2010) et se fixe sur le corécepteur de Nodal, qui participe au maintien de leur pluripotence (Bowles and Koopman, 2010; Spiller et al., 2012). Chez la souris, l'activine A et le TGF- β sécrétés par ces cellules supportrices sont impliqués dans la régulation de leur prolifération et le maintien de leur quiescence (Mendis et al., 2011; Moreno et al., 2010).

Vascularisation spécifique du testicule

On a vu précédemment que les migrations des différentes populations cellulaires précédant la différenciation sexuelle étaient fondamentales pour le modelage et le développement de la crête génitale puis de la gonade (voir « La crête génitale »). Plus tard, la migration des cellules endothéliales du mésonéphros est un processus spécifique de la gonade mâle et fondamentale pour la formation le vaisseau cœlomique, principale artère testiculaire qui persistera à l'âge adulte. Cette migration est initiée grâce à la sécrétion d'activine B et de son monomère, l'inhibine B par les cellules de Sertoli. L'expression de ces gènes est inhibée par les signaux WNT lors de la différenciation ovarienne (Yao et al., 2006).

Régulation paracrine de la différenciation des autres populations cellulaires somatiques

Les cellules de Sertoli sécrètent également DHH, facteur de signalisation, dont le récepteur PTCH1 est exprimé par les cellules interstitielles environnantes, précurseurs des autres populations somatiques (Bitgood and McMahon 1995, Pierucci-Alves et al. 2001). De ce fait, DHH pourrait amplifier la différenciation des cellules de Leydig et pérیتubulaires (Svingen and Koopman, 2013), ces dernières étant impliquées dans la formation des cordons sexuels. Chez la souris, une invalidation du gène entraîne l'altération de la différenciation des cellules de Leydig et des cellules pérیتubulaires (Clark, et al. 2000, Yao, et al. 2002). Chez l'Homme, les mutations sont associées avec des agénésies gonadiques (Canto, et al. 2004, Canto, et al. 2005, Umehara, et al. 2000).

Implication dans la stéroïdogénèse

La production de testostérone est communément attribuée exclusivement aux cellules de Leydig dans le testicule humain adulte. Dans le testicule fœtal, des études effectuées chez la souris ont su mettre en évidence une possible coopération entre les cellules de Leydig fœtales et les cellules de Sertoli. En effet, l'étude des cellules de Leydig isolées des autres cellules somatiques testiculaires dans des souris Sf1-GFP a permis de montrer que, si la grande partie de la machinerie de la stéroïdogénèse est bien exprimée par les cellules de Leydig, le gène HSD17B3 est quant à lui exprimé par les cellules de Sertoli (Shima et al., 2013). Ces résultats laissent donc supposer l'implication de ces dernières dans la production de la testostérone.

1.2.4.1.1.3. Cellules pérیتubulaires

Les cellules pérیتubulaires sont des cellules présentant des caractéristiques similaires aux myofibroblastes, des cellules musculaires. L'origine des cellules mésenchymateuses précurseurs des cellules pérیتubulaires est encore mal décrite, mais l'hypothèse fut émise qu'elles partageraient une origine commune avec les cellules de Leydig (Wainwright and Wilhelm, 2010) (voir « Cellules de

Leydig "). Une limitation majeure de l'étude de leur origine et des processus de différenciation est l'absence de marqueur spécifique à cette population cellulaire au début du développement fœtal. Chez la souris, même si leur morphologie est facilement reconnaissable à partir de 13,5 jpc, les marqueurs identifiés sont également exprimés par toutes les cellules interstitielles (Jeanes et al., 2005). Les cellules de Leydig et péricubulaires expriment toutes deux PTCH1 (Yao et al., 2002), récepteur du facteur de signalisation DHH sécrété par les cellules de Sertoli. Des études sur des souris KO du le gène *Dhh* ont observé une altération des cellules péricubulaires et de la différenciation de Leydig fœtale (Clark et al., 2000; Pierucci-Alves et al., 2001; Yao et al., 2002). Les cellules de Sertoli semblent donc avoir un rôle dans la différenciation des cellules péricubulaires.

Après la formation des cordons sexuels primitifs, ces cellules allongées vont se différencier et participer de concert avec les cellules de Sertoli pour former la membrane basale des cordons sexuels et progressivement acquérir leur capacité contractile. Elles deviennent reconnaissables au cours de la 10^{ème} SD (Ostrer et al., 2007), où elles forment alors une mono-couche. Plus tard, elles se différencieront pour former une bicouche autour des cordons testiculaires.

1.2.4.1.2. Tissu interstitiel

L'espace ou tissu interstitiel est un tissu conjonctif lâche contenant les vaisseaux sanguins et lymphatiques. Il contient également les cellules de Leydig, en charge de la synthèse des androgènes, des fibroblastes, des éléments figurés du sang et des cellules immunitaires telles que macrophages et mastocytes.

1.2.4.1.2.1. Progéniteurs

Les cellules précurseuses des futures cellules interstitielles proviennent de l'épithélium cœlomique suite à son ingression dans la gonade. Ces cellules expriment *WT1*, mais contrairement aux cellules de Sertoli qui initient l'expression de *SOX9* et induisent la détermination sexuelle, ces cellules vont activer le signal Notch et réprimer l'expression de *WT1* (Liu et al., 2016; Tang et al., 2008). D'autres cellules précurseuses des cellules interstitielles proviennent du mésonéphros voisin, migrant vers la gonade (Cool et al., 2008; DeFalco et al., 2011; Jeays-Ward, 2003; Martineau et al., 1997; Merchant-Larios et al., 1993). Bien que le rôle désigné de ce compartiment tissulaire soit la stéroïdogénèse, la majorité des cellules qui le composent ne sont pas des cellules stéroïdogéniques.

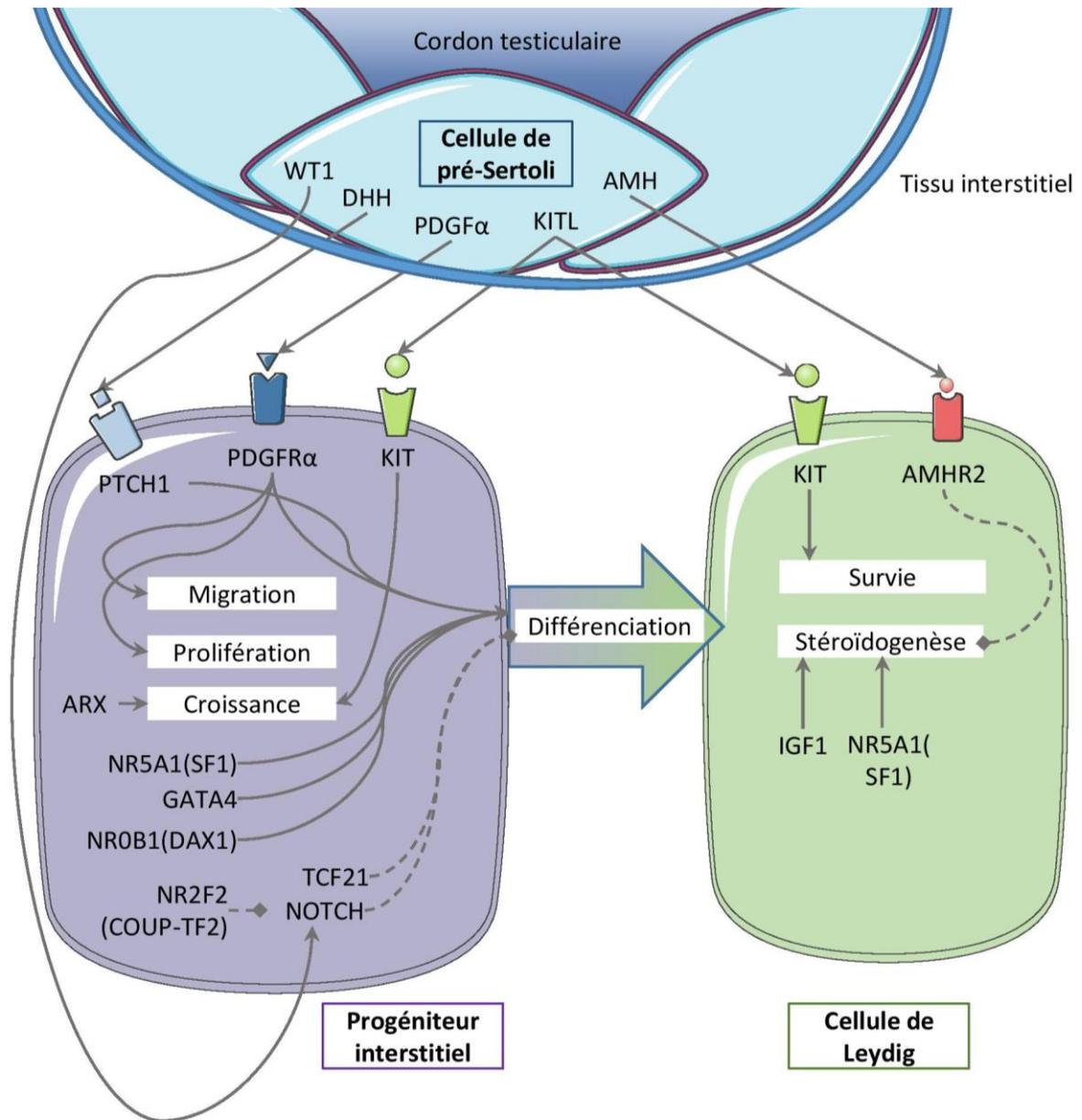


Figure 14. Mécanismes de régulation de la différenciation des cellules de Leydig fœtales (adapté de Wen et al., 2016)

Elles n'expriment ni les enzymes impliquées dans cette voie ni NR5A1(SF1), mais expriment le récepteur Hedgehog PTCH1 et d'autres marqueurs associés aux précurseurs, comme ARX et NR2F2 (COUP-TF2) (van den Driesche et al., 2012; Kilcoyne et al., 2014).

1.2.4.1.2.2. Cellules de Leydig

Dans le testicule fœtal humain, cet ensemble de cellules précurseuses est important, car elles contrôlent et évaluent le nombre de cellules de Leydig permettant le développement sexuel mâle via la testostérone. La différenciation des cellules stéroïdogéniques est sous le contrôle d'une balance entre inhibition et activation. La maintenance du stock de cellules précurseuses dans le tissu interstitiel semble être contrôlée par le gène *PODI (TCF21)*, qui réprime l'expression de *NR5A1 (SF1)* (Cui et al., 2004). ARX quant à lui participe à la survie des précurseurs, contribuant indirectement à la différenciation des cellules de Leydig (Miyabayashi et al., 2013). Chez l'homme, des mutations sur ce gène sont associées à des phénotypes de « sous-virilisation » (Kitamura et al., 2002; Sirisena et al., 2014). Le système Notch est impliqué dans le contrôle de la balance entre cellules précurseuses et cellules de Leydig, car son inactivation entraîne une augmentation des cellules de Leydig fœtales et une diminution des cellules précurseuses (Liu et al., 2016; Tang et al., 2008) chez la souris. Chez l'homme, des processus similaires semblent être en jeu dès lors qu'une mutation dans le gène *MAMLD1*, responsable de l'activation des gènes cibles de NOTCH, a été associée à une dysgénésie gonadique complète 46,XY (Ruiz-Arana et al., 2015), telle que l'hypospadias (Fukami et al., 2006). Enfin, le nombre de précurseurs exprimant *NR2F2 (COUP-TF2)* diminue lorsque les cellules de Leydig commencent à se développer, suggérant un recrutement des précurseurs vers la voie des Leydig (van den Driesche et al., 2012). Son rôle dans le contrôle du ratio cellules stéroïdogènes/non-stéroïdogènes pendant la vie fœtale n'est pas encore défini.

Lorsque ces cellules s'engagent dans la voie de différenciation des cellules de Leydig, elles inhibent progressivement l'expression de marqueurs de cellules précurseuses pour sur-exprimer *NR5A1 (SF1)* et des enzymes stéroïdogènes (DeFalco et al., 2011; Liu et al., 2016; Miyabayashi et al., 2013).

- *Différenciation des cellules de Leydig fœtales*

L'activation de la différenciation des cellules de Leydig est contrôlée par des signaux paracrines des Sertoli, place initiale de la différenciation testiculaire, à partir de la 7^{ème} SD (Figure 14). DHH étant sécrété par les Sertoli, ciblant le récepteur PTCH1 sur les cellules précurseuses du tissu interstitiel. Cela entraîne l'expression des facteurs de transcription GLI, dont GLII.

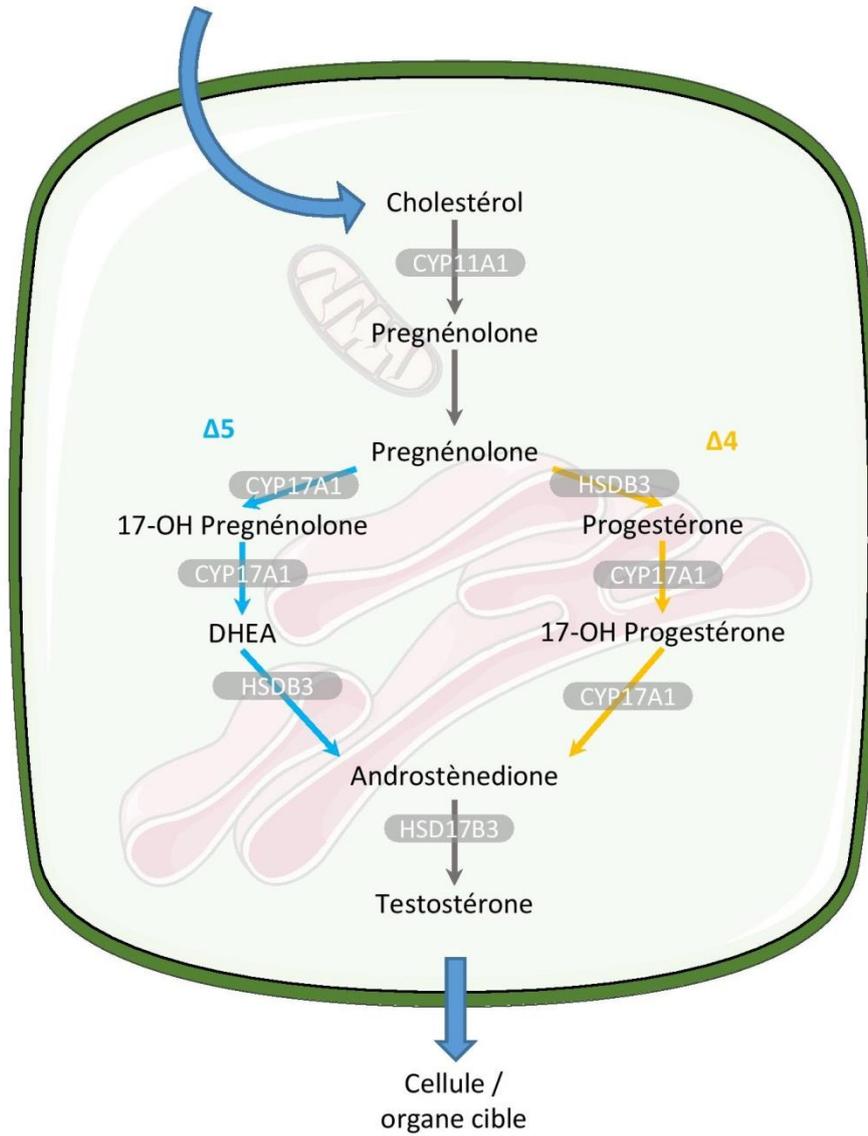


Figure 15. Stéroïdogénèse dans la cellule de Leydig.

Chez l'homme, l'expression de ce gène coïncide avec la différenciation des cellules de Leydig (Mamsen et al., 2017). DHH induit également l'expression des marqueurs de cellules de Leydig, *NR5A1 (SF1)* et *CYP11A1* (Park et al., 2007; Yao et al., 2002). Chez l'homme, des mutations dans les différents composants de la voie de signalisation de Hedgehog sont impliquées dans de nombreux cas de dysgénésie gonadique (Callier et al., 2014; Canto et al., 2004, 2005; Das et al.; Umehara et al., 2000; Werner et al., 2015). Cependant, DHH n'est pas le seul facteur induisant la différenciation des cellules de Leydig, étant donné que certaines d'entre elles se différencient même en son absence (Liu et al., 2016; Yao et al., 2002). La voie des PDGF (Platelet-derived growth factor) est également impliquée. Le PDGFA sécrété également par les cellules de Sertoli trouve son récepteur exprimé par les cellules interstitielles du testicule et les souris mutantes délétées du récepteur *Pdgfra* montrée une altération des cellules de Leydig fœtales (Brennan et al., 2003). Elles atteignent leur pic de développement entre la 17^{ième} et la 21^{ième} SD (Svechnikov et al., 2010).

- *Rôle endocrine pendant la vie fœtale*

Synthèse des stéroïdes

La stéroïdogénèse correspond à une succession de réactions enzymatiques qui transforment le cholestérol en stéroïdes (Figure 15). Ces réactions sont principalement catalysées par plusieurs cytochromes P450 (CYP) et des hydroxystéroïdes déshydrogénases (HSD).

Le testicule fœtal humain exprime les récepteurs aux LDL (low density lipoprotein) à partir de la 10^{ième} SD et le récepteur SRB1 (SCARB1) aux HDL (high-density lipoprotein), son expression étant régulée par *NR5A1 (SF1)* (Cao et al., 1999). La gonade est donc capable de synthétiser *de novo* le cholestérol à partir de lipoprotéines circulantes entre la 10^{ième} et 20^{ième} SD (Carr et al., 1983). Globalement, les cellules de Leydig disposent de plusieurs alternatives pour obtenir le cholestérol nécessaire à la stéroïdogénèse, ce qui pourrait expliquer pourquoi l'inhibition d'une de ces voies n'a pas d'effet majeur sur la synthèse de testostérone (Scott et al., 2009).

Le cholestérol est tout d'abord transporté vers la mitochondrie par StAR (Stocco 2001, Stocco and Clark 1996), dont l'expression est régulée par la LH (Ascoli et al., 2002) et *NR5A1(SF1)* (Dubé et al., 2009), puis il est converti en pregnénolone grâce au clivage de sa chaîne latérale par *CYP11A1* (Hume et al., 1984). La pregnénolone produite passe de la mitochondrie au réticulum endoplasmique lisse où le reste de la stéroïdogénèse se produit (Zirkin and Chen, 2000). Il existe deux voies principales de biosynthèse utilisant la 3 β -HSD (*HSD3B2*) et *CYP17A1*. D'une part, la voie $\Delta 5$ produit les intermédiaires 17 α -hydroxypregnénolone et la DHEA (Déhydroépiandrosterone) et d'autre part, la voie $\Delta 4$ produit la progestérone et la 17 α -hydroxyprogestérone (Coffey et al., 1972).

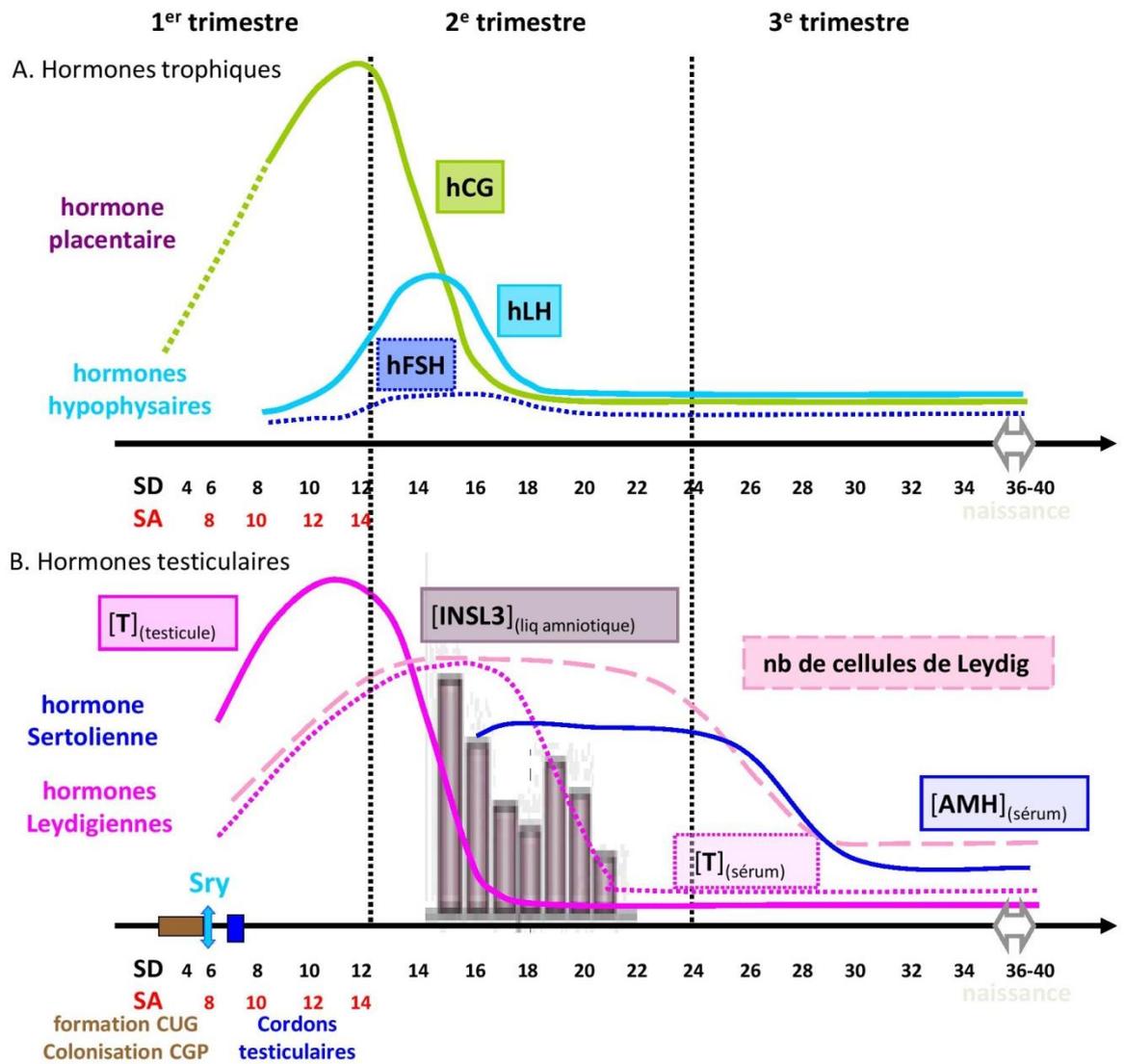


Figure 16. Évolution du nombre de cellules de Leydig, de la quantité de testostérone (testiculaire et plasmatique) et des taux de gonadotrophines sériques au cours de la vie fœtale chez l'Homme.

SD : Semaine de développement embryonnaire ; SA : Semaine d'aménorrhée ; T : testostérone ; INSL3 : insulin-life factor 3 ; AMH : hormone anti-müllérienne ; hCG : hormone chorionique humaine ; LH : hormone lutéinisante ; FSH : hormone folliculo-stimulante. (Adapté de Rouiller-Fabre et al., 2009).

Ces deux voies aboutiront à la formation d'androsténédione, elle-même transformée en testostérone par 17β -HSD (*HSD17B3*) (Payne and Hales, 2004). En fonction de l'espèce ou de l'organe, la biosynthèse des hormones sexuelles suit la voie $\Delta 5$ ou $\Delta 4$. Chez l'homme, les gonades privilégient la voie $\Delta 5$, alors que les glandes surrénales privilégient la voie $\Delta 4$, par déficience partielle ou totale en CYP17A1. Chez les rongeurs, la voie $\Delta 4$ est prépondérante, tandis que chez les primates, les porcs et le lapin, c'est la voie $\Delta 5$ (Fluck, et al. 2003, Mathieu, et al. 2002). Chez la souris, *Hsd17b* est exprimé dans les cellules de Sertoli fœtales (Shima et al., 2013), ce qui laisse supposer une coopération entre ces deux populations cellulaires somatiques lors de la stéroïdogénèse. Les mécanismes en jeu pendant le développement du testicule humain n'ont pas encore été définis.

Une voie alternative de la stéroïdogénèse aboutit à la production de la dihydrotestostérone (DHT), stéroïde puissant capital pour la masculinisation du tractus génital externe mâle, a été identifiée chez le jeune marsupial (Wilson, et al. 2003). L'androstanediol est convertie en DHT par une étape qui court-circuite la biosynthèse classique des androgènes. L'identification de mutations chez l'Homme du gène *AKR1C2* (Fluck, et al. 2011) s'accompagnant de désordres de virilisation du tractus génital externe a ouvert la possibilité de l'existence de cette voie alternative de synthèse de DHT, métabolite plus actif que la testostérone, dans le testicule fœtal humain. Cette voie n'est à ce jour pas complètement caractérisée chez l'Homme.

Synthèse de l'insulin-like factor 3

La production d'insuline-like factor 3, INSL3, par les cellules de Leydig fœtales débute entre la 9^{ème} et la 11^{ème} SD, ce qui est concomitant avec la première détection des androgènes fœtaux. Pendant cette période, INSL3 a pour rôle d'induire la première phase de descente testiculaire, appelée phase abdomino-inguinale, entre la 8^{ème} et la 16^{ème} SD (Payne and Hales, 2004). Avant l'initiation de la descente, le testicule est relié par sa face ventrale à la région inguinale grâce à un ligament fœtal appelé gubernaculum. Les cellules du gubernaculum expriment *RXFP2*, récepteur à l'INSL3 qui sera produite par les cellules de Leydig. Lors de son activation, *RXFP2* couplé à une protéine Gs active l'adénylate cyclase et donc la production d'AMPc. Dès lors, le ligament va s'épaissir, son élasticité va diminuer, orientant ainsi le testicule vers la région inguinale. L'INSL3 possède une action synergique avec les autres androgènes pendant cette étape cruciale du développement sexuel (Kubota et al. 2002). Bien que le rôle central de l'INSL3 pour la descente testiculaire soit établi, peu de mutations sur ce gène ont été associées à des cas de cryptorchidie (revues dans Foresta, et al. 2008), la caractérisation des facteurs induisant une descente complète des testicules reste à établir.

1.2.4.1.3. Régulation endocrine : l'axe hypothalamo-hypophysaire gonadique

La GnRH est une hormone synthétisée par des neurones spécialisés de l'hypothalamus à partir de la 10^{ème} SD. Elle est sécrétée de manière pulsatile par l'extrémité axonale des neurones hypothalamiques, ciblant les cellules gonadotrope de l'hypophyse (Kaprra and Huhtaniemi, 2017).

Sous l'action de la GnRH, les cellules gonadotropes de l'antéhypophyse libèrent deux hormones glycoprotéiques dimériques dans la circulation sanguine : la LH et la FSH (Kaprra and Huhtaniemi, 2017). Ces deux hormones ont un rôle majeur dans la régulation de la fonction endocrine testiculaire. Toutes les gonadotrophines (LH, FSH) et l'hCG partagent une même sous unité α et se différencient par leurs sous unités β spécifiques.

L'hCG est une hormone indispensable à l'établissement et au maintien de la grossesse (Rahman and Rao, 2009). Sa sécrétion par le trophoblaste (composant de l'embryon) dès le 7^{ème} jour de grossesse et par le placenta est utilisée comme marqueur de l'établissement de la grossesse. Ses niveaux augmentent ensuite, jusqu'à atteindre un maximum aux alentours de la 10^{ème} SD. Il s'ensuit une diminution jusqu'au 3^{ème} mois, les niveaux restant stables pour le reste de la grossesse. Il a été montré *in vitro* que l'hCG était indispensable aux cellules de Leydig pour produire un niveau de testostérone suffisant pour assurer la masculinisation du tractus génital masculin (Rabinovici and Jaffe, 1990).

la LH devient détectable à partir de la 12^{ème} SD, alors que le niveau d'hCG commence à décroître, et rencontrera son pic de production au 16^{ème} SD (Clements et al., 1976). C'est entre la 15^{ème} et la 20^{ème} SD que la production des androgènes (plus particulièrement de la testostérone) et sa régulation deviennent dépendantes de l'activité de la glande hypophysaire du fœtus (Lambrot et al., 2009; O'Shaughnessy et al., 2007; del Valle et al., 2017).

Au niveau fœtal, la LH hypophysaire et l'hCG placentaire agissent sur le récepteur testiculaire LHCGR exprimé par les cellules de Leydig pour stimuler la sécrétion de testostérone, mais aussi d'INSL3 (Huhtaniemi et al., 1977; Toppari et al., 2007). Le récepteur LHCGR est donc fondamentalement important pour un développement normal de la fonction de reproduction chez le mâle. Des mutations délétères sont associées à des ADS, dues à un manque d'androgènes (Themmen and Huhtaniemi, 2000). Des mesures d'affinité pour le récepteur ont montré que l'hCG possède 2 à 6 fois d'affinité que la LH, suggérant que lors de la concomitance de production des deux hormones entre la 15^{ème} et la 20^{ème} SD, l'hCG jouerait un rôle plus important que la LH dans la stéroïdogénèse (Dufau et al., 1972; Lee and Ryan, 1973). Contrairement au phénomène de désensibilisation aux gonadotrophines observé dans le testicule adulte, via l'internalisation du LHCGR suite à une

exposition prolongée aux hormones, les cellules de Leydig fœtales ne semblent pas suivre le même mécanisme de régulation (Huhtaniemi et al., 1981, 1982; Warren et al., 1982).

La FSH a un rôle dans la régulation de l'activité paracrine des cellules de Sertoli adultes (voir « L'axe hypothalamo-hypophysaire »), mais son implication sur le développement testiculaire fœtal n'est pas clairement établie (Tostain et al., 2004). Le rôle des hormones dérivées de la glande pituitaire n'est pas réellement bien établi, mais il apparaît comme étant limité à la dernière partie de la grossesse.

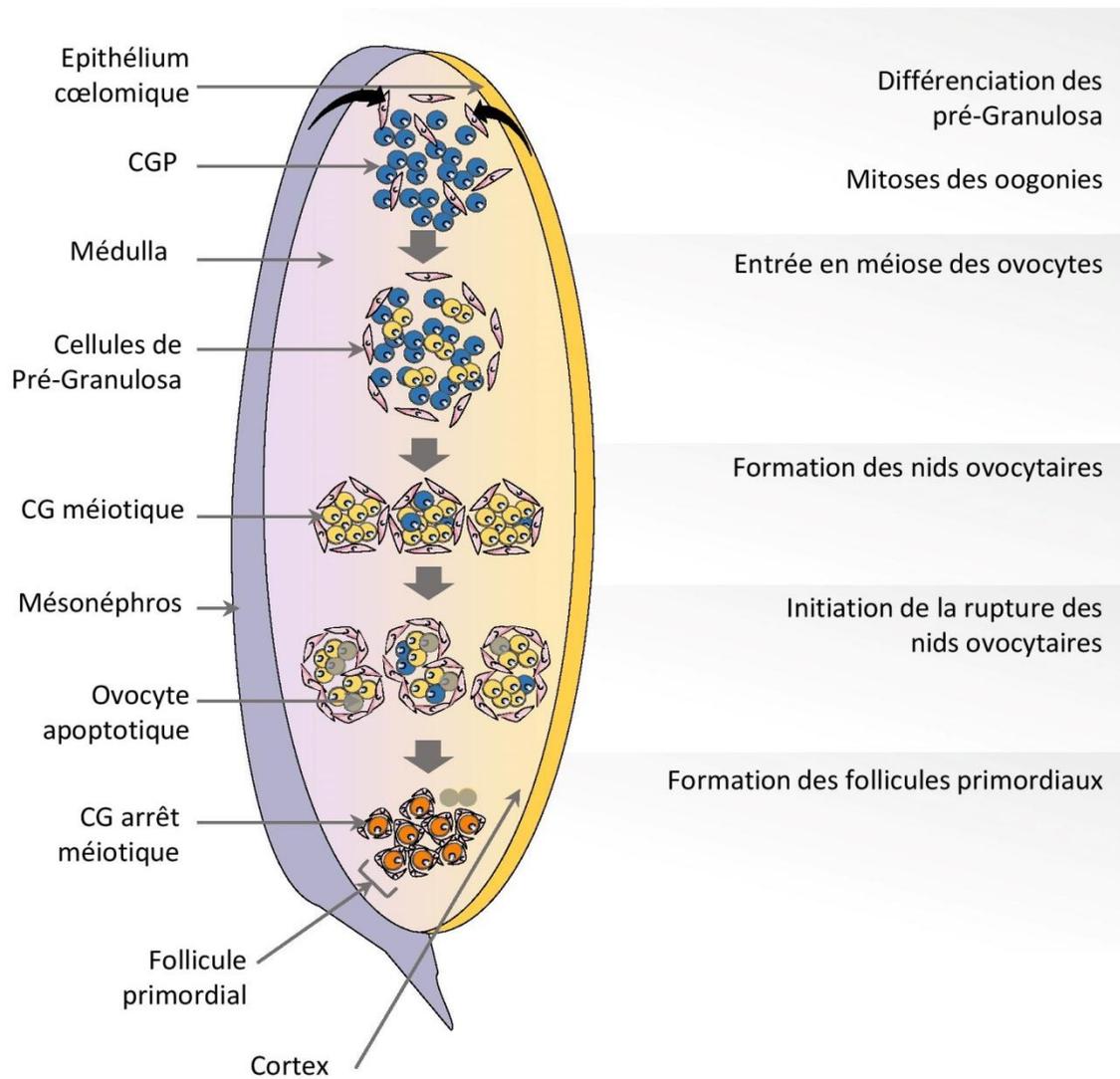


Figure 17. Formation des nids ovocytaires et des follicules primordiaux dans l'ovaire foetal.

1.2.4.2. Développement ovarien

Pendant la vie fœtale, la différenciation ovarienne s'effectue en deux étapes morphogénétiques distinctes, la formation des cordons ovariens puis la formation des follicules primordiaux. Après la différenciation sexuelle, le testicule et l'ovaire peuvent être rapidement distingués par leur taille. En effet, le testicule subira rapidement une réorganisation structurale drastique (cordons testiculaires et vascularisation), qui entraînera une augmentation de sa taille. La structure de l'ovaire reste quant à elle similaire à celle de la gonade indifférenciée, présentant un mélange de cellules somatiques et germinales en son sein.

1.2.4.2.1. Formation des cordons ovariens.

Les cordons ovariens sont des structures épithéliales formées pas des ovocytes en amas, entourés de cellules épithéliales, précurseurs des cellules de la granulosa. Ils sont bordés d'une membrane basale, contiguë à l'épithélium de surface ovarien, lui-même en contact avec l'épithélium cœlomique, source de nouveaux précurseurs de granulosa. Les cordons sont bordés par les cellules mésenchymateuses.

1.2.4.2.1.1. Cellules de pré-granulosa

Ces analogucellules précurseurs des cellules de la granulosa sont souvent considérées comme analogues des cellules de Sertoli dans le testicule, mais elles subissent moins de bouleversements morphogénétiques que leurs équivalents masculins (Jameson et al., 2012b; Munger et al., 2013; Nef et al., 2005) (voir « Cellules de Sertoli »). En effet, elles auraient la même origine que les cellules de Sertoli, mais n'ayant pas exprimé le gène *SRY* lors de la différenciation sexuelle, auraient entamé leur différenciation vers la voie femelle. Ces observations ont été confirmées chez la souris grâce à des expérimentations de lignage cellulaire de souris transgénique exprimant la GFP (Albrecht and Eicher, 2001; Ito et al., 2006). Cependant, il a également été observé que les cellules de pré-granulosa situées dans le cortex (zone externe) de l'ovaire dériveraient de l'épithélium de surface ovarien chez la souris (Mork et al., 2012). Leur origine n'est donc pas encore tout à fait établie.

La formation des cellules de pré-granulosa est gouvernée par les gènes *FOXL2*, *WNT4*, *RSPO1*, *CTNBI* (codant pour la β -catenin), comme décrite dans le chapitre de la « Trajectoire femelle » de la différenciation gonadique. Suite à cette première vague de différenciation, l'augmentation de l'expression de gènes codant pour *CDKN1B* et la diminution induite de la cycline 1 (*CCNA1*) entraîne un arrêt mitotique des cellules de la granulosa de la médulla. De son côté, l'épithélium cœlomique continue de proliférer, fournissant un ensemble de cellules

précurseurs pour la seconde vague de différenciation des cellules de la granulosa (Mork et al., 2012). Contrairement au testicule, l'épithélium coelomique continue d'exprimer les récepteurs de RSPO1, LGR4 et LGR5, ce qui maintient sa capacité à répondre au ligand et donc à maintenir le recrutement continu des cellules précurseurs vers une différenciation en cellule de la granulosa. Les granulosa médullaires différenciées perdent peu à peu l'expression de *LGR5* (Gustin et al., 2016).

Les cellules de la granulosa guident la différenciation ovarienne grâce à leur activité paracrine. D'une part, elles secrètent la follistatine (FST) grâce à la régulation de WNT4/ β -catenin et FOXL2 (Kashimada et al., 2011; Yao et al., 2004b). La follistatine inhibe la migration des cellules endothéliales dans l'ovaire fœtale en bloquant la fixation de l'activine à leurs récepteurs (Yao et al., 2006).

À la différence des cellules de Sertoli, des signaux pro-ovariens sont détectés dans les cellules de la granulosa après la naissance, via les signaux des œstrogènes. En effet, chez la souris, les œstrogènes sont cruciaux pour la maintenance des cellules supportrices, illustrée par la transdifférenciation de ces cellules en cellules de Sertoli dans le cas de la perte des récepteurs aux œstrogènes, ER α et ER β (Couse et al., 1999).

1.2.4.2.1.2. Cellules germinales

Dans l'ovaire, les mécanismes de différenciation les plus significatifs ont lieu dans la lignée germinale, subissant une prolifération par mitose puis une étape méiotique, accompagnée par une vague d'apoptose pour réguler leur nombre. Cela constitue une différence majeure avec le développement testiculaire : alors que la lignée germinale n'est pas requise pour le développement testiculaire, elle est indispensable pour la formation et le maintien des structures folliculaires et donc le développement ovarien (Behringer et al., 1990; Couse et al., 1999; McLaren, 1984; Reynaud et al., 2004; Singh and Carr, 1966).

- *Ovogenèse – mitose*

Dans l'ovaire, les cellules germinales primordiales (CGP) ayant colonisées la gonade deviendront ovogonies et proliféreront de la 5^{ème} SD jusqu'à la 10^{ème} SD, certains jusqu'à 16^{ème} SD puis entreront en méiose à partir de la 13^{ème} SD (Cohen and Holloway, 2010; Kerr et al., 2008; Motta, 1997). En effet, l'entrée en méiose est effectuée de manière asynchrone, certaines cellules germinales exprimant des marqueurs de pluripotence jusqu'à la 16^{ème} SD, alors que d'autres entrent en méiose dès la 13^{ème} SD (Kerr et al., 2008). Ces cellules germinales exprimeront des marqueurs tel que OCT4 et LIN28 (Childs et al., 2012) et l'induction en mitose se propagera de manière radiale dans l'ovaire, débutant à la périphérie pour se diriger vers le centre médullaire de l'ovaire.

Cette phase mitotique permet aux cellules germinales d'effectuer une division cellulaire au cours de laquelle les chromosomes sont dupliqués puis divisés, pour ainsi créer des cellules filles portant l'information génétique conservée. Cette mitose s'effectue de manière très rapide et la cytokinèse imparfaite à la fin de ce processus induit la formation de nids ovocytaires (Edson et al., 2009; Lei and Spradling, 2016; Pepling, 2012). Ces nids ovocytaires sont entourés de cellules pré-granulosa et constitue donc les cordons ovariens, entourés d'une membrane basale continue.

- *Oocytes – méiose*

L'entrée en méiose des cellules germinales contenue dans les cordons ovariens s'effectue au cours de la 13^{ième} SD (Carninci et al., 2005; Motta, 1997) de manière asynchrone. La production d'acide rétinoïque dans le tissu somatique environnant, mésonéphros ou ovaire fœtal lui-même, est l'inducteur principal de l'entrée en méiose des ovogonies (Childs et al., 2011; Le Bouffant et al., 2010). L'acide rétinoïque est un dérivé actif de la vitamine A et sa liaison avec ses récepteurs RARs et RXRs induit l'expression de *STRA8*, qui promeut la transition des ovogonies vers la division méiotique. L'entrée en méiose des cellules germinales est donc contrôlée par un système homéostatique entre synthèses de l'acide rétinoïque par les rétinaldéhydes déshydrogénases (RALDHS) dans la gonade femelle et sa dégradation par les CYP26 (A1, B1, C1), retrouvées dans le testicule fœtal (voir « Quiescence des cellules germinales »). À partir de la 17^{ième} SD, les cellules germinales femelles subissent un arrêt méiotique en prophase I, à l'étape diplotène, où elles demeureront jusqu'à leur activation à partir de la puberté. Le cycle méiotique ne reprendra que lors de l'ovulation.

Contrairement à la méiose chez le mâle adulte, une ovogonie diploïde à 2 chromatides ne fournit pas 4 cellules filles haploïdes à 1 chromatide (Evans and Robinson, 2011). La première division engendre une cellule fille apte à subir la seconde division et un globule polaire, cellule résiduelle issue de la répartition hétérogène du matériel cytosolique et vouée à la dégénérescence. Le même phénomène se produit en deuxième division de méiose, après l'ovulation à l'âge adulte : un seul ovocyte mature est obtenu, haploïde et contenant suffisamment de matériel cytosolique, notamment les mitochondries et le nucléole. Un second globule polaire contenant le reste du matériel génétique est émis, voué comme le précédent à la dégénérescence.

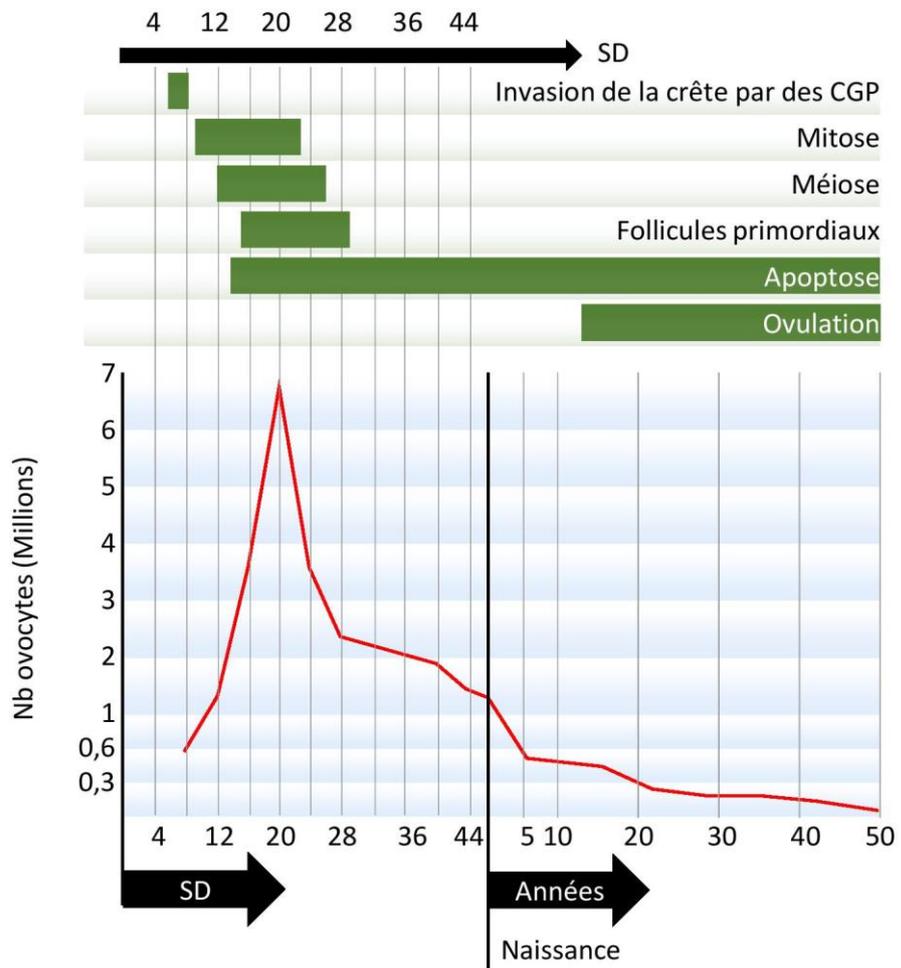


Figure 18. Nombre de cellules germinales au cours de la vie fœtale jusqu'à la vie adulte

1.2.4.2.2. Formation des follicules primordiaux

Ce processus, pourtant capital pour la fertilité de la femme, est étonnamment mal caractérisé. Cependant, certaines études ont su identifier certains gènes impliqués (Fowler et al., 2009), mais les mécanismes moléculaires qu'ils gouvernent restent à définir.

À la différence de certaines espèces mammifères comme le rat, le processus d'histogénèse folliculaire a lieu intégralement pendant le développement fœtal chez l'humain, à partir de la 15^{ème} SD (Forabosco and Sforza, 2007). Cette étape clé conditionnera la santé reproductive de la femme, car le stock initial de follicules primordiaux constitué pendant la vie fœtale n'est pas renouvelé après la naissance et constituera donc le seul stock de gamètes féminins. Cependant, ce dogme est remis en question par de nouveaux résultats obtenus dans les ovaires adultes de mammifères (Porras-Gómez and Moreno-Mendoza, 2017).

Les cellules de pré-granulosa entourant les cordons ovariens, contenant plusieurs ovocytes, s'insinuent au sein de la structure pour individualiser certaines cellules germinales et ainsi former des structures contenant un seul ovocyte entouré de cellules épithéliales puis d'une membrane basale, appelée dès lors follicules. La majorité des ovocytes qui ne seront pas inclus dans un follicule seront dégradés par apoptose. En effet, parmi les 7 millions d'ovocytes présents dans l'ovaire à mi-gestation, seulement 20 % persisteront à la naissance (Figure 18). Ce processus de rupture des nids ovariens commence dans la partie médullaire et s'étend vers la surface corticale de l'ovaire (De Felici et al., 2005; Mazaud et al., 2005; Pepling et al., 2010). Comme décrit précédemment, les cellules de pré-granulosa subissent deux vagues de différenciation (voir « Cellules de pré-granulosa »). Chez la souris, cette asynchronie implique également deux vagues de formation de follicules primordiaux : la plus précoce, ayant lieu dans la région médullaire de l'ovaire et dont les follicules sont destinés à mourir et la seconde, ayant lieu dans le cortex ovarien (Zhang et al., 2014). Ces observations ont été confirmées chez l'homme (De Pol et al., 1997).

Le dialogue entre les cellules somatiques et germinales permet la régulation paracrine de la formation folliculaire. Il a été démontré que la voie de signalisation Kit joue un rôle primordial dans la rupture des syncytia ovocytaires et donc dans la formation de follicules primordiaux (Findlay et al., 2002; Jones and Pepling, 2013). L'expression de KITL dans les cellules somatiques pourrait être régulé par l'activine A, un membre de la famille des TGF- β impliqué dans la folliculogénèse (Childs and Anderson, 2009; Knight and Glister, 2003, 2006; Shimasaki et al., 2004). Chez l'homme, les ovocytes expriment l'activine A avant la rupture des syncytia ce qui inhibe l'expression de KITL dans les cellules somatiques avoisinantes, empêchant ainsi l'induction de l'expression de c-kit dans les cellules germinales elle-même (Tingen et al., 2009). Après la rupture, l'expression décroissante de

l'activin A entraîne l'activation du système KIT et promeut l'assemblage des follicules primordiaux (Deng et al., 2003; Jones and Pepling, 2013).

Chez la souris, l'inactivation de la voie de signalisation NOTCH aboutit à une perturbation de la formation des follicules primordiaux (Trombly et al., 2009). De plus, l'inactivation de Notch2 dans les cellules de la granulosa entraîne l'altération de la fragmentation des nids ovocytaires (Xu and Gridley, 2013). Les Ngf (Nerve Growth Factor) sont également impliqués dans l'assemblage des follicules primordiaux, par un mécanisme encore mal connu (Kerr et al., 2009). Les autres neurotrophines Bdnf (brain derived neurotropic factor), Nt3 et Nt 4-5 sont également impliquées. Elles sont majoritairement exprimées par les cellules somatiques, mais leurs récepteurs sont portés par les cellules germinales (pour revue, Dissen et al., 2009).

Chez les modèles murins, de multiples facteurs de transcription, tels que *Hnrnpk*, sont régulateurs de nombreuses protéines impliquées notamment dans le cycle cellulaire et leur absence réduit la formation des follicules primordiaux (Wang et al., 2011). *Fxna*, dont la délétion conduit également à la persistance de cordons ovariens non fragmentés sans dérégulation de l'apoptose ou défaut de prolifération des cellules de la granulosa, est nécessaire pour coordonner l'organisation des cellules en follicules primordiaux (Garcia-Rudaz et al., 2007). *Figla* est connu pour réguler l'expression des gènes *Zp* à l'origine de la mise en place de la zone pellucide (portion de la matrice extracellulaire apparaissant au stade primaire du développement folliculaire) et sa délétion empêche la formation des follicules primordiaux (Soyal et al., 2000). *Nobox* est également impliqué dans la formation folliculaire. En son absence, les follicules sont beaucoup plus lents à s'assembler. Toutefois, *Nobox* est essentiel pour la suite du développement folliculaire (Rajkovic et al., 2004).

1.2.4.2.3. Régulation endocrine : un rôle à définir.

La production et le rôle des œstrogènes dans l'ovaire fœtal humain n'est pas éclairci aujourd'hui, bien qu'il ait été démontré que l'ovaire était capable de convertir l'androgène en œstrogène à partir de la 10^{ième} SD (George and Wilson, 1978). Plusieurs acteurs de la voie stéroïdogénique ont été observés dans l'ovaire fœtal (Cole et al., 2006; Fowler et al., 2009; Gould et al., 2000; Juengel et al., 2006; Vaskivuo et al., 2005; Wilson and McPhaul, 1996). Globalement, l'ovaire fœtal exprime toute la machinerie nécessaire à la production et à la réponse aux signaux œstrogéniques et androgéniques, y compris *CYP11A1*, *CYP17A1* et *HSD3B2*, étonnamment exprimés dans les cellules somatiques et les oocytes lors de la formation des follicules primordiaux (Fowler et al., 2011).

De même, les gonadotrophines (FSH, LH) et l'hCG qui interviennent dans le développement du testicule (voir « Régulation endocrine : l'axe hypothalamo-hypophysaire gonadique ») n'ont pas de rôle identifié dans le développement fœtal de l'ovaire.

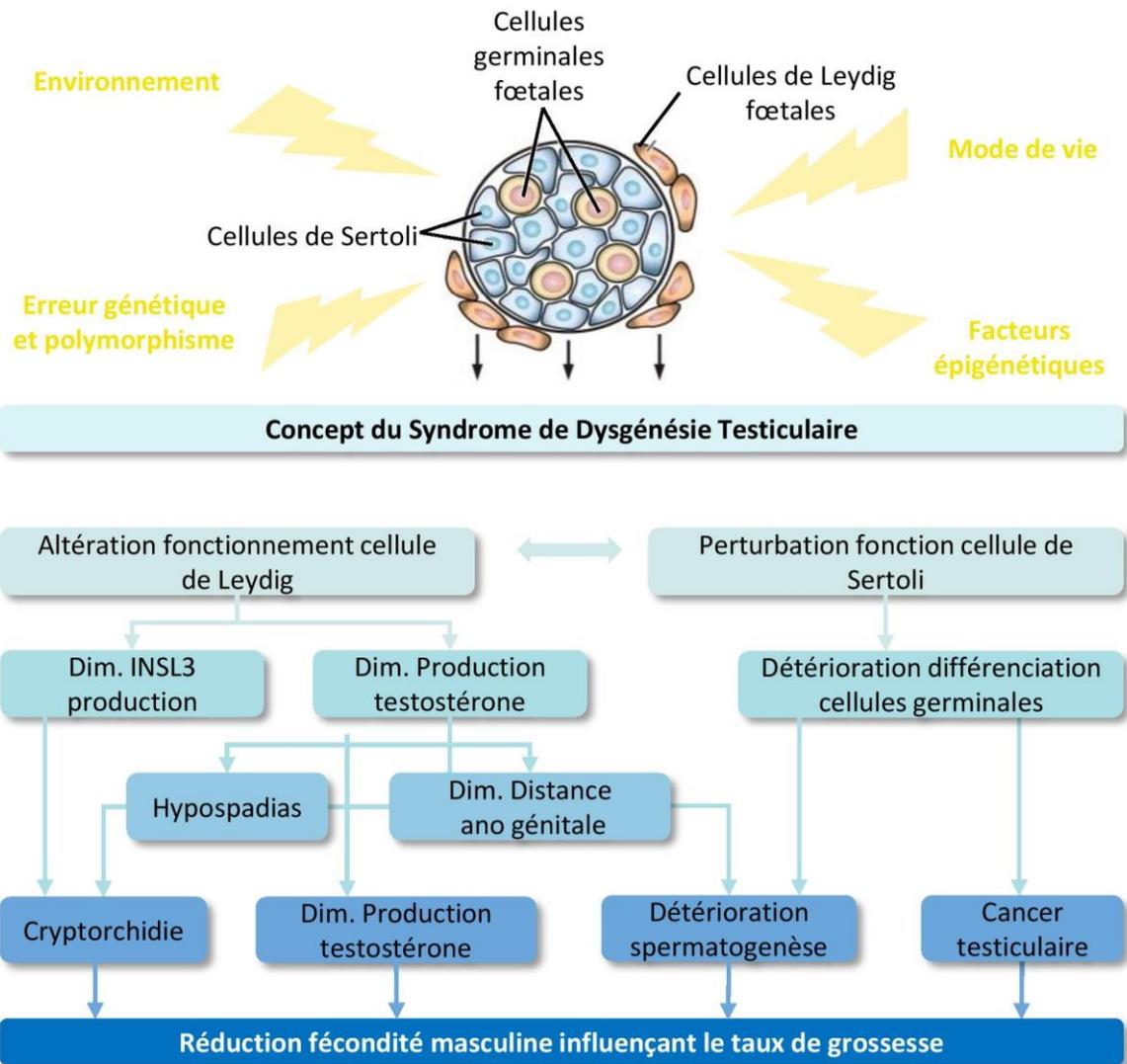


Figure 19. Schéma du syndrome de dysgénésie testiculaire selon Skakkebaek *et al.* 2016

Le chercheur danois Niels Erik Skakkebaek et ses collaborateurs ont formulé l'hypothèse que les différents phénomènes observés sur la reproduction chez l'homme pourraient être réunis et expliqués par une même entité : le Syndrome de Dysgénésie Testiculaire (*Testicular Dysgenesis Syndrome*, TDS).

1.2.5. Les troubles du développement sexuel

Les anomalies du développement sexuel (ADS) sont définies comme des « anomalies de différenciation sexuelle qui peuvent apparaître dans les gonades, l'appareil génital et les organes sexuels externes », incluant des inversions du sexe complètes jusqu'aux défauts mineurs du tractus urogénital. L'incidence globale des ADS est estimée à 1/4500-5500 naissances (Lee et al., 2016), certaines pathologies étant plus fréquentes que les autres, comme l'aneuploïdie des chromosomes sexuels, l'hyperplasie congénitale des glandes surrénales ou le syndrome d'insensibilité aux androgènes. Cependant, la plupart des ADS sont des pathologies idiopathiques et de ce fait, difficiles à diagnostiquer. De plus, les mutations sur des gènes clés du développement sexuels peuvent induire des phénotypes variés (Brauner et al., 2016). On estime aujourd'hui que seulement 13 % des patients atteints de ADS reçoivent un diagnostic clinique et génétique définitif, mais l'utilisation de technologies de séquençage à haut-débit tend à améliorer l'efficacité des diagnostics (Eggers et al., 2016) et donc la prise en charge des patients.

1.2.5.1. Concept de TDS

L'augmentation de l'incidence de pathologies du tractus génital mâle observée depuis quelques années a amené une équipe danoise, menée par Niels Skakkebaek, a énoncé en 2001 le concept de syndrome de dysgénésie testiculaire (Skakkebaek et al., 2001). Ce syndrome regroupe la cryptorchidie, l'hypospadias, le cancer testiculaire ainsi qu'une qualité spermatique réduite et émet l'hypothèse que l'ensemble de ces troubles pourrait trouver une origine commune, pendant la vie fœtale. Les facteurs impliqués dans ces désordres incluent les susceptibilités génétiques, c'est-à-dire des mutations chez certains gènes clés du développement (décrites dans « La différenciation sexuelle » et « Développement testiculaire »), mais également les expositions environnementales, plus particulièrement aux perturbateurs endocriniens (Figure 19). En 2016, le même groupe danois a mis à jour leurs données et inclus dorénavant des facteurs maternels, tels que le stress ou la nutrition, dans les causes potentielles (Skakkebaek et al., 2016). Cependant, comme énoncé précédemment, certaines études épidémiologiques sont régulièrement remises en doute et le concept de TDS est encore débattu à ce jour.

De manière analogue au concept de TDS, le syndrome de dysgénésie ovarienne a été décrit pour la première fois en 2011 (Buck Louis et al., 2011). Elle émet l'hypothèse que certains désordres spécifiques de l'appareil génital féminin trouvent également leur origine pendant la vie fœtale, toujours dus à des susceptibilités génétiques ou exposition à des composés environnementaux. Ils incluent une fertilité réduite, le syndrome d'ovaires polykystiques, une insuffisance ovarienne prématurée et les cancers reproductifs. Ce concept ne sera pas développé d'avantage car les données

sont limitées et l'analyse toxicogénomique effectuée dans cette thèse ne concerne que le testicule fœtal humain.

1.2.5.2. Les anomalies du tractus urogénital mâle et désordres de la fertilité

1.2.5.2.1. Cryptorchidie

Pendant la vie fœtale, les testicules se développent dans l'abdomen, puis vont migrer vers le scrotum avant la naissance grâce à la descente testiculaire. Cette descente est divisée en deux phases : la descente abdomino-inguinale, régulée par l'INSL3, et la descente inguino-scrotale, régulée par la testostérone. Ces deux hormones sont produites par les cellules de Leydig fœtales (Virtanen and Toppari, 2014).

La cryptorchidie désigne la non-descente d'un ou des deux testicules dans le scrotum avant 3 mois postnatal et affecte 1 à 8 % des nouveaux nés garçons (Gaspari et al., 2011; Paulozzi, 1999; Virtanen and Toppari, 2008). Dans la plupart des cas recensés, la descente testiculaire a lieu spontanément après la naissance et seulement 1 à 2 % des garçons sont toujours atteints à 1 an. Étant donné les difficultés de diagnostics et les disparités géographiques de prévalence, il est difficile d'établir l'évolution de l'occurrence de cette pathologie (Serrano et al., 2013).

1.2.5.2.2. Hypospadias

L'hypospadias est une anomalie congénitale de la fermeture de l'urètre, affection la plus fréquente du pénis avec une incidence d'environ 3 %. Il est caractérisé par une localisation anormale de son ouverture sur la surface ventrale de la verge, ou au niveau du scrotum ou encore du périnée. Récemment, une étude longitudinale, danoise a montré une augmentation de la prévalence de l'hypospadias de 2.40 % par an, indépendamment de l'âge maternel (Lund et al., 2009). Cependant, ces observations ont été remises en question par d'autres études menées aux USA, au Royaume-Uni (Fisch et al., 2010) et en Europe (Bergman et al., 2015).

1.2.5.2.3. Les cancers du testicule

Bien que cancer du testicule soit le plus fréquent chez l'homme jeune (15-35ans), il ne représente seulement 1 à 2 % des cancers à l'échelle de la population masculine. Cependant, ces chiffres peuvent varier en fonction du pays étudié, et les pays industrialisés voient s'accroître le nombre de cas depuis le milieu du 20^{ème} siècle (Chia et al., 2010; Huyghe et al., 2007; Purdue et al., 2005). Les cancers des cellules germinales, y compris du séminome, représentent la grande majorité des cas observés. La cryptorchidie constitue un facteur de risque du cancer du testicule (Cook et al., 2010), ainsi que des antécédents familiaux de cryptorchidie, de cancers testiculaires, de la prostate ou du sein (Serrano et al., 2013; Walschaerts et al., 2007), un petit poids de naissance, la survenue

d'une hernie inguinale, un accouchement prématuré ou encore une grossesse gémellaire (Cook et al., 2010). Cependant, ces facteurs ne permettent pas d'expliquer l'augmentation dramatique de son incidence ces dernières décennies, le nombre de cas ayant doublé (Shanmugalingam et al., 2013). Désormais, la communauté scientifique considère la pollution environnementale et l'exposition des populations comme une des causes possibles des cancers testiculaires.

1.2.5.2.4. La qualité spermatique

La santé reproductive de l'homme peut également être évaluée grâce à la qualité et à la quantité des gamètes qu'il produit. Plusieurs études à travers le monde ont permis d'établir à 100 millions de spermatozoïdes par millilitre la norme de la quantité spermatique d'un homme sain (David et al., 1979; Macleod and Gold, 1951).

Dans les années 90, une méta-analyse de 61 articles effectuée par le groupe de Niels Erik Skakkebaek a mis en évidence une diminution régulière de la production et de la densité spermatique entre 1938 et 1990. Bien que critiquée, cette étude a permis une prise de conscience internationale sur ce problème de santé publique, et a donné l'impulsion pour l'analyse rétrospective des données accumulées jusqu'alors. Les études effectuées sur différents continents, prenant en compte les données recueillies dans différents pays ou ville, ont su démontrer une grande disparité géographique de la production spermatique, observée en Amérique du Nord (Swan et al., 2003) et en Europe (Jørgensen et al., 2001). La même hétérogénéité a été observée à l'échelle de la France, en considérant plusieurs paramètres spermatiques (volume de l'éjaculat, concentration et mobilité spermatiques) (Auger and Jouannet, 1997). Depuis une trentaine d'années, de nombreuses études ont tenté de démontrer les variations de différents paramètres spermatiques, mais aucun consensus n'a pu être établi à ce jour. La revue de Niels Erik Skakkebaek publiée en 2016 répertorie les études publiées au fil des années et fait un état des lieux des potentiels facteurs confondants pouvant amener à ces résultats discordants (Skakkebaek et al., 2016). Il est toujours difficile aujourd'hui de définir l'évolution de la qualité spermatique des hommes.

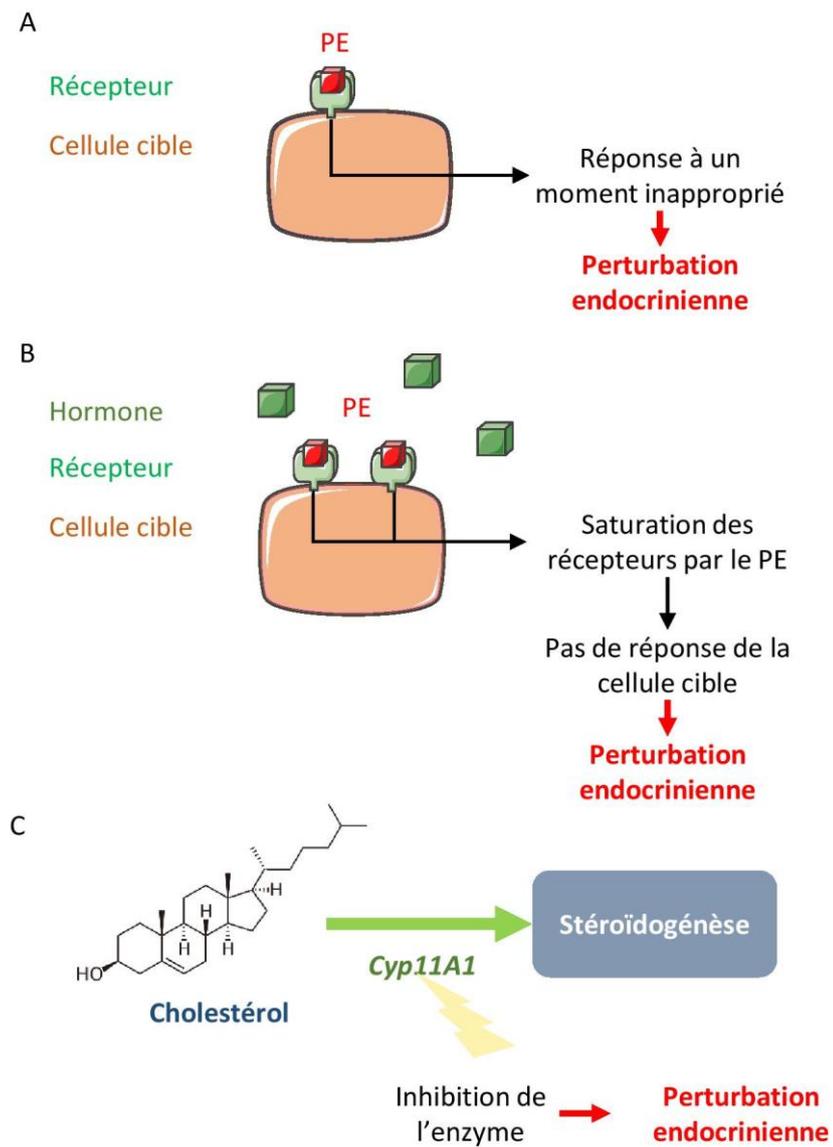


Figure 20. Différents modes d'action des Perturbateurs endocriens

(A) Action hormono-mimétique du perturbateur endocrinien (PE). (B) Action antagoniste du PE. (C) Effet direct sur l'hormone via la perturbation d'enzyme de conversion (ici sur l'enzyme de clivage de la chaîne latérale du cholestérol, inhibant la testostérone).

1.2.5.3. Perturbation endocrinienne et les gonades fœtales

Bien que des études aient déjà décrit des cas de perturbations endocriniennes dans les années 40, sans en énoncer clairement le concept (Wurster, 2015), ce n'est qu'en 1962 de Rachel Carson alerte l'opinion sur l'impact de certaines substances sur la santé animale, y compris l'Homme, dans son livre « Silent Spring » (Carson, 1962). Les années 80 et 90 verront la publication de nombreuses études sur l'impact de produits chimiques sur la reproduction animale à travers le monde. En Angleterre, les poissons mâles exposés aux rejets des stations de traitement des eaux usées présentent une inversion du sexe (Sumpter and Jobling, 1995) et aux États-Unis, les alligators présentent une diminution de testostérone, avec des conséquences phénotypiques majeures, après leur exposition accidentelle au diclofol (Guillette et al., 1994).

Le concept de perturbation endocrinienne voit le jour grâce au travail collaboratif et interdisciplinaire d'une vingtaine de scientifiques, réunis par Theo Colborn et Pete Myers en 1991 à Wingspread. La déclaration de Wingspread fut rédigée à la suite de cette rencontre et énonce déjà les grandes lignes de ce nouveau concept toxicologique, mais ce n'est que 5 ans plus tard que Theo Colborn et Pete Myers publient *Our Stolen Future*, qui explicite le principe de perturbation endocrinienne et évoque sans détour la possibilité que l'humanité mette en danger « sa fertilité, son intelligence et sa survie » par sa consommation excessive de produits chimiques (Colborn et al., 1996).

Au fur et à mesure que la communauté scientifique s'est emparée de ce nouveau champ de recherche, les connaissances sur les perturbateurs endocriniens se sont étoffées, mais ont révélé une complexité mécanistique telle qu'il est difficile aujourd'hui de définir clairement les contours de la perturbation endocrinienne. Le 15 juin 2016, l'UE a présenté sa dernière définition : « **un perturbateur endocrinien est une substance ou un mélange exogène qui a un mode d'action endocrinien et qui est connu pour exercer des effets adverses pour la santé humaine sur un organisme intact, sa descendance ou ses sous populations comme conséquence de son effet mode d'action hormonale.** ». Plusieurs conséquences sous-jacentes ont déjà été soulevées par la communauté scientifique. Cette définition n'inclut que la population humaine, excluant de fait le reste du règne animal. De plus, elle n'inclut que les produits « connus » pour avoir des effets délétères, ce qui exclut des nomenclatures la notion de présomption et de suspicion. La définition de l'OMS ne cristallise pas les mêmes critiques, car elle décrit un PE comme « **une substance ou un mélange exogène, possédant des propriétés susceptibles d'induire une perturbation endocrinienne dans un organisme intact, chez ses descendants ou au sein de (sous)-populations.** ».

En France, l'Agence Nationale de Sécurité Sanitaire de l'alimentation, de l'environnement et du travail (ANSES) se range à l'avis de l'OMS, en définissant un PE comme « **tout produit chimique susceptible d'interagir directement avec le système endocrinien et par voie de conséquence de produire un effet sur ce dernier et d'impacter les organes et les tissus** », n'incluant donc pas de notion d'espèces impactées et en conservant la présomption d'effets. Il est évident que ces discordances des différents organismes en charge de la législation sur la production de produits chimiques compliquent l'homogénéisation des règles de production et de mises sur le marché à travers le monde.

Le système endocrinien gouverne l'activité des organes et cellules cibles grâce à l'excrétion d'hormones dans la circulation sanguine par les organes endocrines. Le mode de production des hormones, leur transport dans la circulation, leur fixation sur les récepteurs de leur cible et les voies de signalisation induites sont autant de mécanismes susceptibles d'être altérés par les perturbateurs endocriniens. On distingue donc plusieurs types de PEs. Tout d'abord, une substance peut inhiber l'activité de certaines enzymes impliquées dans les voies de synthèse des hormones, tel le kétoconazole altérant la synthèse de testostérone via l'inhibition des cytochromes P450 (Miossec et al., 1997). Ensuite, le transport de l'hormone vers sa cible peut être altéré, comme les phtalates perturbent la circulation des hormones lipophiles comme la Sex Hormone-Binding Globulin (SHBG) (Sheikh et al., 2016). Mais un PE peut également contrarier l'activité des organes cibles en interagissant directement avec eux et sans perturber l'hormone en elle-même. D'une part, certaines substances ont la capacité de se fixer sur la cible grâce à leur similarité de structure avec l'hormone physiologique. Ces agonistes peuvent donc mimer l'action d'une hormone et induire une réponse non physiologique des cibles, comme le chlordécone peut induire une activité œstrogénique (Waring and Harris, 2005). D'autre part, certaines substances ont un effet inverse, se fixant sur les récepteurs et bloquant ainsi l'accès de l'hormone native à ses cibles. Cela induira un effet antagoniste, où le signal hormonal ne pourra être capté par l'organe a stimulé.

L'une des crises sanitaires les plus importantes provoquée par un perturbateur endocrinien est celle provoquée par le diéthylstilbestrol (DES) dans le monde entier. Dans les années 1970, des millions de femmes enceintes ont été traitées au DES afin d'éviter les avortements spontanés. Les effets du DES ne sont pas visibles sur la femme enceinte, mais des études épidémiologiques ont su démontrer l'augmentation de l'occurrence des malformations de l'appareil génital, des cancers du sein et de la stérilité chez les enfants nés des mères exposées (Eskenazi et al., 2009; Klip et al., 2002). Ainsi, le DES s'est avéré être l'archétype du PE présentant des effets transgénérationnels, aboutissant à son retrait du marché français en 1977.

Comme exposé dans la seconde partie de l'introduction (voir « La différenciation gonadique »), le développement des gonades est gouverné par des mécanismes de régulation moléculaires, permettant la coordination des signaux entre les différentes populations cellulaires en développement et l'activité endocrine de celles-ci. Le rôle des stéroïdes produits par le testicule est bien défini dans la masculinisation du tractus génital. Une altération du développement des cellules de Leydig aurait des effets délétères à long terme sur la santé reproductive de l'homme. L'extrapolation des résultats obtenus chez le rat a permis de définir une fenêtre de susceptibilité du testicule humain aux PE entre la 8^{ième} et la 12^{ième} SD (Welsh et al., 2008). Certaines études ont en effet démontré que la production hormonale du testicule fœtal humain pouvait être altérée après exposition à certains composés chimiques, classés comme anti-androgénique (Toppari et al., 1996). Ces molécules ont la capacité de perturber le système androgénique, y compris la production de testostérone (pour revue Albert and Jégou, 2014). Au cours de la dernière décennie, les équipes de l'Irset ont démontré l'effet direct de certains composés sur la stéroïdogénèse des testicules fœtaux humains, comme le bisphénol A (Ben Maamar et al., 2015), l'indométacine, le paracétamol (van den Driesche et al., 2015; Jégou, 2015), l'aspirine (Mazaud-Guittot et al., 2013) et l'ibuprofène (Ben Maamar et al., 2017). Il a également été démontré que certains produits induisent des effets différents sur les testicules fœtaux murins et humains, notamment les phtalates (Johnson et al., 2012). Cependant, une revue de la littérature a mis en évidence que les méthodes d'exposition ou d'évaluation de l'effet pouvaient grandement influencer les résultats des études toxicologiques et aboutir à des conclusions divergentes (Albert and Jégou, 2014). Le cas des phtalates illustre bien qu'il est difficile d'établir des normes en ce qui concerne l'évaluation de la perturbation endocrinienne, et qu'il serait nécessaire d'évaluer l'effet de chaque PE suspecté sur tissu humain au vu des spécificités interspèces potentielles.

Chez la femme en revanche, étant donné que les mécanismes endocriniens régissant les différentes étapes du développement ovarien sont encore mal connus, il est difficile d'appréhender si un effet toxique observé est dû à un mécanisme de perturbation endocrinienne ou non. L'établissement de la réserve de follicules primordiaux constituant une étape clé pour la future santé reproductive de la femme, une altération de leur formation pourrait avoir des conséquences irréversibles sur la fertilité. Des études toxicologiques organotypiques ont su démontrer l'effet délétère de certains composés sur le développement de l'ovaire humain, comme l'ibuprofène (Leverrier-Penna et al., 2018), le bisphénol A (Briño-Enríquez et al., 2012; Zhang et al., 2012) et le diéthylstilbestrol (Akbas et al., 2004; Bromer et al., 2009).

1.3. Analyses transcriptomiques du développement gonadique

Comme évoqué au début de l'introduction, le transcriptome peut être défini à l'échelle d'un tissu, d'une population cellulaire isolée ou d'une cellule unique (voir « La transcriptomique »).

Les études travaillant à l'échelle de la gonade entière ont l'avantage de conserver l'intégrité du tissu et des interactions cellulaires. Ces études nécessitent moins d'étapes de préparation des échantillons à séquencer, le transcriptome sera peu affecté par les différentes procédures expérimentales. Cependant, séquencer l'organe entier empêche de distinguer le transcriptome de chaque population cellulaire présente et les populations dominantes peuvent occulter les cellules présentes en plus faible proportion. Les études transcriptomiques effectuées sur des populations cellulaires enrichies peuvent surpasser ce problème et améliorer la sensibilité du séquençage. Cependant, les protocoles de purification peuvent avoir un impact sur l'expression des gènes. De plus, les protocoles de purification peuvent donner lieu à des contaminations de l'échantillon par d'autres populations cellulaires. L'analyse à l'échelle de la cellule unique permet de décrire la dynamique du transcriptome au sein d'une population cellulaire hétérogène et peut être utilisée pour caractériser chaque type cellulaire d'un échantillon, y compris les plus minoritaires et appréhender le programme transcriptionnel gouvernant la différenciation des cellules (voir « **Erreur ! Source du renvoi introuvable.** »).

1.3.1. Premières études du développement gonadiques grâce aux puces à ADN

Depuis une quinzaine d'années, les différentes technologies de transcriptomique ont permis de caractériser la dynamique transcriptionnelle intervenant dans la différenciation des gonades fœtales. Les puces à ADN ont été utilisées par de nombreuses études à l'échelle de la gonade totale, à différents temps de développement et dans de nombreuses espèces (Combes et al., 2011; Coveney et al., 2008; Garcia-Ortiz et al., 2009; Houmard et al., 2009; Liu et al., 2015; Mamsen et al., 2017; Munger et al., 2013; Nicol and Yao, 2015; Ottolenghi et al., 2007; Rolland et al., 2011; Small et al., 2005; del Valle et al., 2017). Les différentes populations cellulaires du testicule et de l'ovaire fœtal ont également été largement décrites, tant la lignée germinale (Jameson et al., 2012b; Molyneaux et al., 2003) que les populations cellulaires somatiques (Beverdam and Koopman, 2006; Bouma et al., 2007; Hummitzsch et al., 2013; Jameson et al., 2012b; Kusaka et al., 2010; Nef et al., 2005). Ces études ont permis de décrypter les bases de la mécanistique moléculaire régulant différents aspects de la différenciation gonadique.

Le RNA-seq a permis des avancées significatives dans notre compréhension des différents processus de différenciation ayant lieu dans la gonade indifférenciée, le testicule ou l'ovaire fœtal.

1.3.2. Etude des cellules germinales primordiales

Au cours de ces dernières années, de nombreuses études ont tenté d'induire la différenciation *in vitro* des CGP à partir de cellules souches embryonnaires humaines (hESCs). Le développement d'une lignée cellulaire « human primordial germ cells-like », hCGPLCs, permet d'une part d'étudier les caractéristiques des cellules germinales indifférenciées humaines, mais également leur processus de différenciation. Pour tester les protocoles mis en place, il a été nécessaire de définir le transcriptome des cellules germinales humaines fœtales, grâce au séquençage des populations purifiées (Gkoutela et al., 2013, 2015; Guo et al., 2015a, 2017; Irie et al., 2015; Li et al., 2017a; von Meyenn and Reik, 2015; Tang et al., 2015). Ainsi, la comparaison des transcriptomes des cellules germinales différenciées *in vitro* et des cellules germinales « natives » a permis d'évaluer la pertinence biologique des protocoles testés. Tout d'abord, la comparaison des transcriptomes des CGP isolées chez l'homme et la souris a pu mettre en évidence la conservation d'un ensemble de gènes impliqués dans la spécification des CGP murine, revue dans (Irie et al., 2018). L'analyse du programme transcriptionnel régulant la différenciation *in vitro* des hCGPLCs a mis en évidence le rôle clé de SOX17 (Irie et al., 2015), spécifique de l'humain (pour revue, Irie et al., 2018; Tang et al., 2016). Grâce à ces études, le modèle d'induction de la différenciation des CGP humains a pu être en partie décrit. De plus, cette nouvelle lignée cellulaire ayant toutes les caractéristiques des CGP représente un bon modèle pour étudier la reprogrammation épigénétique qu'elle subit (revue dans Tang et al., 2016). Enfin, les CGP étant les cellules progénitrices des futurs spermatozoïdes et ovocytes, l'opportunité d'étudier leur différenciation *in vitro* facilitera à l'avenir notre compréhension de la fertilité humaine et des pathologies associées.

L'utilisation récente de technologies de séquençage à haut-débit (RNA-seq) a permis d'affiner la dynamique du transcriptome permettant la différenciation des cellules germinales humaines pendant la vie fœtale (Gkoutela et al., 2015). Notamment, cela a permis de mettre en évidence l'expression significative du gène non codant *XIST* (Gkoutela et al., 2015) dans les cellules germinales femelles, mais également mâle pendant la vie fœtale. *XIST* est connu comme l'un des principaux effecteurs de l'inactivation du chromosome X chez la femelle mammifère (Ohhata et al., 2011). En « tapissant » le chromosome X à partir duquel il est synthétisé, l'ARN non codant rend inaccessible le chromosome X à toute machinerie transcriptionnelle, le rendant inactif. Ce processus est nécessaire pour rétablir la balance dans les cellules femelles (XX) de la quantité de gènes transcrits à partir de ce chromosome, présent en une seule copie chez le mâle. La fonction de ce facteur dans les cellules germinales fœtales mâles n'est pas encore définie.

1.3.3. Atout du RNA-seq pour l'étude de la différenciation sexuelle chez les espèces non-modèles

La différenciation sexuelle a également été étudiée dans de nombreuses espèces. Les mécanismes de la différenciation sexuelle sont diversifiés dans le règne animal (Bachtrog et al., 2014). Les poulets, comme beaucoup d'espèces aviaires, portent les chromosomes sexuels Z et W, les mâles portant deux chromosomes ZZ et la femelle ZW. Le séquençage haut-débit a permis dans un premier temps de compléter l'annotation des gènes portés par W (Ayers et al., 2013), puis d'étudier la dynamique d'expression des gènes lors de la différenciation sexuelle (Ayers et al., 2015). La différenciation température-dépendante des gonades de tortues a également été décryptée, le RNA-seq offrant la possibilité d'identifier *de novo* les transcrits impliqués dans ce processus chez cette espèce non modèle (Czerwinski et al., 2016).

1.3.4. Etude du développement gonadique chez les espèces modèles : complémentarité du RNA-seq et scRNA-seq

Chez la souris, le programme transcriptionnel en jeu pendant le développement des gonades a été largement étudié par puces à ADN, le RNA-seq n'a été employé seulement dans quelques études. L'étude complète de la dynamique du transcriptome au cours de la différenciation du testicule et de l'ovaire n'a que très récemment été publiée (Zhao et al., 2018) et tire parti du séquençage pour s'intéresser à la dynamique de l'épissage alternative lors du processus.

D'autres études ont choisi de travailler sur des populations cellulaires isolées à partir de testicules fœtaux de souris. D'une part, le RNA-seq a été utilisé pour tenter de caractériser les progéniteurs des cellules de Leydig (Inoue et al., 2016). D'autre part, (McClelland et al., 2015) a choisi de comparer les cellules de Sertoli, de Leydig et les cellules interstitielles de souris transgéniques Sfl-eGFP (Nr5a1) à différents stades de développement pour identifier les gènes spécifiquement impliqués dans la différenciation de chacune de ces populations cellulaires. L'apparition du scRNA-seq a permis d'élargir le champ des possibles quant à l'étude de la différenciation de chaque lignée cellulaire lors du développement gonadique. L'étude de (Stévant et al., 2018) a été précurseure dans le domaine, en étudiant la dynamique du transcriptome à l'échelle de la cellule unique des cellules somatiques se différenciant dans le testicule de souris. L'existence d'un précurseur commun aux cellules supportrices et interstitielles a été observé pour la première fois par (Karl and Capel, 1998), et cette approche innovante de transcriptomique a permis d'observer à l'échelle transcriptionnelle cette dynamique de différenciation à partir d'une même cellule multipotente.

L'ovaire a été relativement peu étudié au regard du testicule. Seulement une étude a étudié spécifiquement le développement ovarien murin et a plus particulièrement caractérisé le programme transcriptionnel régulant la prophase lors de la méiose des cellules germinales (Soh et al., 2015).

Chez l'humain, une étude précurseure a caractérisé les différents transcriptomes des cellules germinales du testicule et de l'ovaire fœtal, ainsi que les cellules somatiques environnantes, à différents temps de développement (Li et al., 2017a). Cette analyse a su distinguer trois populations de cellules germinales parmi les cellules testiculaires fœtales : les cellules en migration (CGP), les cellules germinales en phase mitotique et les cellules germinales en arrêt mitotique. Dans l'ovaire fœtal, quatre types cellulaires germinales ont pu être identifiés : les cellules en phase mitotique, les cellules répondantes à l'acide rétinoïque (RA), les cellules en prophase (méiose) et les cellules en cours d'ovogenèse. En ce qui concerne les cellules somatiques, quatre types cellulaires ont pu être identifiés : les cellules endothéliales, les cellules de Sertoli, les cellules de Leydig et leur précurseur. Dans l'ovaire, les cellules endothéliales, les cellules de la granulosa précoce, les cellules de la granulosa « murale » et les cellules de la granulosa tardives ont également pu être discriminées sur la base de leurs transcriptomes. Cette étude n'a séquencé que 2000 cellules, limitant ainsi l'analyse, les interprétations et les découvertes. Cependant, ces résultats sont prometteurs quant à la pertinence d'une telle technologie dans l'étude du développement gonadique et le scRNA-seq promet d'ouvrir de nouveaux horizons dans notre compréhension du processus.

2. Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

2.1. Introduction

À partir d'une gonade bipotentielle, le développement des gonades fœtales humaines est une étape cruciale pour la future santé reproductive de l'Homme. Les connaissances actuelles sur les mécanismes transcriptionnels gouvernant cette différenciation chez l'humain ne sont encore que parcellaires. Les études des gonades fœtales humaines sont peu nombreuses, tant les verrous éthiques, législatifs, pratiques et techniques rendent le matériel biologique rare et délicat à manipuler, voire à cultiver. Le modèle murin est légitimement utilisé pour l'étude du développement gonadique chez les mammifères. Mais les différences interspèces doivent conduire à une prudence critique vis-à-vis des extrapolations des résultats vers l'humain (Maheshwari and Fowler, 2008; von Meyenn et al., 2016). Incontestablement, l'avènement des technologies de transcriptomique haut-débit représente une nouvelle étape dans l'étude de la différenciation des gonades, et certaines études ont déjà employé le RNA-seq afin d'étudier les cellules germinales primordiales et leur développement (Gkoutela et al., 2013, 2015; Guo et al., 2015b; Irie et al., 2015). Ce n'est qu'en 2017 que deux études utilisant les puces à ADN ont ouvert la voie vers la caractérisation de la dynamique d'expression des gènes pendant le développement des gonades humaines (Mamsen et al., 2017; del Valle et al., 2017). A l'exception des *Tilling arrays* (Mockler et al., 2005; Samanta et al., 2007; Yazaki et al., 2007), cette technologie ne permet de quantifier qu'un ensemble prédéfini de gènes, et ne permet pas la découverte de nouveaux gènes ou de nouveaux transcrits.

L'un des objectifs de ma thèse est de décrire l'ensemble du programme transcriptionnel guidant la différenciation et le développement des gonades fœtales humaines. Cette étude est inédite tant au niveau de la technologie utilisée qu'au niveau de large fenêtre de développement étudiée. La description de la dynamique d'expression des gènes entre la 6^{ième} semaine de développement (SD), semaine de la gonade bipotentielle, et la 17^{ième} SD permet d'appréhender l'implication des gènes à plusieurs étapes de différenciation du testicule et de l'ovaire. L'un des objectifs est également de caractériser de nouveaux gènes et de nouvelles isoformes de gènes codants et non-codants, et de tenter d'éclaircir les limites de l'extrapolation entre l'homme et le modèle murin.

2.2. Manuscrit en préparation

Dynamics of the transcriptional landscape during human fetal gonad differentiation and development

Running title:

5 **Summary sentence:**

Keywords: Human gonad development; sex differentiation; transcriptional profiling; novel unannotated transcripts; lncRNAs; expression conservation

Authors and affiliations:

Estelle Lecluze¹, Antoine D. Rolland¹, Millissia. Ben Maamar^{1,2}, Bertrand Evrard¹, Isabelle Coiffec¹,
10 Panagiotis Filis³, Paul A. Fowler³, Séverine Mazaud-Guittot¹, Bernard Jégou^{1,*}, Frédéric Chalmel^{1,*,#}

¹ Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR_S1085, F-35000 Rennes, France.

² Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman,
15 WA 99164-4236, USA

³ Institute of Medical Sciences, School of Medicine, Medical Sciences & Nutrition, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK.

[#]To whom correspondence should be addressed

* These authors made an equal contribution

20 **Grant support:** This work was supported by l'Institut national de la santé et de la recherche médicale
(Inserm); l'Université de Rennes 1; l'Ecole des hautes études en santé publique (EHESP); the grant
XXXXXX awarded to B.J. [grant number XXXXXX].

Correspondence: Frédéric Chalmel, Inserm U1085-Irset, Université de Rennes 1, 263 av. du Général
Leclerc, 35042 Rennes cedex, France; Tel: +33 (0)2 23 23 58 02; Fax: +33 (0)2 23 23 50 55; Email:
25 frederic.chalmel@inserm.fr.

Abstract

Mammalian sex determination is a complex process driven by an orchestrated and sexually dimorphic
expression programs in developing fetal gonads. A comprehensive view of these transcriptional
30 landscapes in humans is essential for deciphering the molecular events at play during the sex-specific
gonad development and for understanding the etiology of unexplained disorders of sex development
(DSDs). Using RNA-sequencing, we performed an in-depth characterization of the human fetal gonadal
transcriptome from 6 to 17 weeks of development, a key period in sex determination and gonad
development. The analysis led to the reconstruction of 12'203 transcripts differentially expressed genes
35 during testicular and/or ovarian development. The comparison of ovaries and testes at the same age
enabled discriminating 7260 genes showing sex-biased expression either in the testis or in the ovary.
After differential expression analysis, we identified syntenic regions common to humans and mice. We
found that a fraction of syntenic lncRNAs also displayed conserved expression during fetal gonad
development. Finally, we next focused on 1'114 genes for which the sex-biased expression was
40 detected as early as 6GW and, among them, 33 lncRNAs and 17 novel genes. All data are also
conveniently accessible through the ReproGenomics Viewer at <http://rgv.genouest.org>.

Introduction

In mammals, sex differentiation stems from a critical decision made in the bipotential fetal genital ridges whether to adopt a testicular or ovarian fate. The expression of the Y-linked transcription factor SRY in Sertoli cells ridges activates the expression of the SOX9 transcription factor which subsequently promotes a highly orchestrated gene expression program leading to the formation of fetal testes (Koopman et al., 1991; Li et al., 2014; Rahmoun et al., 2017; Sekido and Lovell-Badge, 2008; Vidal et al., 2001). In the absence of SRY, RSPO1, WNT4, CTNNB1 and FOXL2 induce another complex cascade of transcriptional events, giving rise to fetal ovaries (Chassot et al., 2008; Liu et al., 2009; Ottolenghi et al., 2007; Schmidt, 2004; Uda et al., 2004; Vainio et al., 1999). To ensure the proper commitment to a given fate, the sex-specific pathways antagonize each other to repress the alternative fate (Bagheri-Fam et al., 2017; Chang et al., 2008; Greenfield, 2015; Jameson et al., 2012; Kashimada et al., 2011; Kim et al., 2006; Maatouk et al., 2008; Wilhelm et al., 2009). Both male and female expression programs underlying sex determination and gonadal differentiation are highly dynamic and complex, and many grey/dark areas exist, especially the noncoding counterpart of this transcriptional landscape that is likely to play critical roles. This is significantly contributing to the fact that the vast majority of disorders of sex development (DSDs) still remain genetically unsolved (Eggers et al., 2016).

Several valuable genome-wide expression studies have been conducted on whole fetal gonads in mammals (Houmard et al., 2009; Mamsen et al., 2017; Munger et al., 2013; Rolland et al., 2011; Small et al., 2005; del Valle et al., 2017) or isolated fetal gonadal cell populations (Beverdam and Koopman, 2006; Bouma et al., 2007, 2010; Nef et al., 2005) using microarray technologies. Recently single-cell transcriptomic analyses of thousand cells have revolutionized our understanding of gonadal cell lineages during fetal life (Guo et al., 2015, 2017; Li et al., 2017; Stévant et al., 2018). Although those technologies hold great potential for further investigations, as microarrays, they are currently limited

65 in their ability to study alternative splicing, decipher the noncoding expression program and discover novel candidate genes relevant for sex differentiation (Haque et al., 2017).

Alternatively, “bulk” RNA-sequencing (RNA-Seq) can be used to circumvent these issues. Several landmark studies have performed RNA-seq in the mouse to investigate the transcriptome of the fetal Leydig cells (Inoue et al., 2016; McClelland et al., 2015), the regulome of the key gonadal transcription
70 factor SOX9 (Rahmoun et al., 2017), and the transcriptome of the mouse developing fetal testes and ovaries at key developmental stages (Zhao et al., 2018). As such, there remains an important place for RNA-Seq analyses of whole-organ development in humans.

In the present study, we performed a strand-specific, ribo-depleted RNA-seq approach to analyze the transcriptome of the human developing fetal testes and ovaries at key developmental stages (from 6
75 to 17 gestational weeks) encompassing the time window from early transcription of the SRY gene in the male to differentiation and primary sex-specific control of the cell lineages in both gonads. Our results provide a comprehensive view of the protein-coding and noncoding transcriptional landscape underlying sex determination including 1114 genes for which the sex-biased expression was detected as early as 6GW and, among them, 33 lncRNAs and 17 novel genes. Furthermore, our analyses
80 uncovered a core group of protein-coding genes and lncRNAs transcribed syntenically during germ cell differentiation suggesting that they play key roles in this process. A graphical display of the corresponding dataset is conveniently available through the ReproGenomics Viewer (RGV, <https://rgv.genouest.org>) (Darde et al., 2015,submitted).

85 **Results**

The transcriptional repertoire of human fetal gonads

To unravel the expression program underlying gonadal differentiation in humans we undertook a large-scale RNA-seq experiment, including 24 fetal testes and 24 fetal ovaries and covering 7 developmental stages, from 6 to 17 gestational weeks (GW) (Figure 1). The sequencing of rRNA-
90 depleted, strand-specific libraries yielded an average of 46 million (46'416'264) 2x50 paired-end reads per sample, out of which 84.2% could be confidently mapped to the human genome. This primary analysis enabled the assembly and of 180'242 non-redundant transcripts and their quantification across all samples (Figure S1).

A stringent refinement protocol was further used to discard assembly artifacts and poor-quality
95 transcripts, leading to a set of 35'194 high-confidence transcripts, corresponding to 13'827 loci, expressed during human gonad development (Figure 2). We next performed a principal component analysis based on these transcripts' expression in order to get insights in samples homogeneity and in transcriptome dynamic (Figure 3). Consistently, the first two components appeared strongly correlated with the developmental stage (component 1) and the genetic sex (component 2) of samples (19.72%
100 and 15.34% of variance, respectively), thus providing a first indication regarding the biological relevance of this dataset.

After comparing the reconstructed transcripts with RefSeq and Ensembl reference annotations, most of them were identified as known (13'673; 38.85%) or novel (18'718; 53.19%) isoforms of annotated protein-coding genes, but they also included known (529, 1.5%) or novel (680, 1.93%)
105 isoforms of annotated lncRNAs. Importantly, this analysis also evidenced transcripts corresponding to completely new intronic (81, 0.23%), intergenic (164, 0.47%) or antisense (73, 0.21%) unannotated loci,

referred as novel unannotated transcripts (NUTs) in the following manuscript. All data can be displayed through the RGV genome browser (Darde et al., 2015, submitted; http://rgv.genouest.org/app/#/browser_genome).

110 Newly identified loci mainly correspond to novel noncoding genes

To characterize the protein-encoding potential (PEP) of novel isoforms and loci identified above, two approaches were combined. First we used four distinct prediction tools and transcripts predicted as potential protein-coding by at least two tools were classified as high PEP. Second, a proteomics informed by transcriptomics approach (PIT) was conducted by querying mass spectrometry data from human fetal gonads (Kim et al., 2014) against a peptide sequence database derived from our reconstructed transcripts.

As expected, 98.52% of mRNAs (known and novel isoforms) displayed a high PEP, and 56% of them were evidenced at the protein level thanks to at least one identified peptide (Figure 4). On the other hand, only 48.06% of lncRNAs (known and novel isoforms) were predicted to have a high PEP, and as little as 8.52% were identified during the PIT analysis, usually through a single peptide identification (Figure 4). When NUTs were finally evaluated, they mostly displayed features similar to that of lncRNAs: 76.1% of them were predicted to have a low PEP while only 0.4% were evidenced at the protein level (Figure 4). The combination of these two approaches therefore strongly suggest that newly identified loci correspond to noncoding genes. Besides, 94 lncRNAs and 1 NUT that were both predicted to have a high PEP and evidenced at the protein level could be good candidates for novel protein-coding genes, and require further experimental validations.

A selected locus (locus ID: XLOC_173972; transcript ID: TCONS_00254891) was identified from one high-confidence 21aa long peptide. It maps on chromosome X (positions 134227400-134232793, minus strand) and is composed of 2 exons with a cumulative exon size of 4483 nt and a maximum ORF

130 size of 121 aa. This transcript exhibits a peak expression in testes from 9GW to 13-14GW in the testis. It was firstly identified as a lncRNA, LINC00087, which changed name and coding status recently, to become SMIM10L2B, a small integral membrane protein 10 like 2B. This novel protein is indexed as “protein predicted” by UniprotKB (P0DMW5), and even if the identified peptide had previously been detected in fetal testes, fetal ovaries, adrenal glands and brain, there is still no direct evidence of the full-length protein. The detection of conserved domain on the predicted protein sequence reveal the presence of a DUF4560 superfamily domain (E-value 1.47e-30), which function is currently uncharacterized. Therefore, the PIT approach revealed the presence of a peptide and its corresponding transcript simultaneously in human fetal gonads, bringing more evidence for the existence of a new protein, which function remain unclear to date.

140

Expression dynamics during human fetal gonad development.

We next focused our analysis on genes that are more likely to play important roles during gonad development according to their differential expression between sexes and/or between developmental stages (Figure 5). We identified 13'145 transcripts (7'633 genes) that were differentially-expressed (fold-change ≥ 2 , LIMMA statistical test $FDR \leq 5\%$) between testes and ovaries (called SDT for Sexually Dimorphic Transcripts), most of which (10'521 transcripts, 6'587 genes) being also developmentally-regulated (fold-change ≥ 2 , LIMMA statistical test $FDR \leq 5\%$). On the other hand, 8'935 transcripts (5'961 genes) were differentially-expressed during gonad development but did not show significant difference between sexes (called NSDT for “Non-Sexually dimorphic Transcripts). These two sets of transcripts were further classified according to peak expression into 14 and 6 expression patterns, respectively (Figure 6; Figure S3).

Several patterns could be distinguished between genes preferentially expressed in testes (Figure 6A-B), including peak expression at 6 to 7 GW (P1; SRY, DMRT1), at 7 GW (P2; SOX9, DHH, INHBB), (P3; NROB1, NR5A1), at 9 to 12 GW (P4; CYP11A1, CYP17A1, STAR, LHCGR, INLS3), at 13 to 17 GW (P6; ACTA2, PTGDS) or with a broader expression throughout testis development (P3; NR5A1, NROB1, and P5; WT1, CLDN11). Similarly, genes preferentially expressed in ovaries showed distinct profiles, including peak expression at 7 GW (P8; RPSO1, AMHR2, NANOG), at 7 to 9 GW (P9; POU5F1, DPPA2/4, LIN28A/B), at 12 GW (P11; FOXL2, DAZL, PIWIL2), at 13 to 14 GW (P12), at 17 GW (P13; MEI1, MEIOB, SPO11, SYCP1-3, and P14; FIGLA, NOBOX, SOHLH1) or with a broader expression throughout ovarian development (P10; DMC1).

In addition to known genes involved in gonad development and listed above, the relevance of expression patterns was also evaluated by Gene Ontology term analysis (Figure 6C). Several broad functions relevant to organogenesis and cell differentiation were enriched in various male (P1, P2, P5, P6) and female (P8) clusters. Noteworthy is that expression P4 was found enriched in genes involved in steroidogenesis, which is consistent with the differentiation and development of Leydig cells from 7 GW onwards. On the other hand, several processes related to meiosis and female germ cell development were found enriched in expression patterns P13 and P14, in agreement with the commitment of ovarian germ cells into meiosis from 12 GW onwards.

Furthermore, developmentally-regulated genes that do not show (statistically-significant) expression differences between testes and ovaries (Figure S3) include transcripts expressed at early stages of gonad differentiation (6 to 7 GW; P1 and P2; WT1, GATA4), following sexual differentiation (7 to 9 GW; P3 and P4; DMRT1, SOX8) or at latter stages of gonadal development (12 to 17 GW; P5 and P6; NR6A1). While these transcripts are also likely to include factors with important roles during gonadal differentiation and development, they were not investigated further in this study. All data, however,

175 are provided as supplementary information (Tables S1) and are also available through the RGV genome browser (Darde et al., 2015, submitted; http://rgv.genouest.org/app/#/browser_genome)

Finally, when investigating the distribution of RNA biotypes within groups of co-expression we found that lncRNAs and NUTs were enriched in expression patterns P12, P13 and P14 (Figure S2) as well as Q1 and Q6 (Figure S4).

180

Searching for factors involved in early human gonad differentiation.

Early steps of testis differentiation are driven by factors induced shortly after the expression of the sex-determining gene SRY, *i.e.* from 5 to 6 GW onwards in humans (Mamsen et al., 2017). Reciprocally, the onset of ovarian development requires early male pathways to be repressed as well as the establishment of a female-specific expression program. In order to highlight factors that could play important roles during these first stages of gonadal differentiation, we focused on transcripts exhibiting early sexual dimorphism, *i.e.* with fold-change of at least 2 between testes and ovaries as early as 6 GW. This filtration step selected 1'479 transcripts, *i.e.* 11.25% of all SDT, and are called eSDT for “early SDT” (Tables S1).

190 While some eSDT could be found in all expression patterns, most of them (61.73%) were logically found in P1, P2, P8 and P9 (Figure S5), whereas these 4 clusters cover only 28.44% of all SDT (Figure S5). The relevance of this filtration is notably evidenced by the selection of SRY (P1), SOX9, DHH, PTCH1 and CYP26B1 (P2), LHX9 (P6), ACVR1B and (P8), AMHR2 (P9) or FOXL2 (P11), among other known genes involved in early gonadal differentiation. Consequently, this narrowing down of SDTs to early ones resulted in an enrichment for genes involved in “Sex differentiation” within expression pattern P2 (Figure S5). These eSDTs therefore represent a promising resource of factors likely to play important roles during human gonad differentiation, including 1459 transcripts corresponding to known protein-

195

coding or noncoding genes, as well as 20 NUTs. Several of such candidates were selected for further investigation (see next section).

200

Cellular investigation of new candidates genes

To investigate further in which cell type(s) selected genes were expressed we performed immunohistochemistry experiments as well as qPCR on FACS-sorted cells. The successful enrichment of Sertoli cells (hEpA+/KIT-), germ cells (KIT+) and interstitial cells (hEpA-/KIT-) from 6-7 GW testes was notably validated by the high expression of *SOX9*, *KIT* and *NR2F2*, respectively (Figure S6A-C), while that of germ cells (KIT+) and somatic cells (KIT-) from 6-8 GW and 10-12 GW ovaries was confirmed by the high expression of *KIT* and of *FOXL2* and *NR2F2*, respectively (Figure S6D-F).

We next investigated genes that display expression profiles similar to that of *SRY* (P1; high expression and clear sexual dimorphism in 6 GW testes). We found that *SRY*, *WLS*, and *THBS1* were mainly expressed by Sertoli cells with a substantial expression in interstitial cells, while *CXCL14* was found to be expressed only in Sertoli cells (Figure 7). Surprisingly, *CCR1* was found to be mainly expressed in germ cells. We also analyzed genes with a *SOX9*-like expression profile (P2; clear sexual dimorphism and peak of expression in 7 GW testes) (Figure 8). We found that *EPHB1*, *FATE1*, *MAGEB1* and *ERBB3* were indeed expressed in Sertoli cells and to a lesser extent in interstitial cells, while *SOX10* was expressed at similar levels in Sertoli cells and interstitial cells. Interestingly, we found that both *CITED1* and a potential corresponding antisens RNA, *TCONS_00249587*, display the same developmental expression pattern as *SOX9* and are also specifically expressed in Sertoli cells. Consistently with qPCR results, *SOX9*, *SOX10*, *KIAA1210*, *EPHB1*, *MAGEB1*, and *FATE1* proteins were indeed all found in cord cells at the histological level, but exhibited varying ratio of expression in Sertoli

220 and germ cells. For instance, *SOX9* was expressed in Sertoli cells, whereas *MAGEB1* was clearly expressed in germ cells too (Figure S7).

We also investigated the cell distribution of female-biased genes that showed early sexual dimorphism, i.e. in 6 GW ovaries (P8, Figure 9) and/or 7 GW ovaries (P9, Figure 10). Similarly to *WT1* and *PTGS1* we found that *NPY* and the NUT *TCONS_00153406* were specifically expressed in somatic cells from 6-9 GW and 10-12 GW ovaries (Figure 9). By contrast, other genes such as *NRXN3*, *CNTN1*, *SOX4* and *SET* also showed a substantial expression in germ cells while another NUT, *TCONS_00113718*, was found to be preferentially expressed in germ cells (Figure 9). In agreement, immunolabeling showed *NRXN3* protein in the nucleus of cells surrounding *KIT*⁺ germ cells, with a pattern very similar to that of *WT1* (Figure 9). Interestingly, the *CNTN1* protein was found in 5-6 DW old ovaries in a subset of epithelial cells surrounding *LIN28*⁺ germ cells, but neither in the surface epithelium nor in ovarian cords adjacent to the mesonephric-gonadal junction (Figure 9). On the other hand, genes from the expression pattern P9 appeared to display expression profiles that are typical of the germ-cell lineage, such as *KIT* or *NANOG* (Figure 10). This germ cell specific expression was further confirmed for two NUTS, *TCONS_00055038* and *TCONS_00042656*, both in 7-9 and 10-12 GW ovaries (Fig. XX). Conversely a third NUT, *TCONS_00224470*, was found to be preferentially expressed in somatic cells, alike *RSPO1* (Figure 10).

The conserved expression program underlying gonadal development

In order to compare transcriptome dynamics during mammalian gonad development, we integrated a mouse RNA-seq dataset of mouse fetal gonads at three developmental stages (Soh et al., 2015). The corresponding reads were first mapped on the mouse genome, yielding an average of 84.7% of aligned reads. The genome coordinates of the 35'194 high-confidence human transcripts assembled

from our dataset were next converted to the mouse genome, allowing to identify 30'491 syntenic regions. These regions then were then used as a reference mouse transcriptome and were quantified thanks to mapped reads. A total of 17'026 syntenic transcripts (9'610 genes) were found significantly expressed during gonad development in both organisms, among which 10'405 transcripts (6'782 genes) were differentially expressed (SDT or NSDT) in human (Figure S8).

Prior to evaluating to which extent the expression of syntenic transcripts is conserved throughout gonadal development, we first had to define which developmental stages would be compared, for testes on the one hand, and for ovaries on the other hand. We therefore assessed all pairwise comparisons and selected those leading to the highest number of correlated transcripts. The retained comparisons for testicular development were early 7, 9 and 12 GW in human with 12,5, 14.5 and 16.5 dpc in mouse, respectively. Regarding the ovary, 9, 12 and 13-14 GW in human were finally compared to 12,5, 14.5 and 16.5 dpc in mouse, respectively. With these matched developmental stages, we highlighted 1'288 syntenic transcripts (1'212 genes) that exhibit a correlated expression profile ($p\text{-value} \leq 0.1$) during mouse and human fetal gonad development, including 739 SDT and 549 NSDT (Figure 11). These genes represent primary candidates for investigating further the functional involvement, i.e. through mouse genetic engineering, during gonadal differentiation in mammals.

260 Discussion

This large scale RNA-seq study is the first one to characterize the transcriptional program at play during human gonad development over such a large developmental window, ranging from early differentiation (i.e. 6 GW) up to Leydig cell transition in the testis of follicle formation in the ovary (i.e. 17 GW). Indeed, recent studies also described the transcriptome from human fetal testes and ovaries, 265 but they did not go beyond 10 GW (Mamsen et al., 2017; del Valle et al., 2017)(Mamsen et al., 2017; del Valle et al., 2017). Furthermore these studies were based on microarray technology, thus restraining the set of studied genes, somehow limiting the characterization of non-coding transcripts, and completely preventing the identification of new genes, unlike in our own analysis. Actually, most of their results were confirmed and completed by our data, such as the newly reported preferential 270 expression of *ETV5* in the ovary, or the over-expression of genes such as *PTCH1* and *GLI1* in the testis (Mamsen et al., 2017). Yet, our data validate the onset of *SRY* expression prior to 6 GW (detected at 5.5 GW in (Mamsen et al., 2017)), but also demonstrate the over-expression of *SOX9* in the testis as early as 6 GW, previously reported at only 48dpc (6,8GW) (Mamsen et al., 2017). This may reflect a higher sensitivity and therefore a more accurate expression profiling (Wang et al., 2014) of RNA-seq as 275 compared to microarrays.

RNA-seq allows to study gene expression at the transcript isoform level, which enables to highlight potential splicing specificities and, more importantly to discover new genes that had never been described. Using a stringent refinement protocol, we defined here a set of 35'194 transcripts that are significantly expressed during gonad development, among, including 32'391 mRNAs and 1'209 280 lncRNAs (known or novel isoforms), as well as 318 entirely new genes. The analysis of testes and at seven time point and with several replicates allowed us to apply a stringent statistical filtration in order to overcome issues related to inter-individual heterogeneity of human samples. While RNA-seq

currently appears the most reliable technology to study the transcriptome, it still appears that some assembly artefacts may occur. Transcript assembly can indeed generate chimeric transcripts or even
285 coin-join genes, i.e. the fusion of transcripts from two different genes into a single one. One striking example in our dataset is that of AMH. Indeed, several transcripts annotated as originating from SF3A2, a gene located in upstream region of AMH, were indeed reconstructed and showed peak expression in early 7 GW testes. After manual inspection of alignment bam files, we concluded that some reconstruction errors had been introduced during assembly steps. The comparison of read density in
290 exonic regions of these two genes across all samples indeed indicate that the expression variations of SF3A2 transcripts are more likely due to reads mapping AMH exons. It should be noted that these transcripts are already detectable at 6GW in the testis, leading to the hypothesis that AMH is specifically induced in the testis as early as 6 GW in the testis.

We decided to classify sexually-dimorphic transcripts into 14 clusters which seemed a good
295 compromise to handle the complexity of all possible expression profiles at play during gonad development. Despite this, and while several known markers seemed to accumulate within specific expression patterns, not all of them were found significantly associated with potentially expected GO terms. This may account for the high heterogeneity and concomitance of many biological processes associated with co-expression groups. In addition to GO terms and known markers, we also
300 characterized expression patterns according to RNA biotype distribution. We notably found that P12 to P14, which are mostly characterized by the commitment to and progress through meiosis, were enriched for lncRNAs and NUTs. Comparable observations have been made within the adult testis, in which lncRNAs were overexpressed in germ cells, from meiosis onwards (Cabili et al., 2011; Chalmel et al., 2014; Laiho et al., 2013). Our data therefore demonstrate the conservation of this phenomenon in
305 female germ cells and suggest that non-coding genes may play particular roles during meiosis.

Our dataset notably enabled to investigate early molecular events of human gonad differentiation. The selection early sexually-dimorphic profiles, i.e. as soon as 6 GW, highlighted 1'479 transcripts, including an unneglectable set of genes which have never been described as potential actors of gonad differentiation. Of note is that testes and ovaries are still morphologically undistinguishable at 6 GW, therefore the expression differences observed at this time are more likely to result from specific transcriptional activation or repression. The GO term analysis performed on eSDT notably showed an enrichment of genes involved in "steroid biosynthetic process", including some well-known markers of Leydig cells. Mamsen and collaborators deduced from their experiment that the onset of steroidogenesis occurred at 53 dpc (i.e. at 7.5 GW), whereas our data indicate a differential expression from 6 GW onwards, suggesting that male-biased expression of steroidogenic factors is induced much earlier than previously thought. Therefore, it appears that the molecular networks underlying Leydig cell differentiation is already ongoing at 6 GW, at least in a significant subset of testicular somatic cells. A similar observation can be made with several primordial germ cell markers, which expression are often higher in the ovary at 6GW. Therefore, many sex-specific transcription pathways required during later stages of development seems to be triggered early during gonad differentiation.

Our analysis focused on SDTs in order to enlighten factors playing potential important roles during sexual differentiation. Nonetheless, NSDTs represent a non-unneglectable counterpart of molecular events driving gonad development. Considering the quantification biases due to analyzing the "bulk" transcriptome from whole gonads, it should be kept in mind that changes in expression observed at late developmental stages are greatly influenced by the fluctuating proportions of various cell lineages within each organ. Considering the heterogeneity of cell composition between testis and ovary, the identification of so many transcripts expressed at similar level until advanced development

stages was unexpected. Still, this common set of synchronous transcription may constitute the basis
330 for the setting-up of organ development, on which sex specific structures can be established.

The identification of relevant factors among the plethora of candidates that are typically
provided by large-scale expression studies is often a tedious task. While it is certainly not a systematic
rule, many known important factors during a given biological process appear evolutionary conserved,
at both the sequence and expression levels. Comparative transcriptomics is an emerging field of
335 research which primary aims at studying genome and gene expression evolution. Such approaches are
also of potential interest to determine if model organisms, including rodents, are relevant for studying
human physiology, by identifying conserved biological processes and molecular pathways. Several
studies have demonstrated that organs exhibiting a high degree of tissue-specific gene expression, such
as the adult brain and the adult testis, also tend to have a more conserved (Chan et al., 2009; Lin et al.,
340 2014; Su et al., 2002; Sudmant et al., 2015). Considering lncRNAs, their sequences are less conserved
(Derrien et al., 2012) and their expression patterns across tissues diverge at a higher rate as compared
to protein-coding genes (Necsulea et al., 2014). The comparative transcriptomic analysis of two
developing organs in two species presented many challenges that we intended to overcome: (i) These
organs are made of several cell types, each present in distinct proportions and expressing their own set
345 of genes, which average constitutes the whole transcriptome; (ii) the development of these organs
involves the appearance of new cell type as well as the proliferation and/or differentiation of existing;
(iii) the kinetics of differentiation processes within the testis and the ovary are known to be different
between the two organisms, spanning a large time lapse in humans, and being faster and more
synchronous in the mouse.

350 To solve these issues we decided to rely neither on prior knowledge regarding orthology
relationships nor on pre-established correspondences between matching developmental stages in

these two species. First, our method based on pairwise genome alignment enlightened interesting features of some co-expression groups. It indeed appears that unmapped genes are overrepresented in expression patterns P8 as well as in P12 and P14, which correspond to transcripts implicated in early
355 ovary differentiation and in meiosis and oogenesis, respectively. This is consistent with the fact that lncRNAs and NUTs, which are poorly conserved, also accumulate in these groups. Second, selected associations between the human and mouse time points, which were based on comparison that maximize the number of correlated transcripts, seem to be relevant to the current knowledge of gonad differentiation. In the testis, 12.5dpc is associated with early 7 GW, right after sex determination and
360 at the onset of Sertoli cell differentiation. 14.5 dpc is associated to 9 GW, corresponding to Leydig cell differentiation. 16.5dpc was associated with 12GW, when both Sertoli cells and Leydig cells are fully developed. In the ovary, all mouse time points are associated with more advanced human developmental stages as compared to the testis, i.e. 9, 12 and 13-14 GW, which is also the delayed and slower ovarian development.

365 Through this analysis we enlightened 1'125 syntenic transcripts which expression dynamics during gonad development is conserved between human and mouse. These included 722 SDT among which 78 eSDT and 6 lncRNAs, and 403 NSDT. A reasonable hypothesis is that such genes may have homologous molecular function in both species, and may therefore be important factors during gonad development, gathering further investigations. Such a comparative transcriptomic approach may be
370 applicable to other biological systems, but it could be improved by using more homogeneous or comparable datasets. Most published works focused on specific processes, e.g. female meiosis (Soh et al., 2015), gonad differentiation (Zhao et al., 2018) or Leydig cell development (McClelland et al., 2015), but none covered the entire gonad development process with RNA-seq.

This large time scale study deciphered the transcriptional program governing many biological processes occurring during gonad development, and its potential of discovery has not been fully explored, making it also a useful resource to answer additional questions in the future.

Materials and Methods

Detailed description of experimental procedures is provided in SI Materials and Methods. RNA-seq data from all sources and all species were processed identically by previously published approaches (Zimmermann et al., 2015) based on the Tuxedo and AMEN suites (Chalmel and Primig, 2008; Pollier et al., 2013). Sequencing data files have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE70361. All data are also conveniently accessible through the ReproGenomics Viewer (Darde et al., 2015, submitted).

List of abbreviations

bp: base pairs; DE: differentially expressed; (l)ncRNAs: (long) noncoding RNAs; miRNA: microRNA; NUTs: novel unannotated transcribed regions; PCA: principal component analysis; PIT: proteomics informed by transcriptomics; RNA-seq: RNA sequencing for expression quantification; qPCR: quantitative polymerase chain reaction.

Acknowledgments

We thank all members of the SEQanswers forums for helpful advice; Steven Salzberg and Cole Trapnell for continuous support with the “Tuxedo” suite; and the UCSC Genome team members. Sequencing

was performed by the IGBMC Microarray and Sequencing platform, member of the France Genomique program.

395 **Authors' contributions**

FC, ADR, and EL wrote the manuscript. FC and ADR supervised the research. EL and FC prepared, analyzed, and interpreted data. ADR, SMG, IC, PF, PAF, MBM and BJ prepared the samples and interpreted sequencing data. BE, ADR and SMG validated expression data. SMG and BJ contributed to the manuscript. All authors approved the final version of the manuscript and declare that they have no

400 competing interests.

Figures

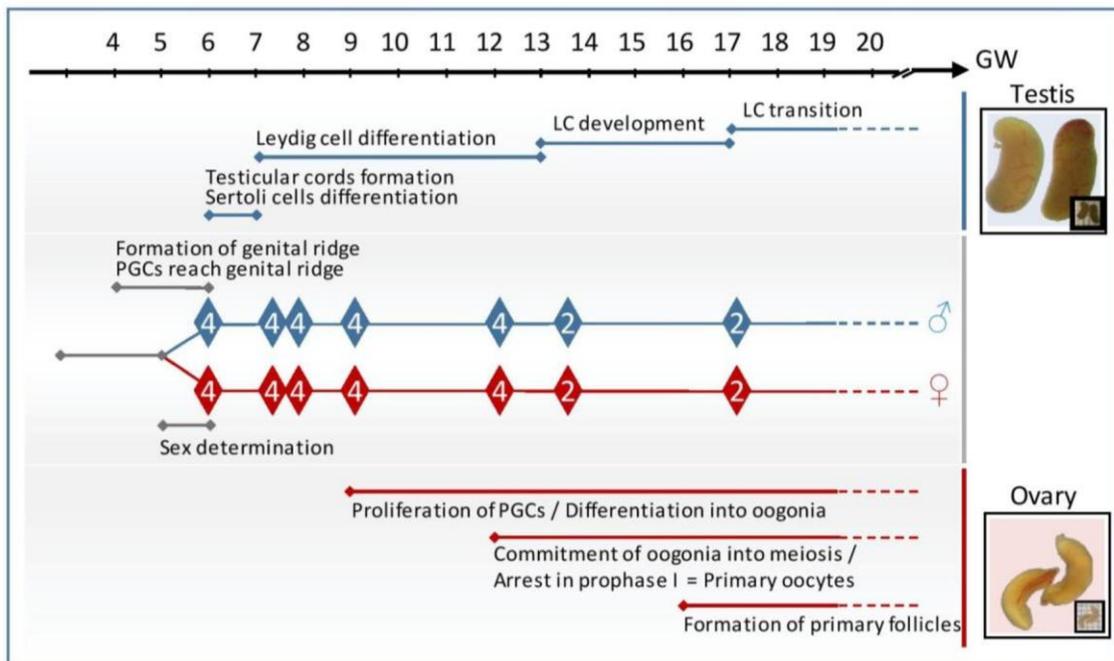


Figure 1. Sample collection to investigate human gonadal development

Human fetal gonads used in this were collected at 7 developmental stages, i.e. at 6, early 7, late 7, 9, 12, 13-14 and 17 gestational weeks (GW). The number of replicates (4 or 2) is indicated for each stage and sex. An overview of the main differentiation processes within human fetal testes and ovaries during the studied time window is also provided. PGC = primordial germ cell; LC = Leydig cell.

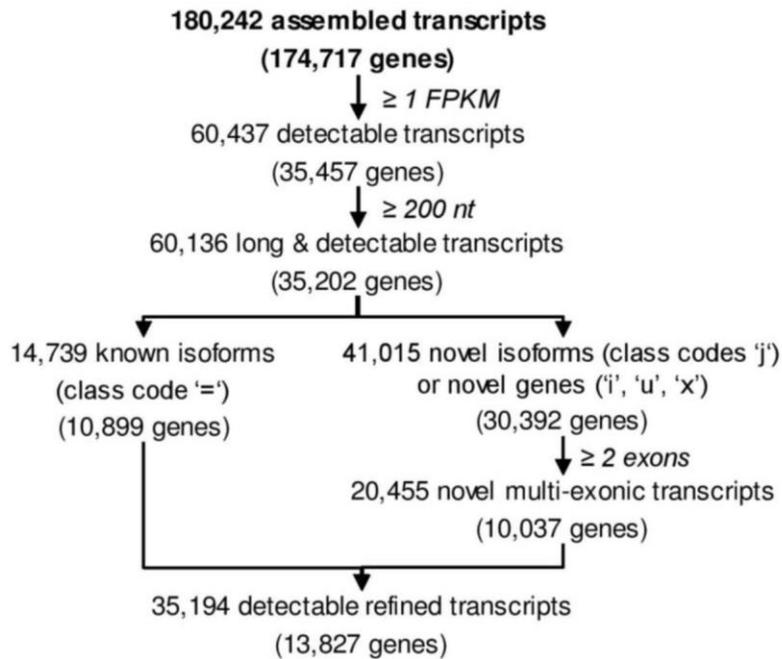


Figure 2. Refinement strategy of assembled transcripts

Following transcript reconstruction with Cufflinks, a refinement strategy was performed to discard sequencing and assembly artifacts: (i) only transcripts with an expression of ≥ 1 FPKM in at least one experimental condition (average value of biological replicates) were considered; (ii) transcripts with a length of less than 200 nucleotides were discarded; (iii) novel transcript isoforms (Cuffcompare class “j”) and genes (classes “i”, “u” and “x”) were required to harbor at least 2 exons. FPKM = fragments per kilobase of exon model per million reads mapped; nt = nucleotide.

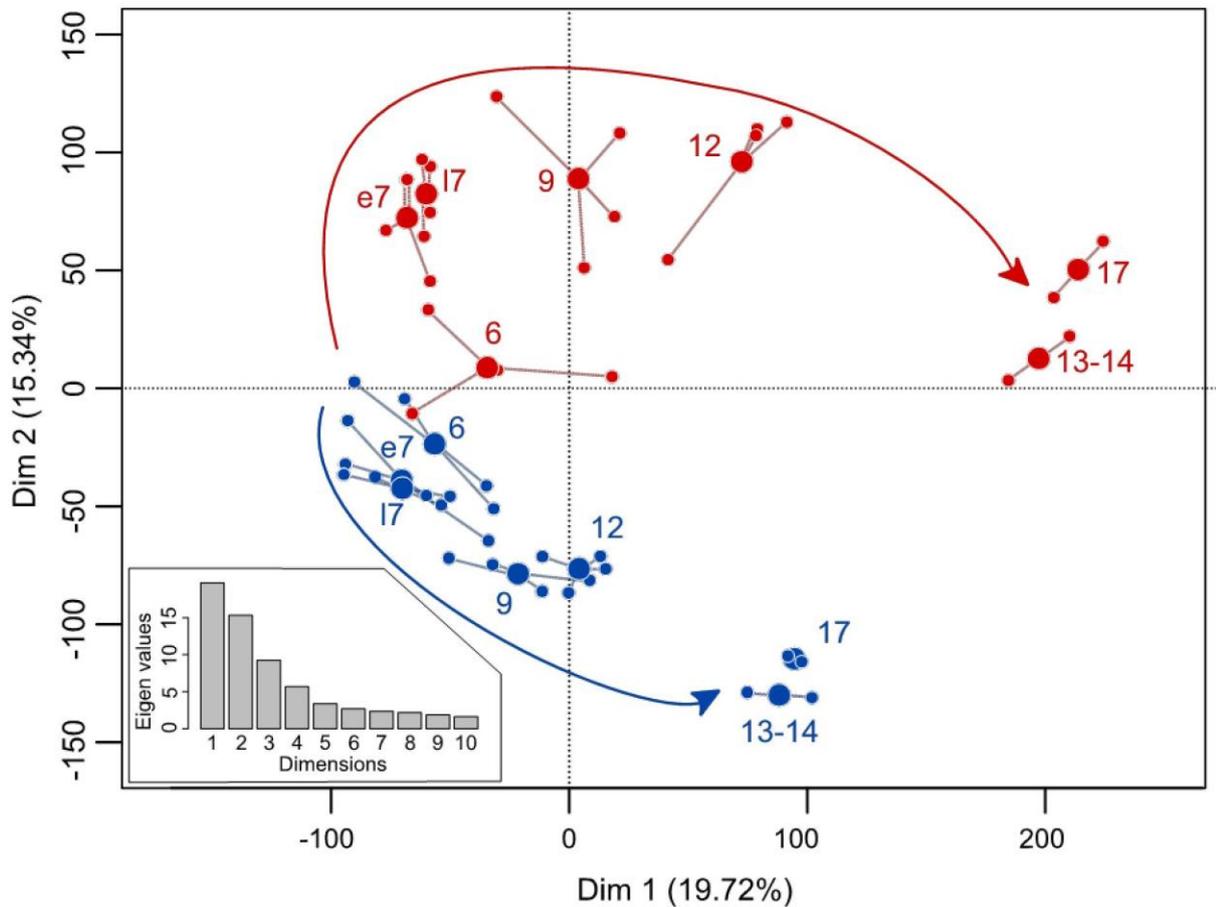


Figure 3. Principal component analysis of refined transcripts

A principal component analysis (PCA) was performed on expression data from 35'194 refined transcripts across all 48 human fetal gonads. Scatter plot represents the position of each samples along the first two dimensions (Dim1: 19.72% explained variability; Dim 2:15.34% explained variability). Smaller dots represent each replicate, and are linked to bigger dots that represent the average expression of transcripts across replicates. The histogram represents the percentage of information carried by each dimension of the PCA. Testis samples are coloured in blue, while ovaries are in red. Time point of each condition is provided in gestational weeks. e7 = early 7 GW; l7 = late 7 GW. The variability between the samples is mainly explained by their age of development (dimension 1) and by their genetic sex (dimension 2). The two arrows highlight the divergence of gonads transcriptomes, from a common origin (at 6 GW) to their distinct fate.

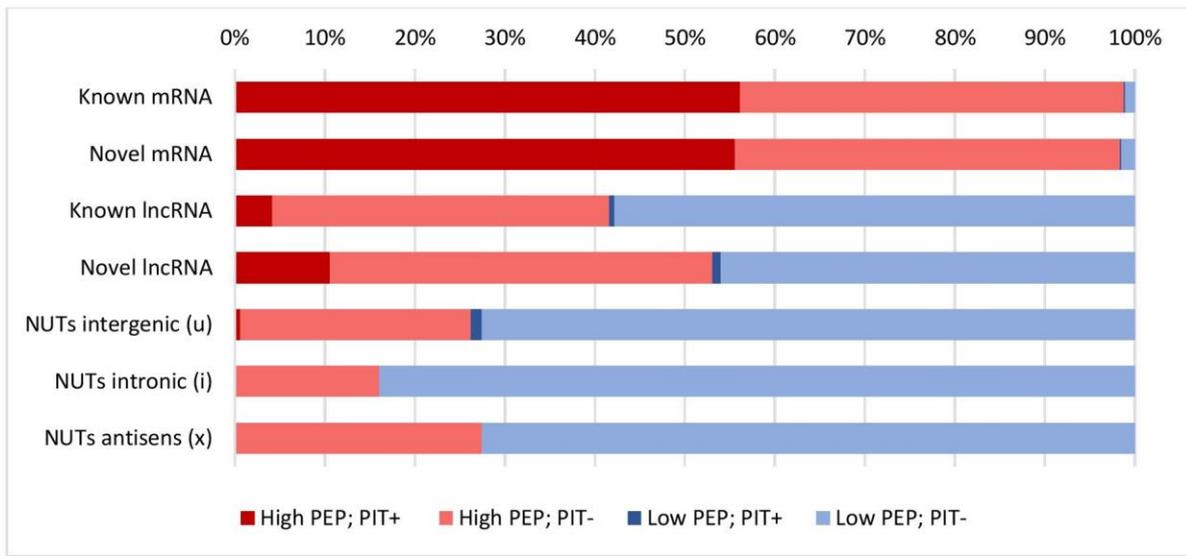


Figure 4. Coding potential assessment of refined transcripts.

Two complementary approaches have been performed to evaluate the coding potential of assembled transcripts. First, the protein-encoding potential (PEP) was predicted by four different prediction tools (CPAT, HMMER, CPC and txCdsPredic). Transcripts predicted as potential protein-coding by at least two tools are classified as “High PEP”, the other ones as “Low PEP”. Second, a Proteomics informed by transcriptomic (PIT) approach was performed to query MS/MS protein data from human fetal gonads against a custom reference protein database that include predicted protein sequences derived from assembled transcripts. The combined results of these two strategies, i.e. Low or High PEP transcripts with (PIT+) or without (PIT-) identified peptide(s) are represented for each transcript biotype.

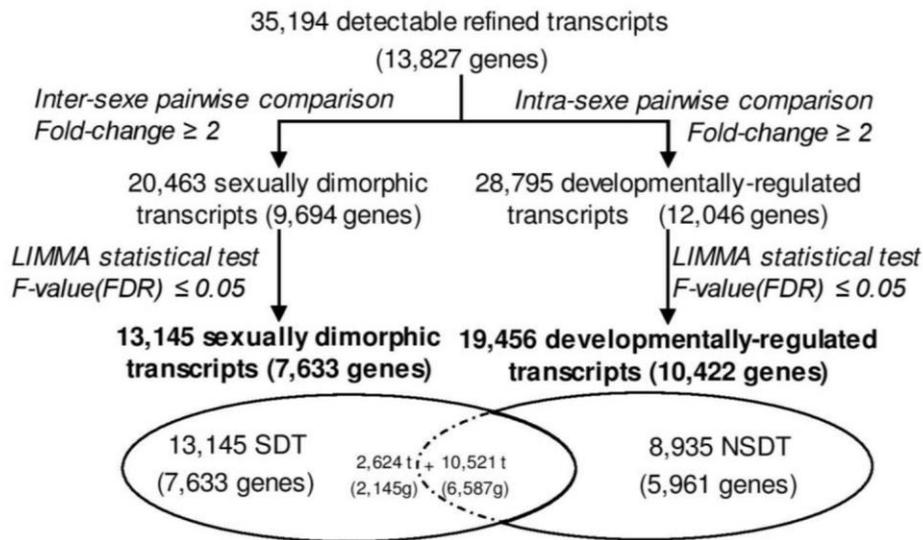


Figure 5. Strategy for differential expression analysis of refined transcripts.

Two statistical filtrations were used to select differentially-expressed transcripts. First, we performed an “intra-sex” comparison in which all developmental stages were compared to each other during testis development on the one hand, and during ovarian development on the other hand (Fold-change ≥ 2 in at least one comparison). Second, we performed an “inter-sex” comparison in which testes and ovaries were compared at each developmental stage (Fold-change ≥ 2 in at least one comparison) Subsequently, a LIMMA statistical test was performed on both sets of transcripts to select those with significant expression variation across replicates (FDR-adjusted F-value of ≤ 0.05). A total of 13,145 transcripts that display “inter-sex” expression variations were defined as “Sexually-dimorphic transcripts” (SDT), while 8,935 developmentally-regulated transcripts that do not exhibit sexual dimorphism were defined as “Non-sexually dimorphic transcripts” (NSDT)

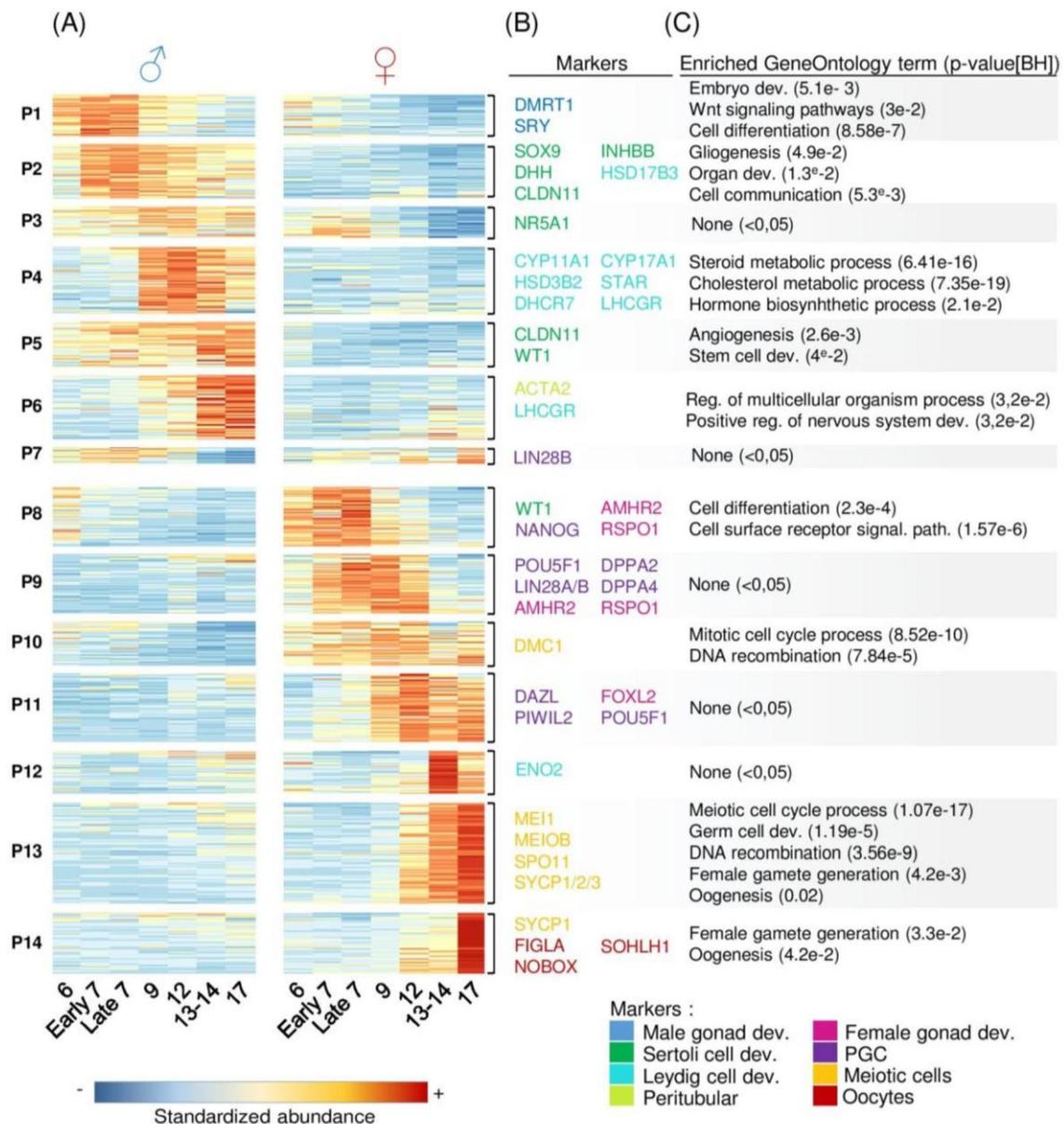


Figure 6. Sexually-dimorphic expression patterns during human gonad development.

(A) Heatmap representation of 13,145 sexually-dimorphic transcripts (SDTs), distributed into 14 expression patterns (P1 to P14), across seven developmental stages for both testes and ovaries. Each row is a transcript, and each column is an experimental condition, *i.e.* the average of testes or ovaries from a given gestational week (GW). The standardized abundance of transcripts is color-coded according to the scale bar, red corresponding to the highest expression level, blue to the lowest. **(B)** Repartition of known markers involved in gonad differentiation and development within SDT expression patterns. Note that several isoforms of a given can be assembled and display distinct expression. **(C)** Gene Ontology terms found to be enriched (BH corrected p-value < 0.05) in each expression pattern.

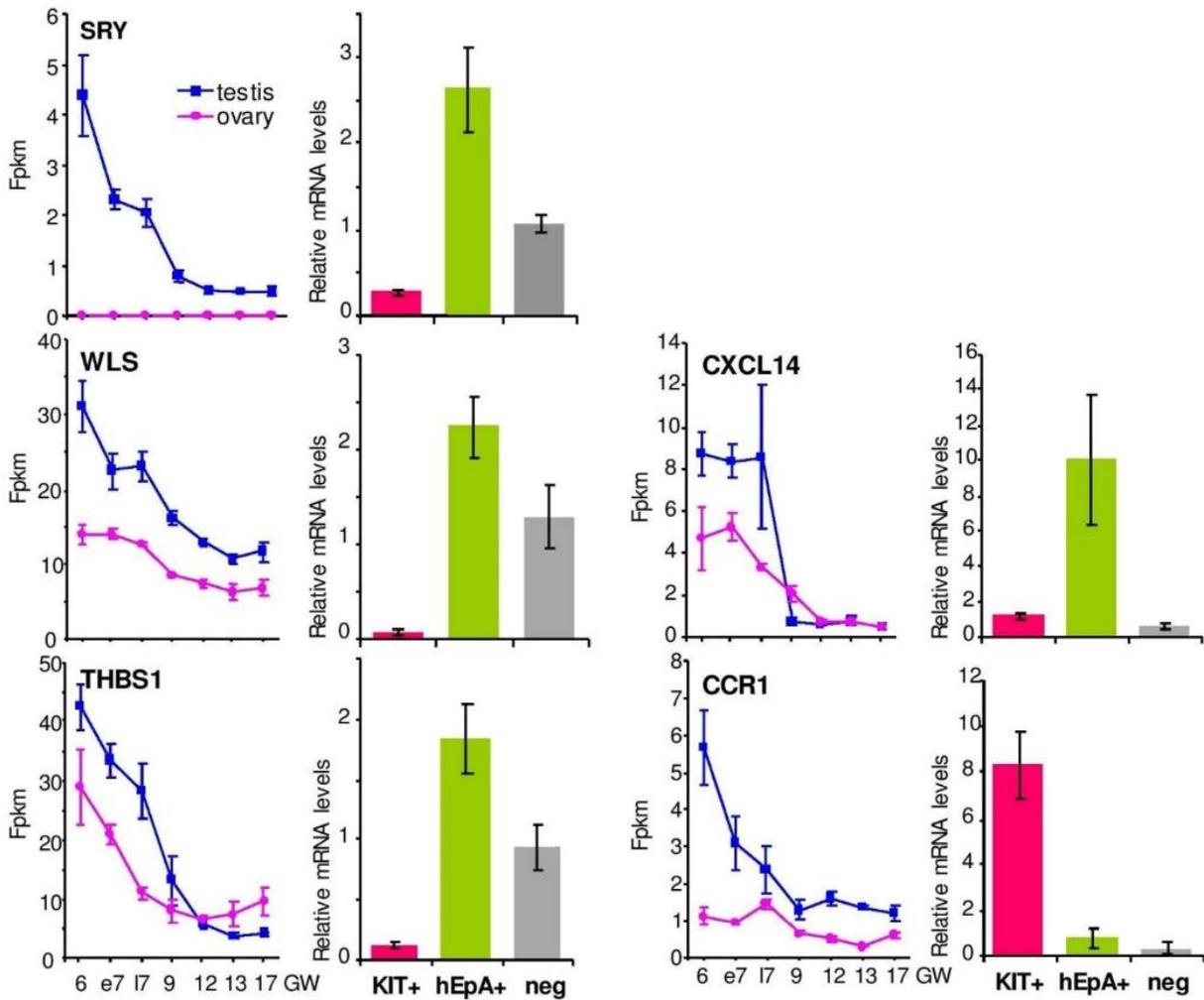


Figure 7. Cellular investigation of eSDTs from expression pattern P1.

Expression levels (left panels) and quantitative RT-PCR (right panels) of sex determining region Y (*SRY*), wntless Wnt ligand secretion mediator (*WLS*), thrombospondin 1 (*THBS1*), C-X-C motif chemokine ligand 14 (*CXCL14*) and C-C motif chemokine receptor 1 (*CCR1*). Expression levels from RNA-seq as a function of age are depicted as blue lines for the testis and pink lines for the ovaries. Each point represents the mean FPKM \pm SEM of the levels measured in 4 (14 GW and younger) and 2 different gonads (15 and 19 GW). Quantitative PCR was performed on the different testicular sorted cell populations of germ cells (KIT+, red bars) Sertoli cells (hEpA+, green bars) and other cells (KIT-/ hEpA-, grey bars). Each column shows a pool of sorted cells from 5 fetal (6.9-7.3 GW) testes. Each bar represents the mean \pm SEM of the fold change in target gene expression relative to the reference genes *RPLP0* and *RPS20*. FPKM = fragments per kilobase of exon model per million reads mapped.

Figure 8. Cellular investigation of eSDTs from expression pattern P2.

Expression levels (line graphic) and quantitative RT-PCR (histograms) of SRY-box 9 (*SOX9*), SRY-box 10 (*SOX10*), EPH receptor B1 (*EPHB1*), fetal and adult testis expressed 1 (*FATE1*), MAGE family member B1 (*MAGEB1*), erb-b2 receptor tyrosine kinase 3 (*ERBB3*), Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 1 (*CITED*) and *CITED* antisense RNA (*TCONS_00249587*). Expression levels from RNA-seq as a function of age are depicted as blue lines for the testis and pink lines for the ovaries. Each point represents the mean FPKM \pm SEM of the levels measured in 4 (14 GW and younger) and 2 different gonads (15 and 19 GW). Quantitative PCR was performed on the different testicular sorted cell populations of germ cells (KIT+, red bars) Sertoli cells (hEpA+, green bars) and other cells (KIT-/ hEpA-, grey bars). Each column shows a pool of sorted cells from 5 fetal (6.9-7.3 GW) testes. Each bar represents the mean \pm SEM of the fold change in target gene expression relative to the reference genes *RPLP0* and *RPS20*. FPKM = fragments per kilobase of exon model per million reads mapped.

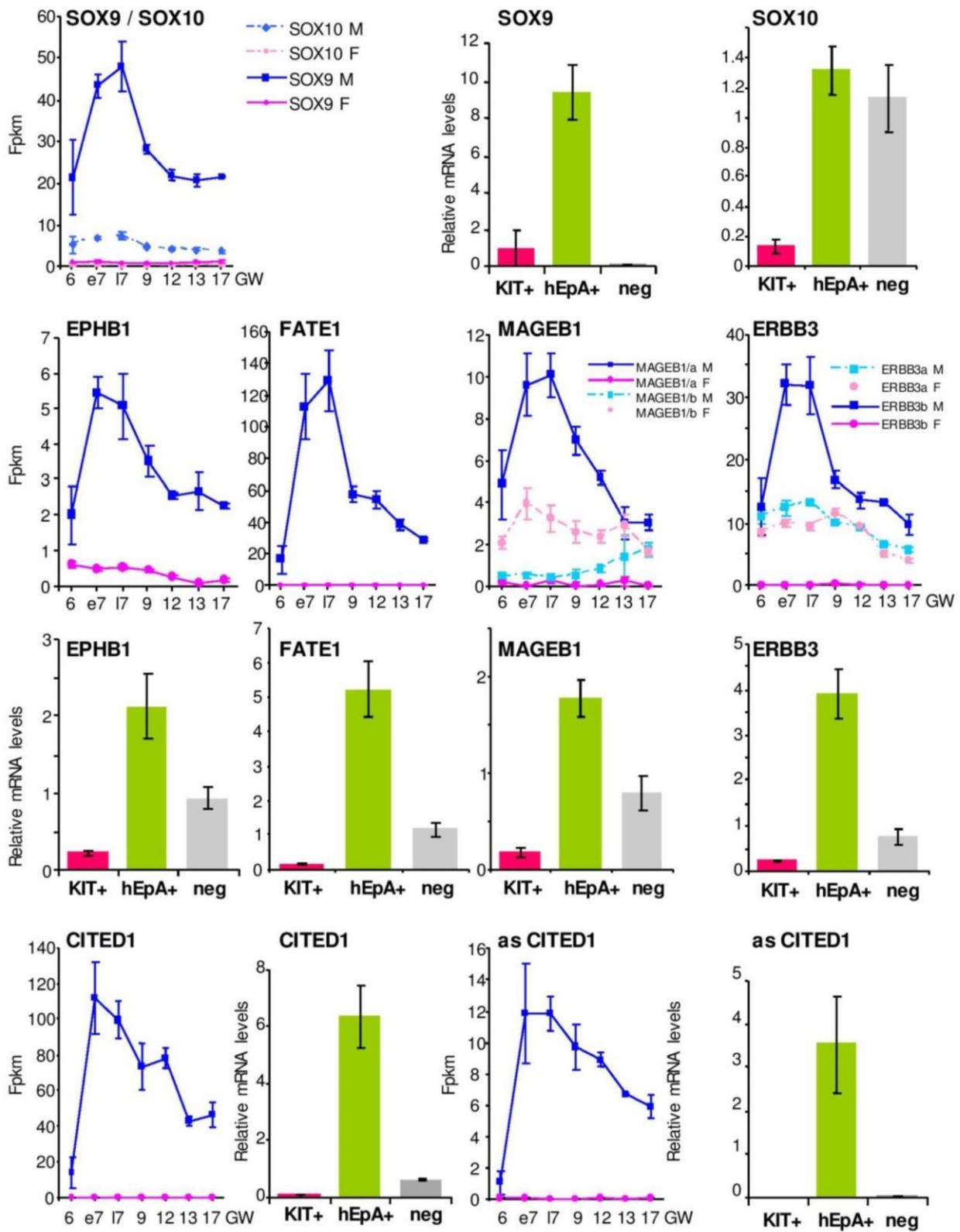


Figure 8. Cellular investigation of eSDTs from expression pattern P2.

Figure 9. Cellular investigation of eSDTs from expression pattern P8.

Expression levels (A, C) (line graphic) and quantitative RT-PCR (B, D) (histograms) of Wilms tumor 1 (*WT1*), prostaglandin-endoperoxide synthase 1 (*PTGS1*), neuropeptide Y (*NPY*), neurexin 3 (*NRXN3*), contactin 1 (*CNTN1*), SRY-box 4 (*SOX4*), SET nuclear proto-oncogene (*SET*) and NUTs *TCONS_00153406* and *TCONS_00113718*. Expression levels from RNA-seq as a function of age are depicted as blue lines for the testis and pink lines for the ovaries. Each point represents the mean FPKM \pm SEM of the levels measured in 4 (14 GW and younger) and 2 different gonads (15 and 19 GW). Quantitative RT-PCR was performed on the ovarian sorted cell populations of germ cells (KIT+, red bars) and other cells (KIT-, grey bars). Each column shows a pool of sorted cells from 7 early differentiating (6.7-8.7 GW, 7-9 GW) and 3 fetal (10.6-11.7 GW, 10-12 GW) ovaries. Each bar represents the mean \pm SEM of the fold change in target gene expression relative to the reference genes *RPLP0* and *RPS20*. FPKM = fragments per kilobase of exon model per million reads mapped. (E-G) Immunofluorescence for (E) WT1 (green) and KIT (red), (F) NRXN3 (green) and KIT (red), and (G) LIN28 (green) and CNTN1 (red) in early differentiating ovaries (6-6.6 GW). Scale bar: 100 μ M.

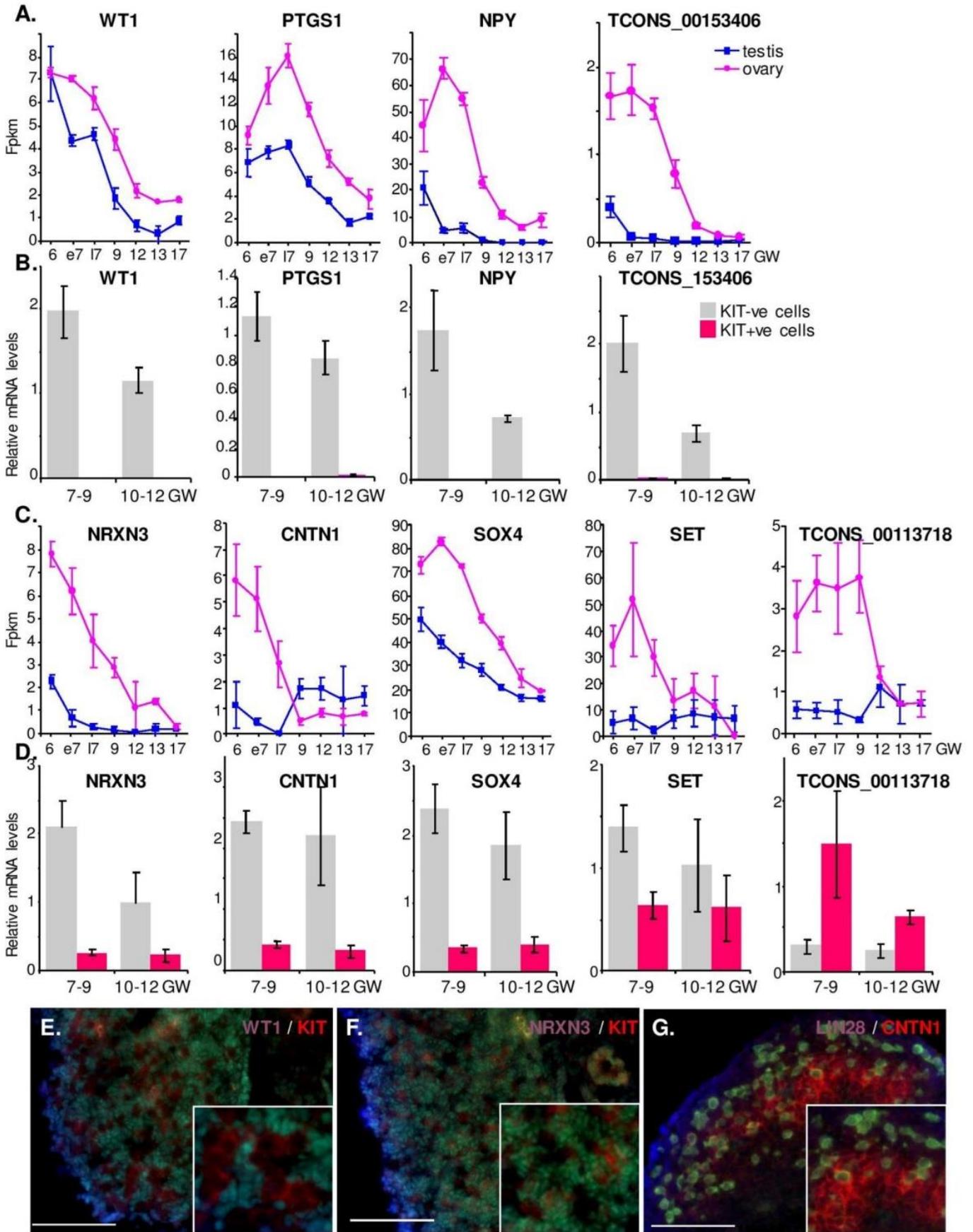


Figure 9. Cellular investigation of eSDTs from expression pattern P8.

Figure 10. Cellular investigation of eSDTs from expression pattern P9.

Expression levels (A, C) (line graphic) and quantitative RT-PCR (B, D) (histograms) of KIT proto-oncogene receptor tyrosine kinase (*KIT*), Nanog homeobox (*NANOG*), POU class 5 homeobox 1 (*POU5F1*), lin-28 homolog A (*LIN28A*), R-spondin 1 (*RSPO1*) and NUTs *TCONS_00055038*, *TCONS_00042656*, and *TCONS_00224470*. Expression levels from RNA-seq are expressed as a function of age are depicted as blue lines for the testis and pink lines for the ovaries. Each point represents the mean FPKM \pm SEM of the levels measured in 4 (14 GW and younger) and 2 different gonads (15 and 19 GW). Quantitative RT-PCR was performed on the ovarian sorted cell populations of germ cells (KIT+, red bars) and other cells (KIT-, grey bars). Each column shows a pool of sorted cells from 7 early differentiating (6.7-8.7 DW, 7-9 GW) and 3 fetal (10.6-11.7 GW, 10-12 GW) ovaries. Each bar represents the mean \pm SEM of the fold change in target gene expression relative to the reference genes *RPLP0* and *RPS20*. FPKM = fragments per kilobase of exon model per million reads mapped.

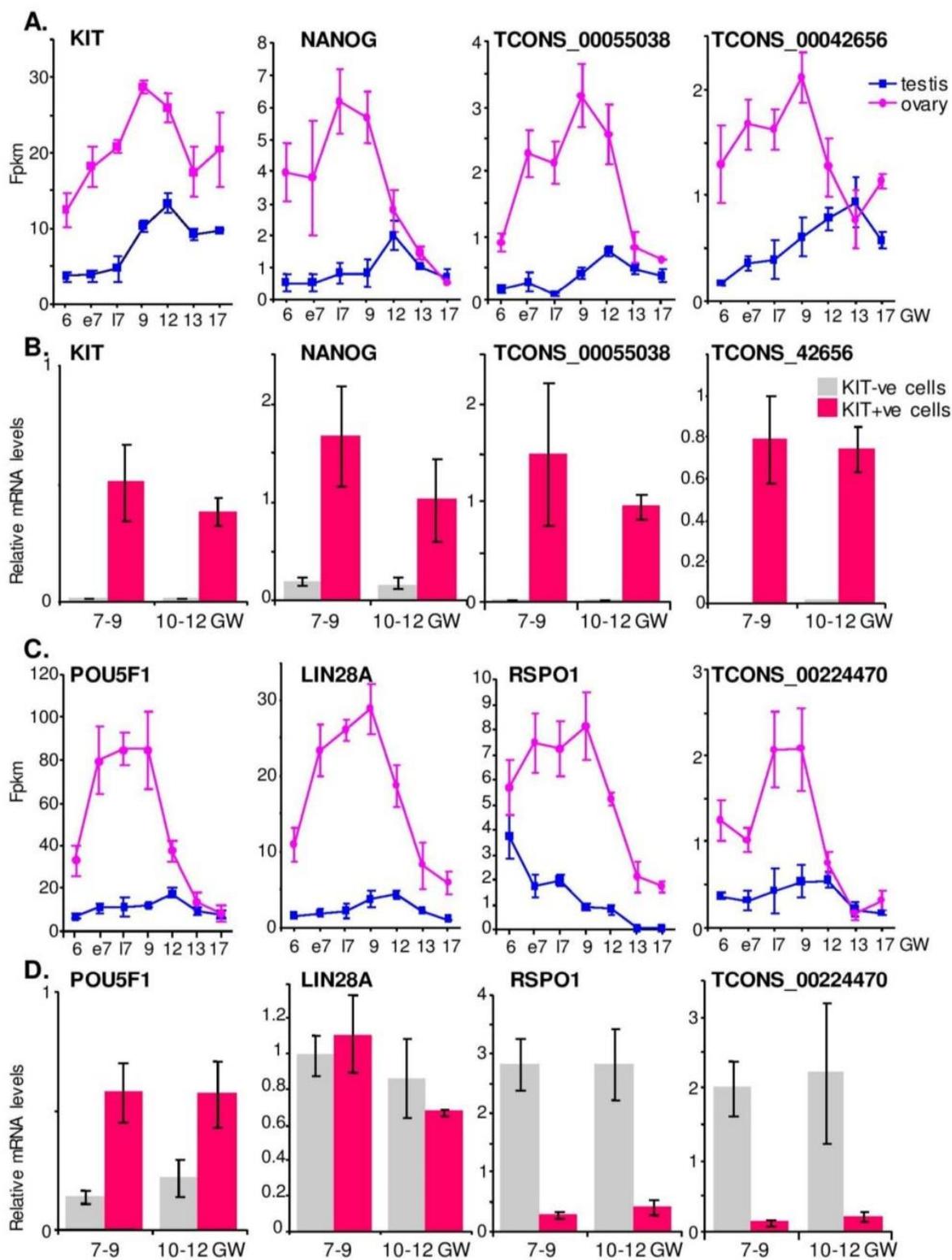


Figure 10. Cellular investigation of eSDTs from expression pattern P9.

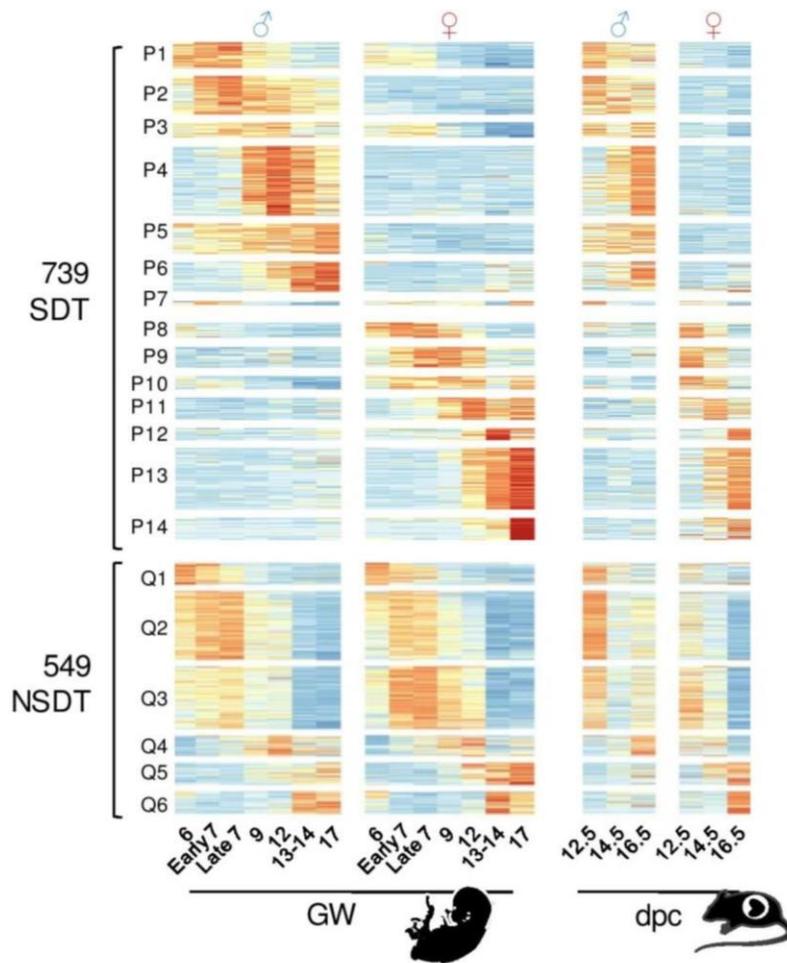


Figure 11. Expression conservation of syntenic transcripts during gonad development in human and mouse.

Heatmap representation of 739 sexually-dimorphic transcripts (SDTs, P1-P14) and 549 non sexually-dimorphic transcripts (NSDTs, Q1-Q6) across seven developmental stages for human testes and ovaries and three developmental stages for mouse testes and ovaries. Each row is a transcript, and each column is an experimental condition, i.e. the average of testes or ovaries from a given a gestational week (GW) or day post-coitum (dpc). The standardized abundance of transcripts is color-coded according to the scale bar, red corresponding to the highest expression level, blue to the lowest.

Heatmap of the conserved SDT and NSDT transcripts. Each row corresponds to a transcript, each column to a condition, testis or ovary at a time of development (Gestational weeks). Color scale represents the standardized abundance of transcripts within each organism, redder meaning the highest expression level, bluer the lower. Enrichment analysis of the unmapped transcripts.

Supplemental figures

	GW	6	e7	17	9	12	13-14	17	Total
Testis	Nb samples	4	4	4	4	4	2	2	24
	Average sequenced pairs of read	50 163 366 +- 5 540 241	43 637 558 +- 3 217 370	45 033 799 +- 4 891 256	44578 922 +- 6124741	61 408 394 +- 31 392 665	46 992435 +-3 747 882	47 295 757 +- 2 131 643	339 110 230
	Average mapped pairs of read	45 088 608 +- 4 567 496	39 066 943 +-2 877 937	38 915 843 +- 6 718 006	36 500 690 +- 3 675 842	46 586 645 +- 20 642 653	40 120 429 +- 2 994 901	38 721 188 +- 4 093 627	285 000 344
	% of mapped reads	89.9	89.5	86.4	81.9	75.9	85.4	81.9	84.4
	Nb reconstructed transcripts	164,931	129,802	103,431	116,903	107,148	164,652	147,287	142,953
Ovary	Nb samples	4	4	4	4	4	2	2	24
	Average sequenced pairs of read	51 756 993 +- 3 476 980	42 038 738 +- 3 950 805	43 167 327 +- 6 221 180	45 461 320 +- 8 410 709	43 085 238 +- 6 565 924	47 163 616 +- 3 195 188	38 044 232 +- 2 869 127	310 717 463
	Average mapped pairs of read	46 047 326 +- 3 200 842	38 087 267 +-3 581 721	33 349 134 +- 3 167 555	37 688 289 +- 10 017 251	34 516 893 +- 5 232 477	38 186 847 +- 1 789 162	33 236 072 +- 2 927 283	261111827
	% of mapped reads	89.0	90.6	77.3	82.9	80.1	81.0	87.4	84.0
	Nb reconstructed transcripts	186,499	121,800	93,201	129,817	112,165	191,983	174,783	168,195

180,242 reconstructed transcripts ←

Figure S1. Statistics for read mapping and transcript assembly

Average number of raw and mapped reads, percentage of mapping and number of assembled transcripts are indicated for each experimental condition. GW = Gestational week. The number of non-redundant assembled transcripts is also provided during testicular and ovarian developments, as well as the overall number of non-redundant assembled transcripts in this study.

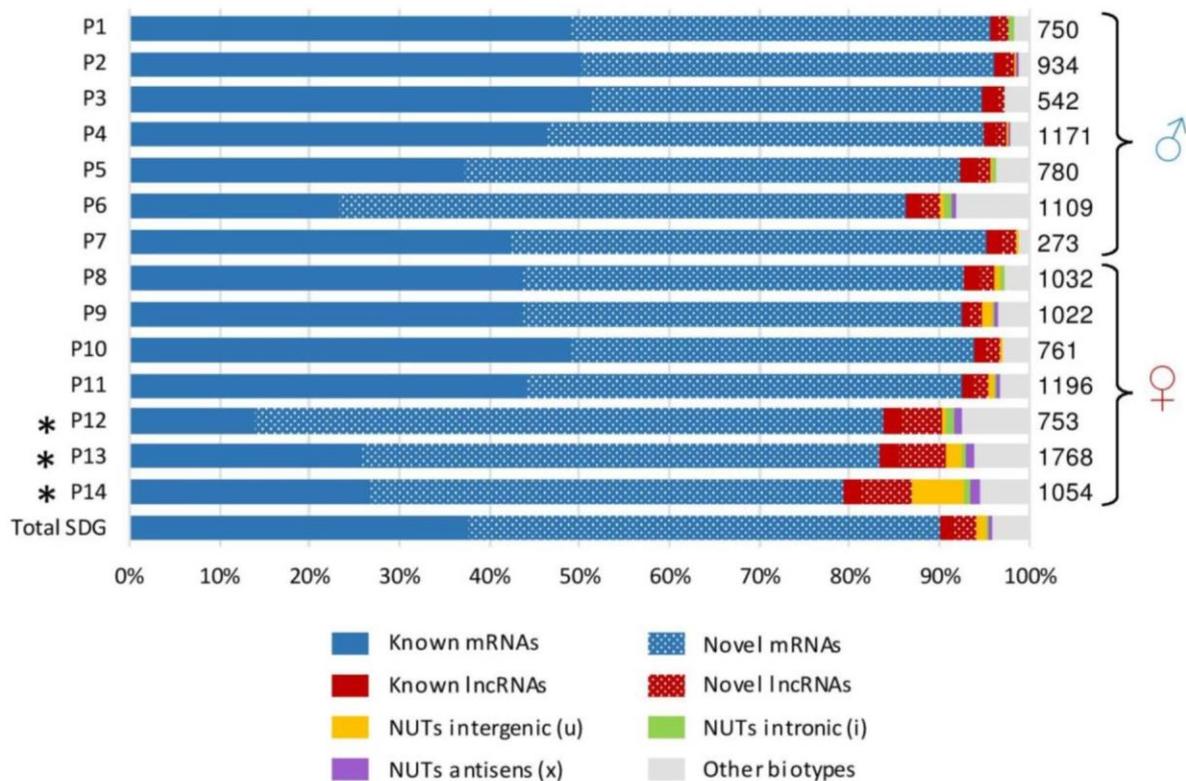


Figure S2. Transcripts biotypes and isoforms status proportion in all SDT and within each clusters.

The comparison of the 13'145 SDT with the human reference transcriptome by Cuffcompare (REF) classified them as known isoform (class code “=”), novel isoforms (class code “j”), novel unannotated transcripts (NUTs) in intronic regions (class code “i”), intergenic regions (“u”), antisense of known transcripts (class code “x”) or other ambiguous biotypes. Proportion of mRNAs, lncRNAs and NUTs in NSDT clusters is given. Total number of transcripts in each clusters is indicated on the right side of the plot. An enrichment analysis using hypergeometric strategy highlighted a significant accumulation of lncRNAs and NUT in P12, P13 and P14 clusters (p-value <0.05) compared to their distribution within the 14 SDT clusters.

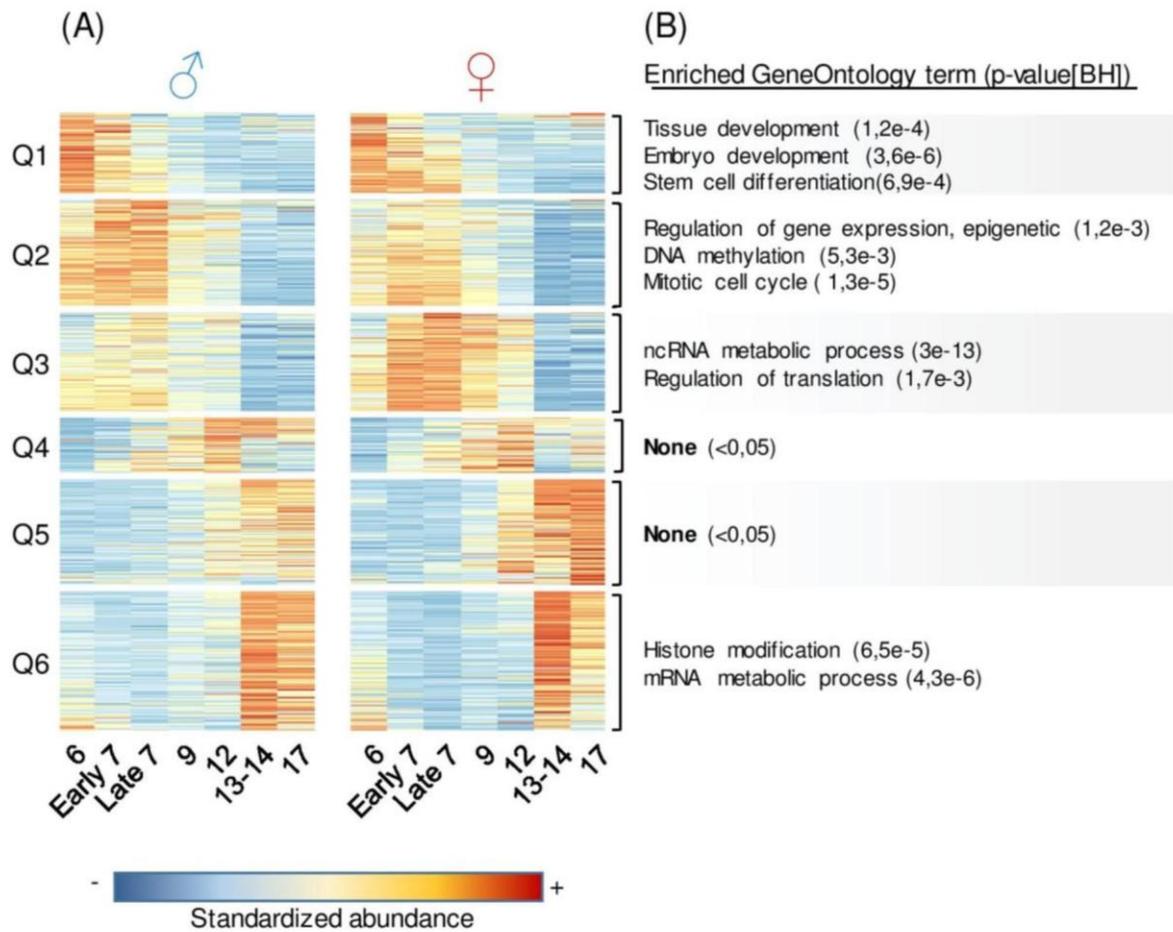


Figure S3. Non sexually-dimorphic expression patterns during human gonad development.

(A) Heatmap representation of 8'935 non sexually-dimorphic transcripts (NSDTs), distributed into 6 expression patterns (P1 to P6), across seven developmental stages for both testes and ovaries. Each row is a transcript, and each column is an experimental condition, *i.e.* the average of testes or ovaries from a given a gestational week (GW). The standardized abundance of transcripts is color-coded according to the scale bar, red corresponding to the highest expression level, blue to the lowest. **(B)** Gene Ontology terms found to be enriched (BH corrected p-value <0.05) in each expression pattern.

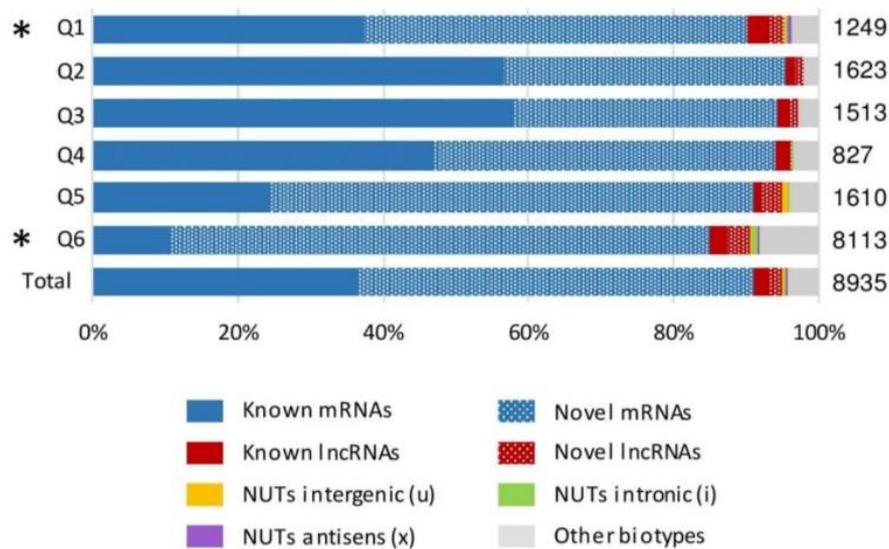


Figure S4. Transcripts biotypes and isoforms status proportion in all NSDT and within each clusters.

The comparison of the 8'935 NSDT with the human reference transcriptome by Cuffcompare (REF) classified them as known isoform (class code “=”), novel isoforms (class code “j”), novel unannotated transcripts (NUTs) in intronic regions (class code “i”), intergenic regions (“u”), antisens of known transcripts (class code “x”) or other ambiguous biotypes. Proportion of mRNAs, lncRNAs and NUTs in NSDT clusters is given. Total number of transcripts in each clusters is indicated on the right side of the plot. An enrichment analysis using hypergeometric strategy highlighted a significant accumulation of lncRNAs and NUT in Q1 and Q6 clusters (p-value <0.05) compared to their distribution within the 6 NSDT clusters.

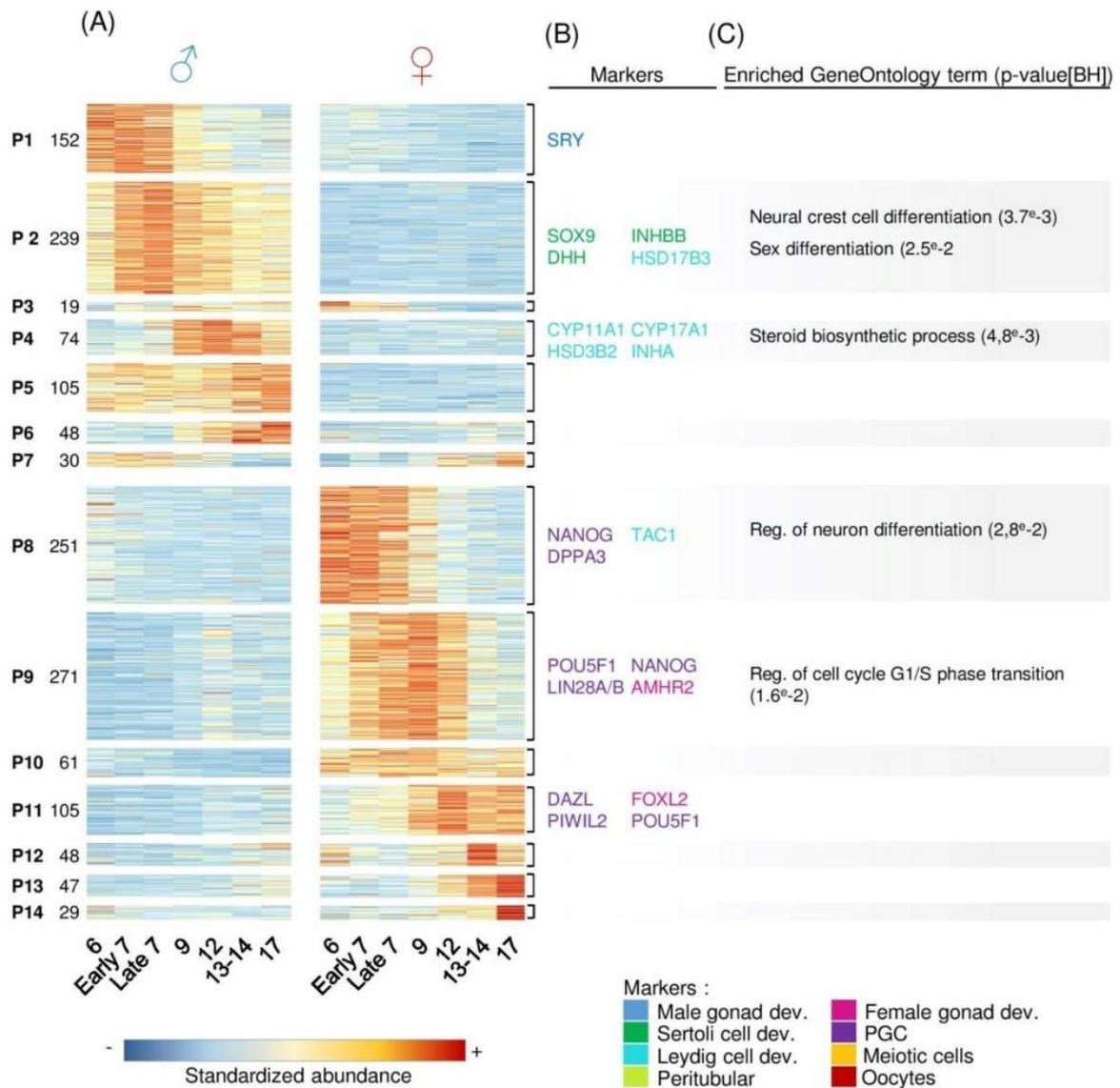


Figure S5. Early sexually-dimorphic expression patterns during human gonad development.

(A) Heatmap representation of 13'145 early sexually-dimorphic transcripts (eSDTs), distributed into 14 expression patterns (P1 to P14), across seven developmental stages for both testes and ovaries. Each row is a transcript, and each column is an experimental condition, i.e. the average of testes or ovaries from a given a gestational week (GW). The standardized abundance of transcripts is color-coded according to the scale bar, red corresponding to the highest expression level, blue to the lowest. **(B)** Repartition of known markers involved in gonad differentiation and development within eSDT expression patterns. Note that several isoforms of a given can be assembled and display distinct expression. **(C)** Gene Ontology terms found to be enriched (BH corrected p-value <0.05) in each expression pattern.

Figure S6. Testicular and ovarian cell-sorting by flow cytometry.

Testicular and ovarian cell sorting by flow cytometry. **(A)** Representative immunofluorescence of hEpA-FITC (green) staining of cord cells and KIT-PE (red) staining of germ cells in testis sections of a 7 DW old embryo. **(B)** Representative Side (SSC) vs Forward (FSC) scatter plot showing hEpA/KIT dot plots according to size (FSC-H) and cellular granularity (SSC-H). Example of gating strategy for flow cytometry sorting of Sertoli cells (hEpA+/KIT-, green dots), germ cells (hEpA-/KIT+, red plots) and other cells types (hEpA-/KIT-, grey plots). **(C)** Quantitative RT-PCR of *KIT*, *NR2F2* and *SOX9* was performed on the different sorted cell populations of germ cells (KIT+, red bars) Sertoli cells (hEpA+, green bars) and other cells (KIT-/ hEpA-, grey bars). Each column shows a pool of sorted cells from 5 fetal (6.9-7.3 GW) testes. Each bar represents the mean \pm SEM of the fold change in target gene expression relative to the reference genes *RPLPO* and *RPS20*. **(D)** Representative immunofluorescence of KIT-PE (red) staining of germ cells in sections of an ovary at 11 GW. **(E)** Representative Side (SSC) vs Forward (FSC) scatter plot showing KIT dot plots according to size (FSC-H) and cellular granularity (SSC-H). Example of gating strategy for flow cytometry sorting of germ cells (KIT+, red plots) and other cells types (KIT-, grey plots). **(C)** Quantitative RT-PCR of *KIT*, *NR2F2* and *FOXL2* was performed on the different sorted cell populations of germ cells (KIT+, red bars) and other cells (KIT-, grey bars). Each column shows a pool of sorted cells from 7 early differentiating (6.7-8.7 GW, 7-9 GW) and 3 fetal (10.6-11.7 GW, 10-12 GW) ovaries. Each bar represents the mean \pm SEM of the fold change in target gene expression relative to the reference genes *RPLPO* and *RPS20*. Scale bars: 100 μ M.

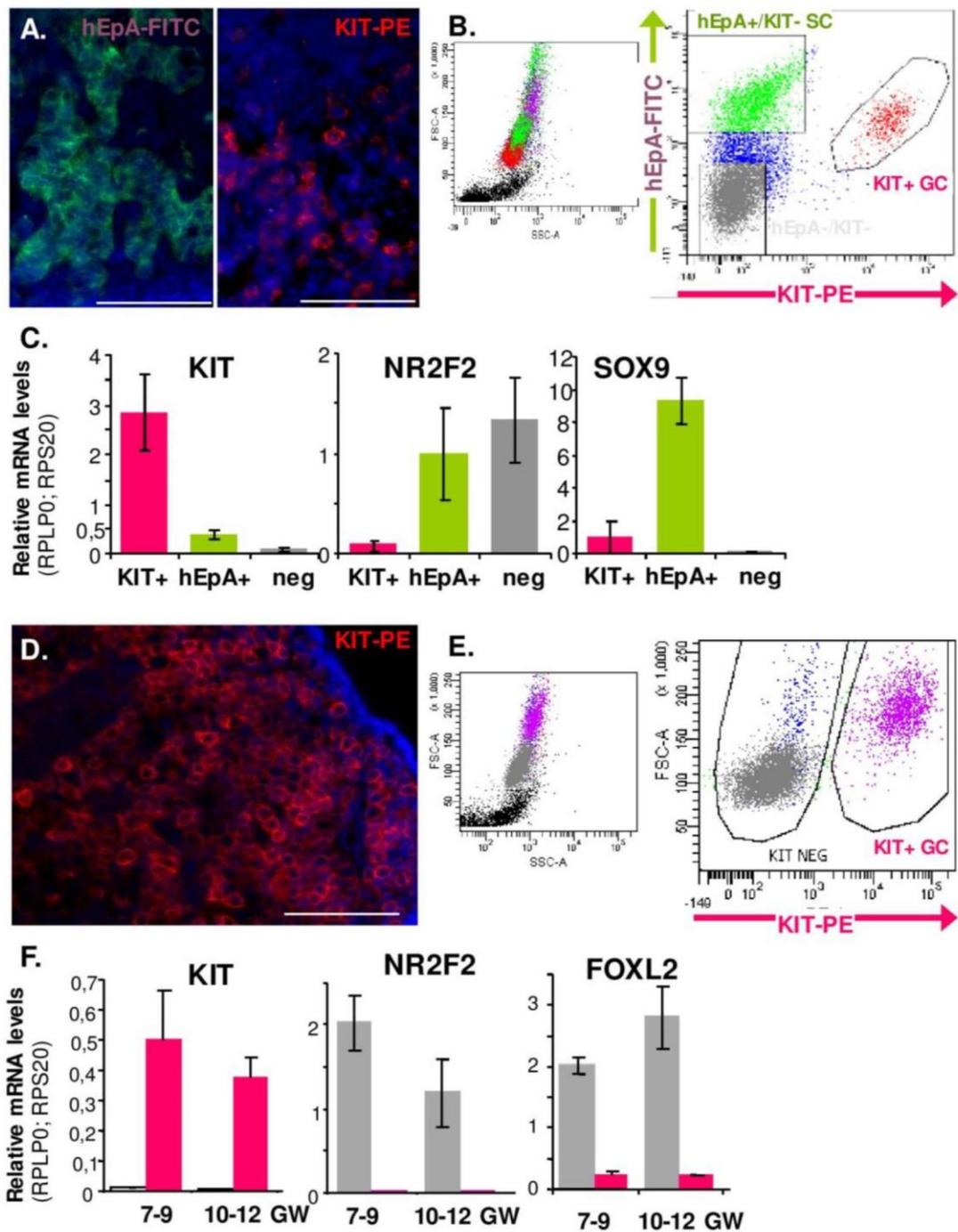


Figure S6. Testicular and ovarian cell-sorting by flow cytometry.

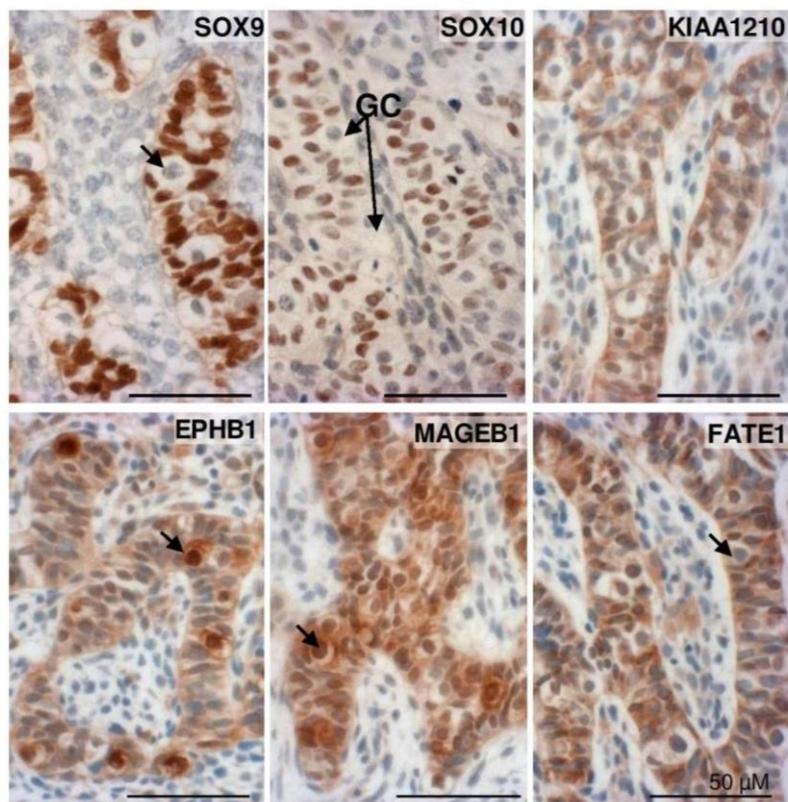


Figure S7. *In situ* analysis of selected candidates at the protein level.

Representative immunohistochemistry of SOX9, SOX10, KIAA1210, EPHB1, MAGEB1 and FATE1 on a 7.1 GW testis. Arrows indicate germ cells (GC). Scale bar: 50 μM.

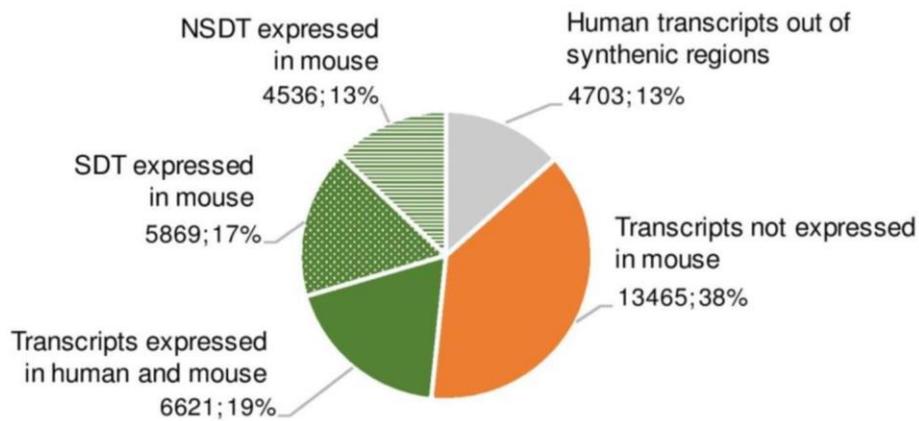


Figure S8. Distribution of refined human transcripts with regard to syntenic conservation in the mouse

Among the 35,194 transcripts expressed during human gonad development, 4703 did not belong to syntenic regions (grey). Among the 17,026 significantly expressed during human and mouse gonad development (green), 5869 are defined as SDT, and 4536 as NSDT. 13,465 transcripts were found to be expressed only during human gonad development (orange).

Supplemental Data Legends

Supplemental data 1 is a searchable table (.xls) containing information on genomic features and expression data for all refined transcripts.

References

- Bagheri-Fam, S., Bird, A.D., Zhao, L., Ryan, J.M., Yong, M., Wilhelm, D., Koopman, P., Eswarakumar, V.P., and Harley, V.R. (2017). Testis Determination Requires a Specific FGFR2 Isoform to Repress FOXL2. *Endocrinology* 158, 3832–3843.
- Beverdam, A., and Koopman, P. (2006). Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. *Hum. Mol. Genet.* 15, 417–431.
- Bouma, G.J., Affourtit, J., Bult, C.J., and Eicher, E.M. (2007). Transcriptional profile of mouse pre-granulosa and Sertoli cells isolated from early-differentiated fetal gonads.
- Bouma, G.J., Hudson, Q.J., Washburn, L.L., and Eicher, E.M. (2010). New Candidate Genes Identified for Controlling Mouse Gonadal Sex Determination and the Early Stages of Granulosa and Sertoli Cell Differentiation1. *Biol. Reprod.* 82, 380–389.
- Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927.
- Chalmel, F., and Primig, M. (2008). The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology. *BMC Bioinformatics* 9, 86.
- Chalmel, F., Lardenois, a., Evrard, B., Rolland, a. D., Sallou, O., Dumargne, M.-C., Coiffec, I., Collin, O., Primig, M., and Jegou, B. (2014). High-Resolution Profiling of Novel Transcribed Regions During Rat Spermatogenesis. *Biol. Reprod.* 91, 5–5.
- Chan, E.T., Quon, G.T., Chua, G., Babak, T., Trochesset, M., Zirngibl, R.A., Aubin, J., Ratcliffe, M.J., Wilde, A., Brudno, M., et al. (2009). Conservation of core gene expression in vertebrate tissues. *J. Biol.* 8, 33.
- Chang, H., Gao, F., Guillou, F., Taketo, M.M., Huff, V., and Behringer, R.R. (2008). Wt1 negatively regulates beta-catenin signaling during testis development. *Dev. Camb. Engl.* 135, 1875–1885.
- Chassot, A.A., Gregoire, E.P., Magliano, M., Lavery, R., and Chaboissier, M.C. (2008). Genetics of Ovarian Differentiation: Rspo1, a Major Player. *Sex. Dev.* 2, 219–227.
- Darde, T.A., Sallou, O., Becker, E., Evrard, B., Monjeaud, C., Le Bras, Y., Jégou, B., Collin, O., Rolland, A.D., and Chalmel, F. (2015). The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Res.* 43, W109-16.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789.
- Eggers, S., Sadedin, S., van den Bergen, J.A., Robevska, G., Ohnesorg, T., Hewitt, J., Lambeth, L., Bouty, A., Knarston, I.M., Tan, T.Y., et al. (2016). Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort. *Genome Biol.* 17, 243.

- Greenfield, A. (2015). Understanding sex determination in the mouse: genetics, epigenetics and the story of mutual antagonisms. *J. Genet.* *94*, 585–590.
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* *161*, 1437–1452.
- Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Goriely, A., et al. (2017). Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling and Metabolic Transitions during Human Spermatogonial Stem Cell Development. *Cell Stem Cell* *21*, 533–546.e6.
- Haque, A., Engel, J., Teichmann, S.A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* *9*, 75.
- Houmard, B., Small, C., Yang, L., Nalwai-Cecchini, T., Cheng, E., Hassold, T., and Griswold, M. (2009). Global Gene Expression in the Human Fetal Testis and Ovary. *Biol Reprod* *443*, biolreprod.108.075747.
- Inoue, M., Shima, Y., Miyabayashi, K., Tokunaga, K., Sato, T., Baba, T., Ohkawa, Y., Akiyama, H., Suyama, M., and Morohashi, K. (2016). Isolation and Characterization of Fetal Leydig Progenitor Cells of Male Mice. *Endocrinology* *157*, 1222–1233.
- Jameson, S.A., Lin, Y.-T., and Capel, B. (2012). Testis development requires the repression of *Wnt4* by *Fgf* signaling. *Dev. Biol.* *370*, 24–32.
- Kashimada, K., Pelosi, E., Chen, H., Schlessinger, D., Wilhelm, D., and Koopman, P. (2011). *FOXL2* and *BMP2* act cooperatively to regulate *follistatin* gene expression during ovarian development. *Endocrinology* *152*, 272–280.
- Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. *Nature* *509*, 575–581.
- Kim, Y., Kobayashi, A., Sekido, R., DiNapoli, L., Brennan, J., Chaboissier, M.-C., Poulat, F., Behringer, R.R., Lovell-Badge, R., and Capel, B. (2006). *Fgf9* and *Wnt4* Act as Antagonistic Signals to Regulate Mammalian Sex Determination. *PLoS Biol.* *4*, e187.
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. *Nature* *351*, 117–121.
- Laiho, A., Kotaja, N., Gyenesei, A., and Sironen, A. (2013). Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *PLoS One* *8*, e61558.
- Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X., et al. (2017). Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* *20*, 891–892.
- Li, Y., Zheng, M., and Lau, Y.-F.C. (2014). The Sex-Determining Factors *SRY* and *SOX9* Regulate Similar Target Genes and Promote Testis Cord Formation during Testicular Differentiation.
- Lin, S., Lin, Y., Nery, J.R., Urich, M. a., Breschi, A., Davis, C. a., Dobin, A., Zaleski, C., Beer, M. a., Chapman, W.C., et al. (2014). Comparison of the transcriptional landscapes between human and mouse tissues. *Proc. Natl. Acad. Sci.* *111*, 201413624.

- Liu, C.-F., Bingham, N., Parker, K., and Yao, H.H.-C. (2009). Sex-specific roles of β -catenin in mouse gonadal development. *Hum. Mol. Genet.* *18*, 405–417.
- Maatouk, D.M., DiNapoli, L., Alvers, A., Parker, K.L., Taketo, M.M., and Capel, B. (2008). Stabilization of β -catenin in XY gonads causes male-to-female sex-reversal. *Hum. Mol. Genet.* *17*, 2949–2955.
- Mamsen, L.S., Ernst, E.H., Borup, R., Larsen, A., Olesen, R.H., Ernst, E., Anderson, R.A., Kristensen, S.G., and Andersen, C.Y. (2017). Temporal expression pattern of genes during the period of sex differentiation in human embryonic gonads. *Sci. Rep.* *7*, 15961.
- McClelland, K.S., Bell, K., Larney, C., Harley, V.R., Sinclair, A.H., Oshlack, A., Koopman, P., and Bowles, J. (2015). Purification and Transcriptomic Analysis of Mouse Fetal Leydig Cells Reveals Candidate Genes for Specification of Gonadal Steroidogenic Cells¹. *Biol. Reprod.* *92*, 1–12.
- Munger, S.C., Natarajan, A., Looger, L.L., Ohler, U., and Capel, B. (2013). Fine Time Course Expression Analysis Identifies Cascades of Activation and Repression and Maps a Putative Regulator of Mammalian Sex Determination. *PLoS Genet.* *9*, e1003630.
- Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grützner, F., and Kaessmann, H. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* *505*, 635–640.
- Nef, S., Schaad, O., Stallings, N.R., Cederroth, C.R., Pitetti, J.-L., Schaer, G., Malki, S., Dubois-Dauphin, M., Boizet-Bonhoure, B., Descombes, P., et al. (2005). Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. *Dev. Biol.* *287*, 361–377.
- Ottolenghi, C., Pelosi, E., Tran, J., Colombino, M., Douglass, E., Nedorezov, T., Cao, A., Forabosco, A., and Schlessinger, D. (2007). Loss of Wnt4 and Foxl2 leads to female-to-male sex reversal extending to germ cells. *Hum. Mol. Genet.* *16*, 2795–2804.
- Pollier, J., Rombauts, S., and Goossens, A. (2013). Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol. Biol. Clifton NJ* *1011*, 305–315.
- Rahmoun, M., Lavery, R., Laurent-Chaballier, S., Bellora, N., Philip, G.K., Rossitto, M., Symon, A., Pailhoux, E., Cammas, F., Chung, J., et al. (2017). In mammalian foetal testes, SOX9 regulates expression of its target genes by binding to genomic regions with conserved signatures. *Nucleic Acids Res.* *45*, 7191–7211.
- Rolland, A.D., Lehmann, K.P., Johnson, K.J., Gaido, K.W., and Koopman, P. (2011). Uncovering gene regulatory networks during mouse fetal germ cell development. *Biol. Reprod.* *84*, 790–800.
- Schmidt, D. (2004). The murine winged-helix transcription factor Foxl2 is required for granulosa cell differentiation and ovary maintenance. *Development* *131*, 933–942.
- Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* *453*, 930–934.
- Small, C.L., Shima, J.E., Uzumcu, M., Skinner, M.K., and Griswold, M.D. (2005). Profiling Gene Expression During the Differentiation and Development of the Murine Embryonic Gonad. *72*, 492–501.

- Soh, Y.Q.S., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet.* *11*, e1005531.
- Stévant, I., Neirijnck, Y., Borel, C., Escoffier, J., Smith, L.B., Antonarakis, S.E., Dermitzakis, E.T., and Nef, S. (2018). Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing. *Cell Rep.* *22*, 1589–1599.
- 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., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci.* *99*, 4465–4470.
- Sudmant, P.H., Alexis, M.S., and Burge, C.B. (2015). Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biol.* *16*, 287.
- Uda, M., Ottolenghi, C., Crisponi, L., Garcia, J.E., Deiana, M., Kimber, W., Forabosco, A., Cao, A., Schlessinger, D., and Pilia, G. (2004). *Foxl2* disruption causes mouse ovarian failure by pervasive blockage of follicle development. *Hum. Mol. Genet.* *13*, 1171–1181.
- Vainio, S., Heikkilä, M., Kispert, A., Chin, N., and McMahon, A.P. (1999). Female development in mammals is regulated by Wnt-4 signalling. *Nature* *397*, 405–409.
- del Valle, I., Buonocore, F., Duncan, A.J., Lin, L., Barenco, M., Parnaik, R., Shah, S., Hubank, M., Gerrelli, D., and Achermann, J.C. (2017). A genomic atlas of human adrenal and gonad development. *Wellcome Open Res.* *2*, 25.
- Vidal, V.P.I., Chaboissier, M.-C., de Rooij, D.G., and Schedl, A. (2001). *Sox9* induces testis development in XX transgenic mice. *Nat. Genet.* *28*, 216–217.
- Wang, C., Gong, B., Bushel, P.R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., Fang, H., Hong, H., Shen, J., Su, Z., et al. (2014). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* *32*, 926–932.
- Wilhelm, D., Washburn, L.L., Truong, V., Fellous, M., Eicher, E.M., and Koopman, P. (2009). Antagonism of the testis- and ovary-determining pathways during ovotestis development in mice. *Mech. Dev.* *126*, 324–336.
- Zhao, L., Wang, C., Lehman, M.L., He, M., An, J., Svingen, T., Spiller, C.M., Ng, E.T., Nelson, C.C., and Koopman, P. (2018). Transcriptomic analysis of mRNA expression and alternative splicing during mouse sex determination. *Mol. Cell. Endocrinol.*
- Zimmermann, C., Stévant, I., Borel, C., Conne, B., Pitetti, J.-L., Calvel, P., Kaessmann, H., Jégou, B., Chalmel, F., and Nef, S. (2015). Research resource: the dynamic transcriptional profile of sertoli cells during the progression of spermatogenesis. *Mol. Endocrinol. Baltim. Md* *29*, 627–642.

Materials and Methods

SI Materials and Methods

Ethical considerations and sample collection

5 *First trimester fetuses.* Human fetuses (6-12 GW) were obtained from legally-induced normally-progressing terminations of pregnancy performed in Rennes University Hospital. Tissues were collected with women's written consent, in accordance with the legal procedure agreed by the National agency for biomedical research (authorization #PFS09-011; Agence de la Biomédecine) and the approval of the Local ethics committee of Rennes Hospital (advice # 11-48). The termination of pregnancy was induced using a
10 standard combined mifepristone and misoprostol protocol, followed by aspiration. Gestational age was determined by ultrasound, and further confirmed by measurement of foot length. The gonads were recovered and dissected free of mesonephros in ice-cold phosphate-buffered saline (PBS) using a binocular microscope (Olympus SZX7, Lille, France). The sex of the gonad was determined by morphological criteria, except for fetuses younger than 7 GW, for which a PCR was performed on genomic DNA using primers
15 specific for SRY (ACAGTAAAGGCAACGTCCAG; ATCTGCGGGAAGCAAAGTGC)(Friel et al., 2002) as well as for AMELX and AMELY (CTGATGGTTGGCCTCAAGCCTGTG; GTGATGGTTGGCCTCAAGCCTGTG) (Akane et al., 1992).

Second trimester fetuses. Human fetuses (13-14 and 17 GW) were obtained from pregnant women after legally induced abortions at the Aberdeen Pregnancy Counselling Service. The collection of fetal material
20 was approved by the National Health Service (NHS) Grampian Research Ethics Committees (REC 04/S0802/21). In all cases, women seeking elective terminations of pregnancy were recruited with full written, informed consent by nurses working independently of the study at the Aberdeen Pregnancy Counselling Service. Maternal data, medications used and self-reported number of cigarettes smoked per

day were recorded. Only fetuses from normally progressing pregnancies (determined at ultrasound scan
25 prior to termination) from women over 16 years of age were collected following termination by RU486
(mifepristone) treatment (200 mg) and prostaglandin-induced delivery, as detailed previously
(O'Shaughnessy et al., 2007). Fetuses were transported to the laboratory within 30 min of delivery,
weighed, sexed and the crown-rump length recorded. Fetal tissues were snap-frozen in liquid nitrogen and
stored at -80°C. Maternal smoking status was confirmed by measurement of fetal plasma cotinine using a
30 commercially available kit (Cozart Plc, Abingdon, Kent, UK).

RNA extraction, library construction and RNA-Sequencing

Total RNA were extracted from human fetal gonads using the RNeasy mini Kit (Qiagen), quantified using a
NanoDrop™ 8000 spectrophotometer (Thermo Scientific) and quality controlled using a 2100
35 Electrophoresis Bioanalyzer (Agilent). Libraries suitable for strand specific high throughput DNA
sequencing were then constructed essentially as previously described (Jégou et al., 2017), i.e. using
“TruSeq Stranded Total RNA with Ribo-Zero Gold Prep Kit” (catalog # RS-122-2301, Illumina Inc.). First, the
removal of rRNA was performed from 500 ng of total RNA using biotinylated, target-specific oligos
combined with Ribo-Zero rRNA removal beads. Following purification, the RNA was fragmented using
40 divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand
cDNA using random primers and reverse transcriptase, followed by second strand cDNA synthesis using
DNA Polymerase I and RNase H. The double stranded cDNA fragments were blunted using T4 DNA
polymerase, Klenow DNA polymerase and T4 PNK. A single 'A' nucleotide was added to the 3' ends of the
blunt DNA fragments using a Klenow fragment (3' to 5' exo minus) enzyme. The cDNA fragments were
45 ligated to double stranded adapters using T4 DNA Ligase. The ligated products were enriched by PCR
amplification (30 sec at 98°C; [10 sec at 98°C, 30 sec at 60°C, 30 sec at 72°C] x 12 cycles; 5 min at 72°C).
Excess of PCR primers was removed by purification using AMPure XP beads (Agencourt Biosciences

Corporation). Final cDNA libraries were quality-checked and quantified using a 2100 Electrophoresis Bioanalyzer (Agilent). The libraries were finally loaded in the flow cell at 7pM concentration and clusters
50 were generated in the Cbot and sequenced in the Illumina HiSeq 2500 as paired-end 2x50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 1.17.20 and CASAVA 1.8.2.

Read mapping, transcript assembly and quantification with the Tuxedo suite

55 *Assembly of a unique set of human reference transcripts.* Ensembl (Yates et al., 2016) and RefSeq (Brown et al., 2015; Pruitt et al., 2014) transcript annotations (GTF format) of human genome (release hg19) were downloaded from the UCSC (Speir et al., 2016) in June 2015, and were merged using Cuffcompare (Pollier et al., 2013) This non-redundant annotation was then used as human reference transcripts (HRT) as previously published (Chalmel et al., 2014). A non-redundant dataset of human splice junctions (HSJ) was
60 also extracted from alignments of human transcripts and expressed sequence tags (ESTs) on the human genome as provided by UCSC.

Read mapping. Reads from individual samples were aligned to the hg19 release of the human genome with TopHat2 (version 2.0.12) (Trapnell et al., 2012) using previously published approaches (Chalmel et al.,
65 2014; Pauli et al., 2012; Trapnell et al., 2012; Zimmermann et al., 2015). Briefly, TopHat2 program was first run for each fastq file, using HRT and HSJ datasets to improve read mapping. Exonic junction outputs were then merged and added to the HSJ set. TopHat2 was next run a second time using the new HSJ dataset to produce a final alignment file (BAM format) for each. Finally, BAM files corresponding to the same experimental condition (fetal testes or ovaries, at a given time-point) were merged and sorted using the
70 Samtools suite (version 2.19.0) (Li et al., 2009).

Transcriptome assembly and quantification. The Cufflinks suite (version 2.2.1, default settings) (Pollier et al., 2013) was used to assemble transcript fragments (or transfrags) for each experimental condition based on merged BAM files. Resulting assembled transcripts were further merged into a non-redundant set of transfrags which were further compared to the HRT dataset using the Cuffcompare program. Finally, Cuffquant was used to estimate abundance of each transcript in each individual sample as fragments per kilobase of exon model per million reads mapped (FPKM), and the normalization of the expression values across samples was performed by Cuffnorm.

Refinement of assembled transcripts. A four-step strategy was used to filter out dubious transcripts as previously described (Chalmel et al., 2014; Prensner et al., 2011) (Figure 2). First, based on the Cuffcompare comparison, we only selected transfrags defined as complete match ('='), potentially novel isoform ('j'), falling entirely within a reference intron ('i') or an intergenic region ('u'), or showing exonic overlap with reference on the opposite strand ('x'). Next, assembled transcripts of less than 200 nucleotides or that were undetectable (<1 FPKM) were discarded. Finally, all transfrags that were annotated as either novel isoforms or novel genes (class codes "j", "i", "u" or "x") and that did not harbor at least two exons (multi-exonic) were filtered out.

Principal Component Analyses. Principal component analysis was performed based on the expression values of refined transcripts using the FactoMineR package to graphically evaluate the distribution of sequenced samples (Lê et al., 2008).

90

Coding potential analysis of assembled transcripts

DNA sequences of the refined transcripts were extracted using TopHat's gffread tool. As recommended in (Chalmel et al., 2014; Zimmermann et al., 2015), the protein-coding potential of each transcript was

estimated by using an empirical integrative approach based on four predictive tools: CPAT (Wang et al.,
95 2013) (using the following cutoff: coding probability >0.364), HMMER (Finn et al., 2011) (E-value <10⁻⁴), CPC
(Kong et al., 2007) (class “coding”) and txCdsPredict (Kuhn et al., 2013) (score >800). Finally, we considered
assembled transcripts to likely encode for proteins if their nucleic sequences were considered as protein-
coding by at least two predictive tools.

100 **Proteomics Informed by Transcriptomics strategy**

As previously published in (Chocu et al., 2014), we made use of a Proteomics Informed by Transcriptomics
(PIT) approach (Evans et al., 2012) to provide evidence at the protein-level for assembled transcripts. First,
we assembled a customized non-redundant protein database by merging the UniProt (Pundir et al., 2015)
and the Ensembl (Yates et al., 2016) (downloaded 2015/10) proteome databases together with the set of
105 predicted proteins derived from the assembled transcripts. Briefly, the refined transcript sequences were
translated into the three-first open reading frames using the EMBOSS’s Transeq program (Rice et al., 2000)
and only the amino acid sequences of at least 20 residues were selected.

We next made use of the human fetal gonad MS/MS proteomics datasets available from the Human
Proteome Map (Kim et al., 2014). First, 131 raw data files (corresponding to three fetal ovary samples and
110 two fetal testis samples) were downloaded from the PRIDE database (accession number PXD000561)
(Vizcaíno et al., 2016) and converted into mgf format using ProteoWizard (Adusumilli and Mallick, 2017).
Subsequent analyses were performed using SearchGUI (Vaudel et al., 2011) (version 3.2.20) and
PeptideShaker (Vaudel et al., 2015) (version 1.16.8). A concatenated target/decoy database was created
from the enriched reference proteome using SearchGUI. Cross-peptide identification was then performed
115 using X!Tandem, OMSSA, and MSGF+ tools, with these following parameters: *i*) precursor ion tolerance
units set at 10 ppm; *ii*) fragment tolerance set at 0.05 Da, *iii*) carbamidomethylation of cysteine defined as

a fixed modification; *iv*) oxidation of methionine and acetylation of protein N-term defined as a variable modification; *v*) only tryptic peptides with up to two missed cleavages; and *vi*) minimum peptide length set to 6 amino acids. Using PeptideShaker, all peptides with at least one validated PSM and a confidence score upper than 80% were kept for further analyses. Finally, only identifications with a False Discovery Rate (FDR) <1% were selected.

Statistical transcript filtration

The set of sexually dimorphic transcripts (SDTs) was defined by filtering transfrags that exhibited a ≥ 2 -fold difference between the male and female gonads (using median expression values of sample replicates) in at least one of the seven developmental stages. A LIMMA statistical test was then used to identify transcripts displaying significant changes between male and female gonads (F-value adjusted with a FDR $\leq 5\%$) (Smyth, 2004). Among SDTs, those showing a differential expression as early as 6GW were designated as the set of early SDTs (or eSDTs). In addition, the set of developmental regulated transcripts (DRTs) were also selected by isolating transfrags that exhibited a ≥ 2 -fold difference between two developmental stages during either male or female gonad development. Similar to the selection of SDTs, a F-value adjusted with a FDR $\leq 5\%$ was then used to identify candidates showing significant variations across developmental stages. Finally, the difference between the sets of DRTs and SDTs allowed us to discriminate non-sexually dimorphic transcripts (NSDTs) corresponding to the set of transfrags showing a developmentally regulated expression pattern across fetal gonad development, but no significant differential expression between male and female gonads.

140 **Cluster and functional analyses**

The resulting SDTs and NSDTs were then clustered into 14 (named P1-14) and 6 (Q1-6) expression patterns using the unsupervised HCPC algorithm (Lê et al., 2008). These clusters were then ordered according to peak expression levels across developmental stages in testes first and then ovaries. Gene Ontology (GO) term enrichments were estimated with the Fisher exact probability, using a Gaussian hypergeometric test implemented in AMEN (Chalmel and Primig, 2008). A GO term was considered significantly associated with a given expression pattern if the FDR-corrected P value was $\leq 5\%$ and the number of genes bearing this annotation was ≥ 5 .

Conservation analysis

150 *Quantification of syntenic transcripts in the human genome*

In order to study the conservation of the expression program during fetal gonad development across species, a mouse RNA-seq dataset (Gene Expression Omnibus identifier: GSE70361) of fetal testes and ovaries at 12.5, 14.5 and 16.5 dpc (Soh et al., 2015) was analyzed anew using the same mapping protocol (see above) on the mouse genome (mm10 release).

155 Genomic coordinates of the refined human transcripts were next converted to syntenic mouse regions using the Crossmap tool (Zhao et al., 2014) from the UCSC chain file (Kuhn et al., 2013) that describes the pairwise alignments between the human and mouse genomes. The abundance (FPKM) of each syntenic transcript in the mouse RNA-seq dataset was finally assessed using the Cufflinks suite (Pollier et al., 2013). A syntenic transcript was considered as “detectable” when its abundance value ≥ 1 FPKM in at least of the
160 developmental stages in the mouse.

Identification of conserved and correlated transcripts using a backtracking method

To define the set of genes showing a conserved and correlated expression profiles throughout fetal gonad development in humans and mice, the first critical step was to properly define the best correspondence
165 scheme between experimental conditions, i.e. between the available developmental stages in both datasets without a priori considerations. We thus calculated a Pearson correlation coefficient for each transcript between every combinations of three sequential developmental stages for Human and the three sequential developmental stages available for the mouse. The sequential combination of three human experimental conditions maximizing the selection of correlated transcripts (P -value of the corresponding
170 Pearson correlation coefficient adjusted with an $FDR \leq 10\%$) was defined as the best correspondence scheme and allowed us to identify the corresponding set of transcripts showing a conserved and correlated expression pattern across species.

FACS sorting and quantitative PCR validation

175 *Single cell dissociation and cell sorting*

Single cell suspensions were obtained from gonads by a standard enzymatic and mechanic digestion procedure. Gonads were cut into small pieces and digested in 0.25% Trypsin-0.02% EDTA (#T4049, Sigma-Aldrich) and 0.05 mg/ml DNase (#DN25, Sigma-Aldrich) for 10 min at 37 °C. Trypsin digestion was stopped by adding 10% fetal bovine serum in M199 media and samples were centrifuged at 350 g for 5 min.
180 Dispersed cells were centrifuged and resuspended in PBS. Cells were counted on a Malassez hemocytometer after labeling of dead cells with Trypan blue. For testes, a plasma membrane labelling of cord cells was performed using a mouse FITC-conjugated anti-human epithelial antigen (clone Ber EP4; Dako # F0860, diluted 1:100) associated with a labelling of germ cells with a mouse R-Phycoerythrin-coupled anti-human KIT/CD117 (clone 104D2, BioLegend, # 313204, diluted 1:100) for 30 min at room
185 temperature (RT). For ovaries, germ cells were labeled with the mouse R-Phycoerythrin-coupled anti-

human KIT/CD117 as described above. Cells were sorted by a flow cytometer cell sorter FACSAriaII (BD Biosciences) equipped with Diva software. Cells were collected in PBS, centrifuged at 500g for 45 min at 4°C and pelleted cells were stored at -80°C until RNA extraction.

190 *Quantitative PCR*

RNA were extracted from cell pellets with PicoPure RNA Isolation Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Total RNAs (100 ng) were reverse transcribed with iScript cDNA synthesis kit (Biorad) and quantitative PCR was performed using the iTaq® universal SYBR green supermix (Biorad) according to manufacturer's instructions in a Cfx384 OneTouch Real-Time PCR system (Biorad). The following amplification program was used: an initial denaturation of 3 min at 95°C, 40 cycles of 10 sec denaturation at 95°C and 30 sec at 62°C for annealing and extension. Dissociation curves were produced using a thermal melting profile performed after the last PCR cycle. To avoid amplification of contaminating genomic DNA, primer pairs flanking an intron were designed. RPLP0 and RPS20 mRNA were used as internal controls for normalization (Table 1). Results were calculated with Bio-Rad CFX Manager 3.1 thanks to the $\Delta\Delta CT$ method as n-fold differences in target gene expression, relative to the reference gene and calibrator sample which is constituted of an equal mixture of all the tested samples for a given organ.

Table 1. List of primers that were used for q-PCR experiments.

Gene	Forward (5'-3')	Reverse (5'-3')	Size (bp)	Reference
RPLP0	TCTACAACCCTGAAGTGCTTGAT	CAATCTGCAGACAGACACTGG	96	PMID: 24743772
RPS20	AACAAGCCGCAACGTAAAATC	ACGATCCCACGTCTTAGAACC	166	PMID: 24743772
CCR1	CAGAAAGCCCCAGAAACAAA	GGTGTGGGAGTTTCCATCC	80	Primer 3
CITED1	TGCACTTGATGTCAAGGGTG	GTTGTAGGAGAGCCATTGG	201	PMID:22703800
CITED1 as	GGGAGAAAAGTAGCCCCAAG	GAAGCAAATGGAGAGACGGA	111	Primer 3
CNTN1	TTGGGAAGATGGTAGCTTGG	TAGTAACAAGGGTTCCAGTGC	113	PMID: 26855587
CXCL14	ATGAAGCAAAGTACCCGCA	TCTCGTCCAGGCGTTGTAC	148	PMID: 24700803

EPHB1	AGAGGAGGGAAAAGGACCAGG	GGTTCCACGGCATCTC	183	PMID: 24121831
ERBB3	ACAGCCCAGATCTGCAC	GTTGGGCGAATGTTCTCATC	78	PMID: 23991224
FATE1	GGCAATTTCCAAGGCATACG	CTAGTCTGCCCACTGCATC	68	PMID: 17761949
FOXL2	GCGAAGTTCCCGTTCTACGA	CTCGTTGAGGCTGAGGTTGT	75	Primer BLAST
KIT	TTCTTACCAGGTGCAAAGG	AAATGCTTTCAGGTGCCATC	209	PMID: 21668453
LIN28A	AAGCTGCACATGGAAGGGTT	CCGCCTCTACTCCCAATAC	138	Primer3
MAGEB1	CTATGGGGAACCCCGTAAGT	GGTTTCAGCATAGGCTCTCG	130	Primer 3
NANOG	TGAACCTCAGCTACAAACAG	TGGTGGTAGGAAGAGTAAAG	154	PMID: 24927536
NPY	CTACATCAACCTCATCACCAGG	TCACCACATTGCAGGGTCT	133	PMID: 27722841
NR2F2	GCCATAGTCTGTTCACCTCA	AATCTCGTCGGCTGGTTG	131	PMID: 24318875
NRXN3	GCTGAGAACAACCCCAATA	ATGCTGGCTGTAGAGCGATT	179	PMID: 28013231
POU5F1	TACTCTCGGTCCCTTTCC	CAAAAACCTGGCACAACCT	131	PMID: 24743772
PTGS1	GTTCTTGCTGTTCTGCTCC	TCACACTGGTAGCGGTCAAG	147	PMID: 27347562
RSPO1	ACACTCCCAGCATCTGAGACCAA	TGCTGAACAGGATGGGAAGAAGGT	146	PMID: 23617070
SET	AAATCAAATGGAATCTGGAAAGG	AAAGAAGCTCTCTGTTCTCATG	101	PMID: 22677993
SOX4	GACCTGCTCGACCTGAACC	CCGGGCTCGAAGTTAAAATCC	107	PMID: 28535514
SOX9	AAGCCTTCATGGTGTGG	TCTCGCTCTCGTTCAGAAGTC	124	Primer3
SOX10	AAGCCTCACATCGACTTCGG	TCCATGTTGGACATTACCTCGT	67	PMID: 23338937
SRY	ACAGTAAAGGCAACGTCCAG	ATCTCGGGGAAGCAAACCTGC	293	PMID: 11869379
TCONS_113718	GCCCAACCACAGAAGGTTT	CTGGGACAGGATGGAGAGG	97	Primer3
TCONS_55038	CCCACCTTCTCCTCCTG	TCAGTGCGAAGAGCCCAA	86	Primer3
TCONS_153406	TCTACTGTTTCTGGAGCTGAAG	TGCAGTAACATCCTCCTCCTC	120	Primer3
TCONS_224470	CCTGGGTACTACTGGTCTTT	GAGCACCTCATTCTTGCT	70	Primer3
TCONS_42565	GCGGCCCTAAGACAAAGAAC	TCTGACCAGAAAATCGCTTC	104	Primer3
THBS1	TTGTCTTTGGAACACACCA	CTGACAGCTCATCACAGG	187	PMID: 27513329
WLS	CCTTGGTTCCAATTATGCT	TTCACTGCTCAGCAAACG	138	Primer3
WT1	CGAGAGCCAGCCCGCTA	TCGAAGGTGACCGTGTGTA	51	PMID: 24534946

Immunohistochemistry and immunofluorescence

205 Upon collection, gonads were fixed either in Bouin fluid fixative or paraformaldehyde 4% for 1 to 2 hours, embedded in paraffin using standard procedures and cut into 5 μ m-thick sections. After dewaxing and rehydration, slides were treated for antigen retrieval with pre-heated 10 mM citrate buffer, pH 6.0 at 80°C for 40 min before cooling at room temperature (RT). Sections were blocked 1 h at RT with 4% bovine serum albumin (BSA) in PBS before the overnight incubation at 4°C with the primary antibody diluted in Dako antibody diluent (Dako Cytomation, Trappes, France). Antibodies and conditions are described in Table 2.

210

Secondary antibodies were goat anti-rabbit antibody biotinylated (E0432, Dako, diluted 1:500); or rabbit anti-mouse biotinylated antibody (E0464, Dako, diluted 1:500). Sections were developed with streptavidin-horseradish peroxidase (Vectastain ABC kit, Vector Laboratories, Burlingame, CA, USA) and 3,3'-diaminobenzidine tetrahydrochloride (Vector Laboratories Inc.) and counterstained with hematoxylin. Stained sections were examined and photographed under light microscopy (Olympus BX51). For cryosectioning, fixed gonads were cryopreserved in PBS-sucrose 20%, embedded in NEG50TM (Allan-Richard Scientific, Thermo Sci.) and cut in 8 μ m-thick sections. Thawed sections were treated for antigen retrieval with citrate buffer as described when necessary (for WT1, NRXN3, CNTN1 and LIN28), rinsed in PBS and incubated overnight at 4°C with primary antibody. Rinsed sections were incubated with the ad hoc fluorescent secondary antibodies (1:500). The second primary antibody was subsequently incubated overnight at 4°C followed by the corresponding secondary antibody. Secondary antibodies were either 488 or 594 Alexa Fluor conjugated antibodies made in chicken for rabbit- and mouse-hosted primary antibodies and in donkey for CNTN1 primary antibody. Sections were mounted in prolong Gold anti-fade reagent with DAPI (Invitrogen, Thermo Fisher Sci). Slides were examined and photographed with an AxioImager microscope equipped with an AxioCam MRc5 camera and the ZEN software (Zeiss, Le Pecq, France).

Table 2. Antibodies used for immunofluorescence and immunohistochemistry.

Antigen		HIER	dilution	Antibody supplier and product number
hEpA-FITC	Mouse FITC-coupled anti-human epithelial antigen	-	1 :100	Dako, F0860
CNTN1	Contactin 1	Ci	1 :100 (IF)	R&D systems, AF904
EPHB1	ephrin type-B receptor 1	Ci	1:2500- 1:5000	Sigma, HPA067740
ERBB3	HER3/ErbB3 (D22C5) XP® Rabbit mAb	Ci	1:250	Cell Signaling Tech., #12708
FATE1	fetal and adult testis expressed 1	Ci	1:1000-1:2500	Sigma, HPA034604

KIAA1210	KIAA1210	Ci	1:200- 1:500	Sigma, HPA048322
KIT/CD117-PE	mouse R-Phycoerythrin-coupled anti-human KIT/CD117; clone 104D2	-	1 :100	BioLegend, 313204
LIN28	lin-28 homolog A	Ci	1 :100 (IF)	Abcam, Ab46020
MAGEB1	melanoma-associated antigen B1	Ci	1:100 - 1:250	Sigma, HPA001193
NRXN3	neurexin 3	Ci	1 :200 / 1 :100 (IF)	Sigma, HPA0002727
SOX9	SRY (sex determining region Y)-box9	Ci	1:100	Millipore, AB5535
SOX10	SRY (sex determining region Y)-box 10	Ci	1:200- 1:500	Sigma, HPA068898
WT1	Wilms tumor 1	Ci	1:100 (IF)	Santa-Cruz Biotech, sc-192

230 **References**

- Adusumilli, R., and Mallick, P. (2017). Data Conversion with ProteoWizard msConvert. In *Methods in Molecular Biology* (Clifton, N.J.), pp. 339–368.
- Akane, A., Seki, S., Shiono, H., Nakamura, H., Hasegawa, M., Kagawa, M., Matsubara, K., Nakahori, Y., Nagafuchi, S., and Nakagome, Y. (1992). Sex determination of forensic samples by dual PCR amplification of an X-Y homologous gene. *Forensic Sci. Int.* *52*, 143–148.
- 235
- Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R., et al. (2015). Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.* *43*, D36–D42.
- Chalmel, F., and Primig, M. (2008). The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology. *BMC Bioinformatics* *9*, 86.
- 240
- Chalmel, F., Lardenois, a., Evrard, B., Rolland, a. D., Sallou, O., Dumargne, M.-C., Coiffec, I., Collin, O., Primig, M., and Jegou, B. (2014). High-Resolution Profiling of Novel Transcribed Regions During Rat Spermatogenesis. *Biol. Reprod.* *91*, 5–5.
- Chocu, S., Evrard, B., Lavigne, R., Rolland, A.D., Aubry, F., Jégou, B., Chalmel, F., and Pineau, C. (2014). Forty-Four Novel Protein-Coding Loci Discovered Using a Proteomics Informed by Transcriptomics (PIT) Approach in Rat Male Germ Cells1. *Biol. Reprod.* *91*, 123–123.
- 245
- Evans, V.C., Barker, G., Heesom, K.J., Fan, J., Bessant, C., and Matthews, D.A. (2012). De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat. Methods* *9*, 1207–1211.
- Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* *39*, W29-37.
- 250
- Friel, A., Houghton, J.A., Glennon, M., Lavery, R., Smith, T., Nolan, A., and Maher, M. (2002). A preliminary report on the implication of RT-PCR detection of DAZ, RBMY1, USP9Y and Protamine-2 mRNA in testicular biopsy samples from azoospermic men. *Int. J. Androl.* *25*, 59–64.
- Jégou, B., Sankararaman, S., Rolland, A.D., Reich, D., and Chalmel, F. (2017). Meiotic Genes Are Enriched in Regions of Reduced Archaic Ancestry. *Mol. Biol. Evol.* *34*, 1974–1980.
- 255
- Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. *Nature* *509*, 575–581.
- Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* *35*, W345-9.
- 260
- Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and associated tools. *Brief. Bioinform.* *14*, 144–161.
- Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* *25*, 1–18.

- 265 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- O’Shaughnessy, P.J., Baker, P.J., Monteiro, A., Cassie, S., Bhattacharya, S., and Fowler, P.A. (2007). Developmental changes in human fetal testicular cell numbers and messenger ribonucleic acid levels during the second trimester. *J. Clin. Endocrinol. Metab.* 92, 4792–4801.
- 270 Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., et al. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591.
- Pollier, J., Rombauts, S., and Goossens, A. (2013). Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol. Biol. Clifton NJ* 1011, 305–315.
- 275 Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I.A., Grasso, C.S., Kominsky, H.D., et al. (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat. Biotechnol.* 29, 742–749.
- 280 Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., et al. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* 42, D756–D763.
- Pundir, S., Magrane, M., Martin, M.J., O’Donovan, C., and UniProt Consortium (2015). Searching and Navigating UniProt Databases. *Curr. Protoc. Bioinforma.* 50, 1.27.1-10.
- 285 Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet. TIG* 16, 276–277.
- Smyth, G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat. Appl. Genet. Mol. Biol.* 3, 1–25.
- 290 Soh, Y.Q.S., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet.* 11, e1005531.
- Speir, M.L., Zweig, A.S., Rosenbloom, K.R., Raney, B.J., Paten, B., Nejad, P., Lee, B.T., Learned, K., Karolchik, D., Hinrichs, A.S., et al. (2016). The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.* 44, D717–D725.
- 295 Trapnell, C., Roberts, A., Goff, L., Petrea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.
- Vaudel, M., Barsnes, H., Berven, F.S., Sickmann, A., and Martens, L. (2011). SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *PROTEOMICS* 11, 996–999.

- 300 Vaudel, M., Burkhart, J.M., Zahedi, R.P., Oveland, E., Berven, F.S., Sickmann, A., Martens, L., and Barsnes, H. (2015). PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* *33*, 22–24.
- Vizcaíno, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., et al. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* *44*, 11033–11033.
- 305 Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* *41*, e74.
- Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. *Nucleic Acids Res.* *44*, D710–D716.
- 310 Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinforma. Oxf. Engl.* *30*, 1006–1007.
- Zimmermann, C., Stévant, I., Borel, C., Conne, B., Pitetti, J.-L., Calvel, P., Kaessmann, H., Jégou, B., Chalmel, F., and Nef, S. (2015). Research Resource: The Dynamic Transcriptional Profile of Sertoli Cells During the Progression of Spermatogenesis. *Mol. Endocrinol.* *29*, 627–642.

315

Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

2.3. Résultats

2.3.1. L'assemblage des transcrits

Après l'étape d'alignement sur le génome, l'assemblage des transcrits à partir des lectures alignées sur le génome humain (version hg19) a été réalisé avec Cufflinks (Trapnell et al., 2012). Comme décrit dans l'introduction (voir « L'assemblage des transcrits »), le développement d'outils d'assemblage reste toujours un challenge bio-informatique. L'un des gènes marqueurs du développement des cellules de Sertoli est le gène de l'AMH, qui joue un rôle central dans la masculinisation du tractus génital mâle (voir « Masculinisation du tractus génital : l'AMH »). Cependant, aucun transcrit correspondant à ce gène n'a été identifié dans l'ensemble des transcrits reconstruits de notre étude. L'examen des fichiers d'alignement (fichiers .bam) dans le navigateur de génome IGV (Thorvaldsdóttir et al., 2013) a permis de mettre en évidence que quatre transcrits assemblés correspondent au gène SF3A2, situé en amont de l'AMH (Figure 21). Il est possible d'observer que des lectures sont alignées à la fois sur le dernier exon de SF3A2 et sur le premier de l'AMH. Ces gènes ont déjà été décrits comme étant contigus (Dresser et al., 1995). L'assembleur a donc reconstruit des transcrits contenant l'ensemble des exons des deux gènes. Il s'agit ici d'erreurs d'assemblage, et non de complètes chimères, mais cela entraîne des erreurs d'identification des isoformes de SF3A2 exprimées et une carence du transcrit de l'AMH identifié. Cela illustre les difficultés que peuvent rencontrer les outils actuellement disponibles pour reconstruire des transcrits et identifier les gènes exprimés.

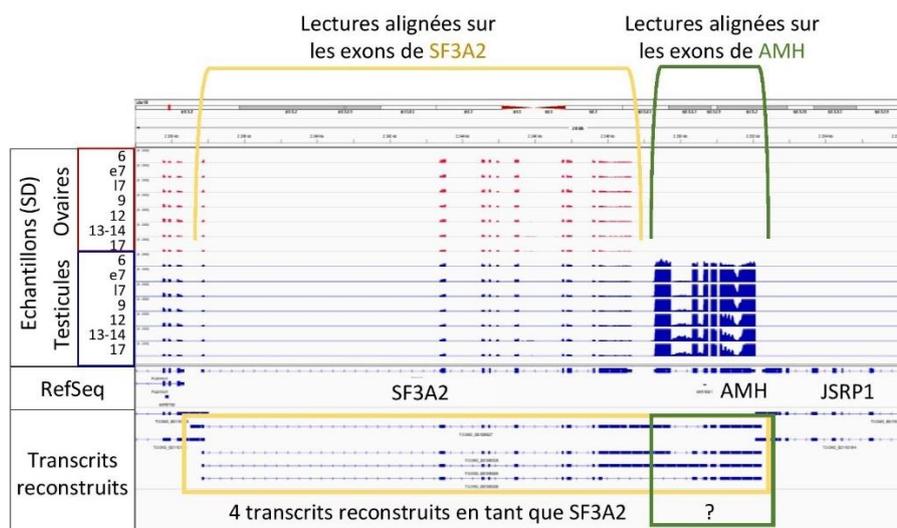


Figure 21. Visualisation de l'alignement des lectures séquencées sur le génome humain hg19 grâce à IGV.

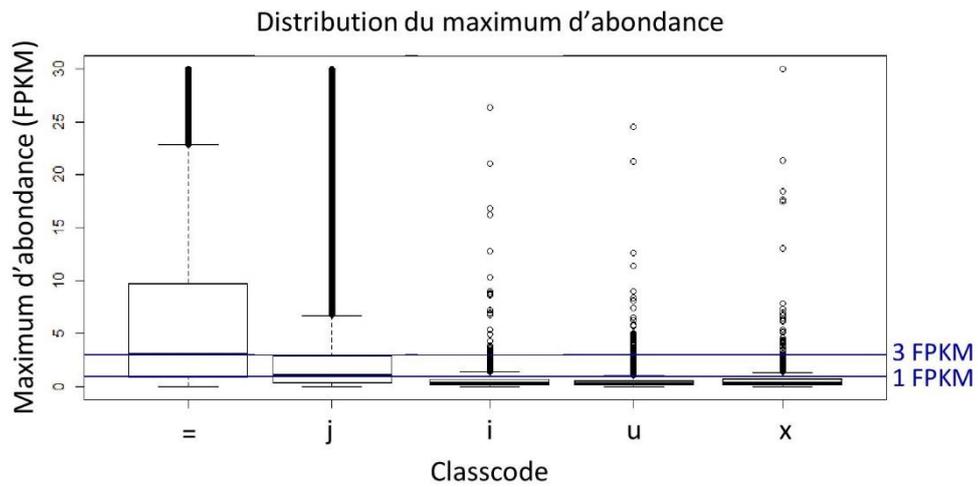


Figure 22. Représentation en « boîte à moustache » du maximum d'abondance des transcrits à travers les échantillons en fonction de leurs classcodes.

Les transcrits « = » correspondent à des isoformes connues; les transcrits « j » correspondent à de nouvelles isoformes de gènes connus; les transcrits « i », « u » et « x » sont de nouvelles isoformes situés dans des régions introniques, intergéniques ou antisens, respectivement. Pour des commodités de visualisation, les valeurs extrêmes (> 30FPKM) ne sont pas représentées.

Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

2.3.2. Le raffinement et le choix du seuil d'expression

Les premières étapes de raffinement sont nécessaires afin d'écarter de l'analyse les transcrits provenant d'erreurs de séquençage ou d'assemblage et de sélectionner les transcrits d'intérêt. La filtration des transcrits de moins de 200 pb et des nouvelles isoformes monoexoniques permet d'éliminer les transcrits chimériques. Initialement, un seuil d'expression de 3 FPKM avait été choisi, seuil au-delà duquel les transcrits sont considérés comme étant détectables (exprimés) dans un échantillon donné.

Dans cette étude, un transcrit détecté à 3 FPKM ou plus au sein d'au moins un échantillon est considéré comme exprimé. Ainsi, sur les 255 915 transcrits reconstruits, seulement 27 721 étaient considérés comme exprimés. Le *classcode* est une catégorisation des transcrits reconstruits par Cufflinks en fonction de leur composition exonique : les transcrits «=» correspondent à des isoformes déjà connues; les transcrits «j» correspondent à de nouvelles isoformes de gènes connus; les transcrits «i», «u» et «x» sont de nouvelles isoformes situées dans des régions introniques, intergéniques ou antisens, respectivement. En observant la distribution du maximum d'expression en fonction du *classcode* des transcrits, ce seuil d'expression est apparu comme trop strict (Figure 22). Ainsi, ce premier seuil ne permettait la sélection que de 12 979 isoformes connues, 14 609 nouvelles isoformes, et seulement 36 isoformes introniques, 36 isoformes en antisens de gènes connus, 61 transcrits intergéniques. Considérant le système dynamique étudié, comprenant une différenciation tissulaire, somme de différenciations cellulaires hétérogènes ainsi qu'une prolifération inégale de chaque population, un faible niveau d'expression peut résulter de plusieurs biais. Par exemple, une faible population cellulaire exprimant significativement le transcrit, une faible expression dans une proportion raisonnable de cellules, ou bien sûre, un artefact d'assemblage. La différenciation globale des gonades humaines n'ayant jamais été caractérisée avant en faisant appel à ce type de technologies qui révèle un exceptionnel potentiel de découverte de gènes. Sur la base d'autres publications dans le domaine (Derrien et al., 2012), le seuil d'expression que nous avons fixé s'est fondé sur un compromis entre une filtration stricte et probabilité de découverte. Avec un seuil à 1 FPKM, 49 904 transcrits étaient sélectionnés, dont 18 733 isoformes connues, 30 308 nouvelles isoformes, 194 isoformes introniques, 186 isoformes en antisens de gènes connus et 483 transcrits intergéniques.

2.3.3. L'analyse d'expression différentielle

Le raffinement analytique choisi a permis de sélectionner 49 904 transcrits, et une deuxième sélection sur leur niveau d'expression dans au moins une condition (moyenne d'expression des

réplicats biologiques) supérieur à 1 FPKM a permis de discriminer 35 194 transcrits, parmi lesquels des transcrits codant pour des protéines, des longs ARN non codants ainsi que des nouveaux gènes/transcrits.

Après filtration statistique, 13 145 transcrits séquencés ont été définis comme « sexuellement dimorphiques », c'est-à-dire qu'ils sont préférentiellement exprimés soit dans le testicule, soit dans l'ovaire, et 8935 comme « sexuellement non dimorphiques », leur niveau d'expression variant de manière similaire au cours du développement de la gonade dans les deux sexes. Une classification (méthode HCPC) a permis de définir 20 groupes de gènes co-exprimés. Une analyse fonctionnelle de ces groupes a mis en évidence des enrichissements significatifs en termes GO associés à chacune de ces classes, cohérents avec la répartition des marqueurs en leur sein. Par exemple, le groupe P4 est enrichi en gènes impliqués dans la stéroïdogénèse et il contient également les gènes marqueurs des cellules de Leydig, comme CYP11A1, CYP17A1, STAR et HSD3B2.

Considérant que la différenciation des gonades jusqu'alors bipotentielles connaît son essor à partir de la 6^{ième} semaine de développement et que le testicule et l'ovaire sont histologiquement indifférenciables à ce stade, l'expression différentielle entre les deux gonades détectées si tôt dans le développement repose sur des mécanismes de régulation transcriptionnelle différents. La sélection de 1479 transcrits présentant un ratio d'expression d'au moins 2 a permis d'identifier ceux qui sont potentiellement impliqués dans la cascade transcriptionnelle initiant la différenciation sexuelle.

2.3.4. La conservation du programme transcriptionnel chez le modèle murin

La souris est le modèle animal le plus utilisé en recherche sur les maladies humaines, il est maintenant bien connu que sa physiologie ne peut être entièrement extrapolée à l'humain. De façon générale, au niveau transcriptionnel, les profils d'expression des gènes sont conservés de façon globale. Toutefois, ce degré de conservation dépend grandement du tissu et des gènes comparés. La plupart des études transcriptomiques comparatives prennent en compte les gènes codant pour des protéines, et de ce fait, occultent le fait que les ARNlnc présentent des propriétés de conservation différentes. Tout d'abord, moins de 10 % des ARNlnc annotés sont orthologues (gènes homologues de différentes espèces ayant évolués à partir d'un ancêtre commun via spéciation), alors que cette propriété est retrouvée chez 75 % des gènes codant pour des protéines (Breschi et al., 2017). Il est important de noter que la plupart des ARNlnc orthologues sont exprimés spécifiquement dans le testicule adulte chez l'homme et la souris (Hezroni et al., 2015; Washietl et al., 2014), et plus particulièrement les gènes les moins conservés (Necsulea et al., 2014). Cette signature transcriptomique particulière qui caractérise la gonade mâle est probablement dû à l'état particulier de la chromatine pendant la spermatogénèse (Soumillon et al., 2013), ce qui participerait à l'évolution

Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

rapide du transcriptome du testicule au regard des autres organes. Étant donné les dynamiques particulières induites par la méiose et la gamétogenèse au niveau épigénomique et transcriptomique, il serait pertinent de déterminer plus précisément les limites du modèle murin, pour mettre en lumière les processus conservés, et donc permettre une extrapolation plus solide des résultats entre les espèces.

2.3.4.1. Intégration des puces à ADN : limites

La première approche pour l'étude de la conservation consistait en l'intégration de données des puces à ADN alors disponibles. L'étude de Small et collaborateurs caractérisait l'expression des gènes de testicules et d'ovaires fœtaux de souris, récoltés à 5 temps de développement (11.5 jpc, 12,5 jpc, 14.5 jpc 16,5 jpc et 18.5 jpc) en duplicat (Small et al., 2005). L'analyse reposait sur les liens d'homologie entre les gènes interrogés sur la puce et les gènes séquencés dans notre étude, grâce à la banque de données HomoloGene (Pruitt et al., 2014). Plusieurs limites techniques sont alors apparues pendant cette analyse : (i) l'asymétrie de l'étude souris (5 conditions mâles, 4 conditions femelles), engendré par la mise à l'écart d'un réplicat femelle suite à son échec au control qualité; (ii) la nécessité de comparer les résultats de deux technologies foncièrement différentes; (iii) les biais techniques particuliers de chaque technologie. Indépendamment, chaque limitation aurait pu être dépassée, mais dans cette étude, cela impliquait beaucoup de conjectures et d'extrapolations qui limitaient la pertinence de la comparaison.

2.3.4.2. Intégration de données de RNA-seq

Dans notre travail, nous avons décidé d'intégrer des données de RNA-seq publiées par Soh et collaborateurs (Soh et al., 2015). Cette étude se concentre sur le développement des cellules germinales et sur la méiose femelle lors du développement de l'ovaire. Pour cela, une analyse par RNA-seq d'ovaires fœtaux de souris mutantes *Stra8*^{-/-}, *Dazl*^{-/-} et des gonades (testicules et ovaires) déplétées en cellules germinales a été effectuée. Dans ce contexte, les auteurs ont également analysé des testicules et ovaires fœtaux au cours de trois périodes de développement (12,5 jpc, 14,5 jpc et 16,5 jpc) en duplicat et par RNA-seq, en tant que contrôles. Ce sont ces 12 échantillons que nous avons utilisés pour l'intégration à notre étude.

2.3.4.2.1. Analyse bio-informatique et conversion des données de RNA-seq souris vers le génome humain.

Afin de quantifier et de caractériser l'expression des gènes souris, les lectures ont été alignées sur le génome souris (mm10) par TopHat2 (Kim et al., 2013). Afin de pouvoir quantifier

l'expression des gènes séquencés chez la souris par rapport aux gènes exprimés chez l'humain, les fichiers .chain fournis par l'UCSC ont été utilisés pour effectuer une conversion entre les deux génomes (Speir et al., 2016). Brièvement, l'alignement complet du génome humain et du génome souris a permis de définir les régions génomiques qui sont conservées et synténiques au niveau de leur séquence. Les coordonnées des transcrits assemblés dans l'étude de RNA-seq humaine ont ainsi été converties vers le génome souris.

La quantification des transcrits synténiques a été effectuée avec les lectures alignées de l'étude de Soh *et al.* Le même seuil d'expression de 1 FPKM a été utilisé fin de sélectionner les transcrits humains exprimés chez la souris. Parmi les 30 491 transcrits synténiques exprimés chez l'homme, 17 026 sont également exprimés pendant le développement des gonades souris.

2.3.4.2.2. Association des stades de développement humains et souris et corrélation des profils d'expression des transcrits

L'une des problématiques pour comparer les profils d'expression des gènes chez les deux espèces est que les deux études ne prenaient pas en compte le même nombre de conditions (sept stades de développement chez l'humain et trois chez la souris). Il était donc ici nécessaire d'associer les trois temps de développement chez la souris avec trois temps de développement humain, afin d'évaluer la corrélation des profils d'expression. L'association optimale maximisant le nombre de gènes dont la cinétique est corrélée entre l'humain et la souris est d'une part, pour le testicule, 7 tardif-9-12 SD comparé à 12.5-14.5-16.5 jpc chez la souris, et d'autre part, pour l'ovaire, 9-12-13, 14 SD comparé à 12.5-14.5-16.5 jpc. Grâce à cette méthode, nous avons pu mettre en évidence que 1288 transcrits présentent un profil d'expression significativement conservé au cours du développement des gonades humaines et souris.

2.4. Discussion et Perspectives

En France et dans la plupart des pays européens, le délai légal d'avortement est limité à 12 SD. La collaboration avec l'équipe de Paul Fowler en Écosse, où le délai légal de l'IVG est de 22 SD, nous a permis de compléter la description du programme transcriptionnel jusqu'aux semaines les plus tardives du développement.

Considérant la complexité, la sensibilité et les résultats obtenus au fil des analyses informatiques, nous avons choisi d'orienter la publication vers les phénomènes les plus précoces de la différenciation sexuelle. Cependant, l'ensemble du projet constitue une ressource précieuse pour l'étude des différents processus de différenciation des populations somatiques et germinales. La différenciation sexuelle prend place pendant la 6^{ème} SD, le testicule et l'ovaire sont encore morphologiquement indifférenciables, tant par leur taille que par leurs populations cellulaires. Au

Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

cours de la 7^{ème} SD, les cellules de Sertoli se différencient et prolifèrent, la composition cellulaire du testicule et de l'ovaire diverge, tant par la nature des populations que par leur proportion. Dès que les phénomènes de prolifération entrent en jeu, il faut garder à l'esprit que la quantification de transcrits reflète leur niveau d'expression et la proportion des cellules qui les expriment au sein de la gonade. La description de la dynamique globale des gènes du moment le moins différencié jusqu'aux âges tardifs permet d'apprécier leur devenir au sein du testicule et de l'ovaire.

Dans l'ovaire, la caractérisation du programme transcriptionnel jusqu'à 17SD nous a permis d'observer l'activation des gènes méiotiques des ovocytes et le début de la folliculogénèse, phénomène rarement observé chez l'humain. Les analyses d'enrichissement ont démontré que les groupes de gènes co-exprimés associés à la méiose et à l'ovogénèse étaient enrichis en ARNlnc et en nouveaux gènes. Le même phénomène a été précédemment observé lors de la spermatogénèse de rat, où les ARNlnc sont préférentiellement exprimés lors de la méiose et de la spermiogénèse (Chalmel et al., 2014). Il serait donc nécessaire d'approfondir l'analyse de la méiose dans l'ovaire fœtal, les nouveaux gènes impliqués et comparer ces résultats avec le testicule adulte. Dans le testicule, de nombreux phénomènes restent également à caractériser. Notamment, ces données ont permis d'approfondir l'étude de la stéroïdogénèse et l'origine du cholestérol au sein de la cellule Leydig fœtale (étude menée par le Dr Séverine Mazaud-Guittot de l'Irset).

2.4.1. L'étude comparative des mécanismes de régulation transcriptionnelle de l'Homme et de la souris

En faisant abstraction des connaissances sur les analogies physiologiques des gonades au cours de leur différenciation entre l'Homme et la souris, nous avons tenté de développer une méthode afin d'associer chaque stade de développement sur la base des données d'expression des gènes. Ainsi, nous avons associé pour le testicule, 7 tardif-9-12 SD à 12.5-14.5-16.5 jpc chez la souris, et d'autre part, pour l'ovaire, 9-12-13, 14 SD à 12.5-14.5-16.5 jpc. La différence d'association entre les temps de développement masculin et féminin peut être expliquée par le fait que le transcriptome testiculaire diverge plus rapidement à partir du moment de la différenciation sexuelle, l'ovaire subissant moins de bouleversement transcriptionnel. L'analyse de corrélation des profils d'expression a abouti à la sélection de 1288 transcrits ayant une dynamique d'expression significativement conservée au cours du développement des gonades humaines et souris. Cet ensemble de gènes dont la séquence et le patron d'expression sont conservés semble être des acteurs importants du développement gonadique mammifère. De plus, il a été démontré que les séquences régulatrices des gènes sont en général moins conservées que leurs exons (Wang and Rekaya, 2009) et que la conservation du profil d'expression

des gènes est due à des processus compensatoires mis en place pour les gènes importants (Breschi et al., 2017). L'analyse approfondie de ces mécanismes conservés présente un fort potentiel d'élucidation des spécificités inter-espèces du développement gonadique. Une étude comparative des motifs de fixation des facteurs de transcription, l'étude de la conservation des séquences promotrices entre les espèces ou l'intégration de données d'épigénétique issues de l'étude de plusieurs espèces pourraient décrypter les processus conservés de régulation.

Une autre étude de RNA-seq a été récemment publiée par l'équipe du Pr Koopman et s'intéresse particulièrement à la différenciation sexuelle souris. Des testicules et ovaires murins à 4 stades de développement ont été séquencés, englobant la différenciation (10,5, 11,5, 12,5, 13,5 jpc). Il serait possible d'appliquer la méthode que nous avons développée en concentrant l'analyse sur les stades de développement humains les plus précoces, et établir quels gènes présentent un profil d'expression conservé pendant les premières étapes de différenciation des testicules et des ovaires.

Chez le rat, nous n'avons recensé aucune étude de RNA-seq sur l'étude des gonades à ce jour. Il serait opportun de générer ce type de données, car de nombreux laboratoires utilisent ce modèle animal pour leurs études toxicologiques. Au-delà de l'aspect fondamental de l'étude, il est impératif de mieux appréhender les similarités et surtout les divergences entre l'Homme et les modèles animaux couramment utilisés en toxicologie afin de savoir si les résultats obtenus sont extrapolables à l'espèce humaine.

2.4.2. L'intégration des données de GWAS

GWAS Catalog (MacArthur et al., 2017) et GWASdb (Li et al., 2012) sont deux bases de données qui répertorient les *Single Nucleotide Polymorphism* (SNP) associés à des pathologies ou phénotypes humains. Les résultats préliminaires de l'intégration de ces données indiquent que 525 transcrits différenciellement exprimés au cours du développement des gonades fœtales humaines ont déjà été associés à des pathologies du tractus génital humain. Au-delà des marqueurs du développement gonadiques et des gènes connus pour être impliqués dans divers processus de différenciation spécifiques des gonades (Gene Ontology, Gene Ontology Consortium, 2015) ou désordres associés (Human Phenotype Ontology, Köhler et al., 2014), 460 transcrits peuvent potentiellement occuper une place importante dans le programme transcriptionnel au vu des conséquences phénotypiques d'une mutation dans leur séquence et méritent une analyse plus approfondie.

Cependant, il est important de souligner que les ressources en termes de mutations de gènes impliqués dans les pathologies humaines sont limitées, car le domaine de la recherche clinique reste réticent quant au partage de ses données. D'une part, GWAS Catalog est enrichi grâce aux

Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

contributions des scientifiques, qui restent limitées, et d'autre part, GWASdb n'a pas été mis à jour depuis octobre 2015.

2.4.3. La caractérisation transcriptomique du développement des gonades fœtales humaines à l'échelle de la cellule unique : scRNA-seq

Le RNA-seq et le scRNA-seq permettent d'appréhender les processus biologiques à des échelles de résolution différentes, en termes de matériel biologique et de sensibilité de séquençage (voir «

RNA-seq à l'échelle de la cellule unique, scRNA-seq »). D'une part, le « bulk » RNA-seq sur gonade totale permet de limiter les modifications induites par le traitement des échantillons, d'effectuer un séquençage profond et brin spécifique, de reconstruire les transcrits, de découvrir de nouvelles isoformes et de nouveaux gènes. D'autre part, le scRNA-seq permet de capturer une partie des cellules gonadiques en différenciation, mais implique également de lourds protocoles de dissociation des organes et de capture de cellule et donc des modifications du transcriptome potentiellement plus importantes. Cependant, cette caractérisation à l'échelle de la cellule unique permet une appréciation

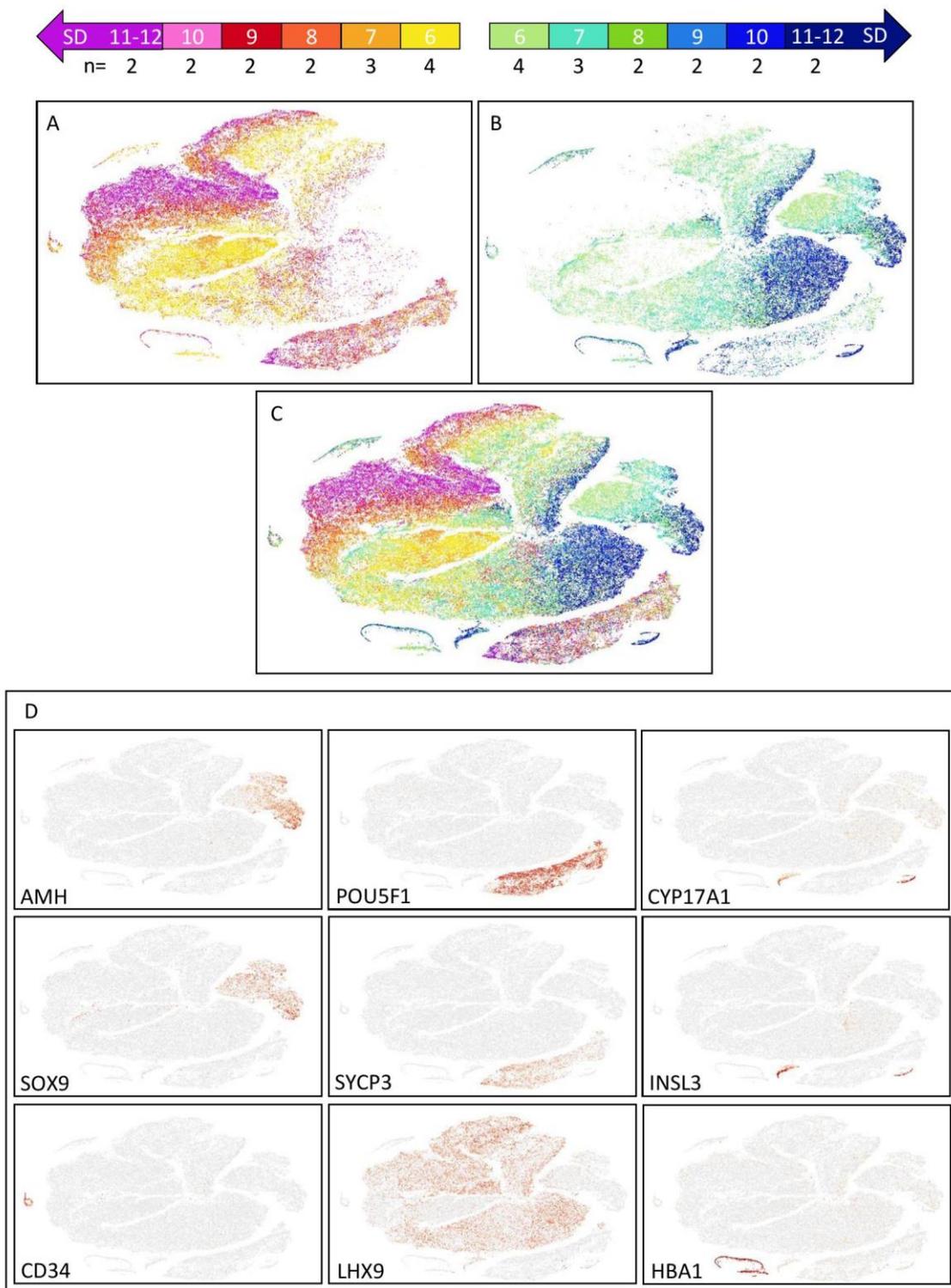


Figure 23. Représentation en nuage de point des résultats de pré-séquence des 90 000 cellules capturées à partir des 30 gonades fœtales humaines.

Les ovaires et testicules fœtaux humains ont été prélevés à 5 temps de développement différents, en quatre réplicats pour les plus jeunes (6 SD), trois réplicats pour les gonades à 7SD, et en duplicat pour les trois autres temps de développement. Après dissociation des tissus, la procédure de capture a permis d'isoler 90 000 cellules à travers l'ensemble des échantillons. Une première étape de pré-séquence a permis d'évaluer la qualité des données. La réduction de la dimensionnalité des données transcriptomiques de chaque cellule a permis une représentation en nuage de point de l'ensemble des cellules. Les cellules isolées à partir des ovaires fœtaux (A) et testicules fœtaux (B) sont représentées et colorées en fonction du stade de développement de la gonade à partir de laquelle elles ont été capturées (même couleur pour chaque réplicat biologique d'un même stade de développement). (D) Expression de certains gènes marqueurs au sein des 90 000 cellules (le rouge le plus intense

Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

représentant une forte expression au sein de la cellule) : des marqueurs des cellules de Sertoli (AMH, SOX9), des cellules de Leydig (CYP17A1, INSL3), des cellules germinales (POU5F1), y compris méiotiques (SYCP3), des macrophages (CD34), des progéniteurs du stroma ovarien (LHX9) et des cellules sanguines (HBA1).

précise de l'hétérogénéité cellulaire et des processus transcriptionnels gouvernant la différenciation de chaque population. Ces deux approches sont donc complémentaires et devraient permettre des avancées sans précédent dans la compréhension du programme transcriptionnel de la différenciation et du développement des gonades fœtales humaines.

Notre équipe a mis en place un projet de scRNA-seq pour l'étude du développement des gonades fœtales humaine, permettant l'analyse transcriptomique des cellules isolées à partir de 30 échantillons de testicules et d'ovaires humains, entre 6 et 12 SD (Figure 23). L'objectif de ce projet est d'identifier l'ensemble des populations cellulaires tout au long du développement des gonades et déterminer leur filiation (ou *lineage*) lors de leur différenciation.

Environ 90 000 cellules ont été capturées à partir des 30 échantillons. Une première étape de pré-séquençage a permis d'évaluer la qualité et la variabilité des échantillons. Cette caractérisation préliminaire en une seule procédure de séquençage permet d'appréhender le nombre de cellules capturées à partir de chaque gonade, d'évaluer le nombre de gènes exprimés dans chacune d'entre elles et de vérifier la qualité des échantillons. Une première analyse par t-SNE permet de réduire la dimensionnalité des données et ainsi de représenter chaque cellule en fonction de son profil transcriptomique en nuage de point (*scatter plot*) (Figure 23) (Maaten and Hinton, 2008). Cela permet de visualiser que le transcriptome des cellules ovariennes (Figure 23A) et des cellules testiculaires (Figure 23B) évolue bien en fonction du stade de développement de la gonade, avec une répartition au centre du nuage des cellules les plus précoces (jaunes pour l'ovaire, vert clair pour le testicule) vers la périphérie du nuage des cellules les plus âgées (violet pour l'ovaire, bleu foncé pour le testicule). La représentation de l'ensemble des cellules dans le même nuage de points (Figure 23C) permet de visualiser la divergence des transcriptomes des cellules ovariennes (vers la gauche) et des cellules testiculaires (vers la droite) à partir d'une origine commune (centre du nuage).

La procédure de séquençage plus profond permettra de caractériser de manière plus fine le transcriptome de chaque cellule, ce qui permettra une analyse puissante de la différenciation de chaque lignée. Cependant, il est déjà possible de visualiser l'expression de certains gènes grâce à ces résultats préliminaires (Figure 23D). L'expression des marqueurs connus permet d'observer le regroupement des différentes populations cellulaires (progéniteurs : LHX9; Sertoli : AMH, SOX9 ; Leydig : INSL3, CYP17A1). Les cellules germinales testiculaires et ovariennes se regroupent au sein d'un même ensemble de cellules, marquées par l'expression du gène POU5F1, bien que certains

oocytes entrent déjà en méiose, exprimant le gène SYCP3. Il est important de noter que, relativement aux connaissances accumulées concernant le développement testiculaire, les connaissances sur le développement ovarien sont parcellaires. Cette première étude présente un très fort potentiel de découverte, notamment dans la découverte de marqueurs spécifiques des différentes populations cellulaires ovariennes.

Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

3. Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

3.1. Introduction

Depuis les années 1990, de nombreuses études suggèrent une hausse de l'incidence de certaines anomalies et pathologies du système reproducteur mâle dans la population humaine des pays industrialisés, se traduisant par : dans certains cas une dégradation aussi bien quantitative que qualitative du sperme (Andersen et al., 2000; Auger et al., 1995; Carlsen et al., 1992; Fisch et al., 1996; Irvine et al., 1996; Jegou and Velez De La Calle, 1993) ; une augmentation de la fréquence du cancer testiculaire (Adami et al., 1994; MØLLER, 1998) ; et une hausse de certaines malformations congénitales du tractus urogénital, telles que la cryptorchidie (non descente du (des) testicule(s) dans le scrotum) et l'hypospadias (non-fermeture de l'urètre) (Campbell et al., 1987; Chilvers et al., 1984; Matlai and Beral, 1985; Paulozzi et al., 1997). En 2001, le Danois Niels Skakkebaek a émis l'hypothèse que ces pathologies pourraient dériver de causes communes combinant susceptibilité(s) génétique(s) et facteurs environnementaux agissant au cours de la vie fœtale : le concept du *Testicular dysgenesis syndrome* (TDS) (Serrano et al., 2013; Skakkebak et al., 2001). Bien que l'hypothèse du TDS soit sujette à débats (Skakkebaek et al., 2016), il n'en demeure pas moins que de nombreuses études, aussi bien épidémiologiques que toxicologiques, suggèrent l'implication de diverses molécules aux propriétés de perturbateur endocrinien (PE) pendant la vie fœtale dans le développement de ces anomalies du développement du testicule (Andersen et al., 2008; Ben Maamar et al., 2015, 2017; Brucker-Davis et al., 2008; Damgaard et al., 2006; Gaudriault et al., 2017; Kristensen et al., 2012; Krysiak-Baltyn et al., 2010, 2012).

L'Organisation Mondiale de la Santé définit les PEs comme étant des « substances exogènes ou des mélanges qui altèrent la/les fonction(s) du système endocrinien et par voie de conséquence causent un effet délétère sur la santé d'un individu, sa descendance ou des sous-populations » (Schug et al., 2016). De par leur définition, les PEs constituent une famille hétérogène de molécules et peuvent interférer avec le système endocrinien par des mécanismes variés : ils peuvent interférer avec la synthèse des hormones, leur transport, leur métabolisation, ou leur interaction avec leurs récepteurs. Dans tous les cas, cela se traduit soit par une altération des niveaux circulants de ces hormones, soit par une altération de l'action de ces hormones sur leurs cellules cibles. Plusieurs catégories de composés chimiques et contaminants ubiquitaires environnementaux sont déjà connues comme ayant des propriétés de PE, tel que les phtalates, les bisphénols, les parabènes et certaines molécules pharmaceutiques (Giulivo et al., 2016). La mise en place des études toxicologiques pour

l'évaluation de leur innocuité est complexe. Les PE peuvent en effet : i) agir à faibles voire très faibles doses, ii) présenter des effets contraires suivant leur concentration (effets « non monotones ») et, iii) induire des effets différents en fonction du stade de développement (notion de fenêtre de susceptibilité) (Vandenberg et al., 2012). D'autre part, l'exposition constante des populations, de la vie fœtale à la vie adulte, est difficilement reproductible.

Afin d'élucider les mécanismes d'action des PE au cours du développement des gonades humaines, l'Irset a mis en place des protocoles de culture *in vitro* de testicules fœtaux humains afin de tester l'effet de PE, en mesurant par exemple la production de testostérone après l'exposition de l'organe à un produit chimique d'intérêt (Ben Maamar et al., 2015, 2017; Mazaud-Guittot et al., 2013). Ce protocole *ex vivo* permet de conserver la complexité du tissu pendant la culture, et ainsi de maintenir sa production de testostérone et de mesurer sa perturbation par un xénobiotique. Les travaux de l'équipe permettent de déterminer si chaque toxique testé perturbe la production de testostérone (ainsi que d'autres hormones stéroïdiennes et peptidiques) au sein des testicules fœtaux humains. Ces études abordent généralement la problématique des mécanismes d'action de ces substances par des approches bas-débit types « gène candidat » en validant l'expression de certains gènes d'intérêt par qPCR et éventuellement immunohistochimie. Plus récemment, mon laboratoire d'accueil a décidé d'aborder ce dernier point par l'utilisation de méthodes de transcriptomique haut-débit afin de décrypter l'ensemble des perturbations transcriptionnelles occasionnées par ces produits chimiques. Dans ce contexte, deux projets distincts ont été mis en place dans l'équipe sur des potentiels PE, l'ibuprofène et le chlordécone, déjà étudiés dans l'équipe, dont j'avais la responsabilité de l'analyse bioinformatique (Ben Maamar et al., 2017; Gaudriault et al., 2017; Mazaud-Guittot et al., 2013). L'objectif était de définir les signatures toxicogénomique des composés, correspondant à l'ensemble des gènes différentiellement exprimés suite à l'exposition d'un échantillon (un organe, une culture cellulaire) à un composé chimique, afin de mieux appréhender les cellules dont l'activité et les processus biologiques sont altérées.

3.1.1. L'ibuprofène

L'ibuprofène est un anti-inflammatoire non stéroïdien du groupe des propioniques dérivés de l'acide arylcarboxylique. Il possède des effets antalgiques, antipyrétiques, anti-inflammatoires et inhibiteurs de courte durée des fonctions plaquettaires. Il est référencé comme médicament essentiel par l'OMS (WHO 2013).

C'est en 1961 que les propriétés antalgiques de l'ibuprofène sont découvertes par le laboratoire Boots, dans le cadre de la recherche d'une alternative à l'aspirine, qui possède de nombreux effets indésirables, notamment gastriques. Il est commercialisé pour la première fois en

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

1969 en Angleterre comme traitement contre l'arthrite rhumatoïde, et deviendra disponible en médicament conseil au milieu des années 1980.

L'ibuprofène inhibe l'activité des cyclo-oxygénases (COX) impliquées dans la synthèse des prostaglandines, médiateurs de l'inflammation. Sa capacité à inhiber la synthèse de thromboxane A₂ lui confère des propriétés d'antiagrégant plaquettaire. D'autres mécanismes d'action ont été montrés ou sont supposés (pour revue, Rainsford, 2009). Lors de la prise d'un comprimé par l'adulte, le pic plasmatique est atteint environ 90 min après ingestion. Une dose thérapeutique de 200 mg entraînera une concentration sérique maximale de 8.2×10^{-5} M, et une dose de 400 mg aboutira à une concentration de 1.45×10^{-4} M. L'ibuprofène est capable de traverser la barrière placentaire, car il a déjà été détecté dans le méconium (premières selles du nouveau-né) (Alano et al., 2001). Néanmoins, aucune donnée n'est disponible quant au pourcentage auquel est exposé le fœtus après consommation de la mère. Aucune malformation du fœtus induite par l'exposition de la mère pendant la grossesse (effet tératogène) n'a été observée chez l'animal. Il est cependant contre-indiqué pendant le troisième trimestre chez la femme et doit être évité pendant les deux premiers, en particulier à cause de sa toxicité rénale. Cependant, à cause de sa libre accessibilité en officine, la prise d'ibuprofène pendant la grossesse est documentée (Kristensen et al., 2012, 2016). Au niveau des malformations du tractus génital, une étude épidémiologique a démontré une association chez l'homme entre hypospadias et prise d'ibuprofène pendant la grossesse (Lind et al. 2013). Récemment, il a également été démontré que l'exposition à l'ibuprofène du testicule fœtal humain induit une altération de la production de la testostérone (Ben Maamar et al., 2017).

Ce premier projet de toxicogénomique avait pour but d'étudier l'effet de l'ibuprofène conjointement avec l'effet de l'aspirine et du paracétamol, antalgiques consommés par les femmes pendant la grossesse et suspectés d'être des perturbateurs endocriniens (Kristensen et al., 2012, 2016; Mazaud-Guittot et al., 2013). L'étude complète comprenait donc l'exposition de quatre testicules fœtaux âgés de 8-9 SD, divisés en quatre sections égales pour les quatre conditions testées (un contrôle, trois traités). Ensuite, les effets de ces trois médicaments ont été évalués par une analyse de l'expression différentielle des gènes induite par l'exposition (décrite dans « La filtration et analyse de l'expression différentielle »). Dans ce contexte, l'aspirine et le paracétamol n'induisait pas de réponse transcriptionnelle significative dans le testicule fœtal en raison d'une importante variabilité inter-individuelle des échantillons (données non présentées). Seule la signature toxicogénomique de l'ibuprofène a pu être définie et sera présentée ici.

3.1.2. Le chlordécone

Le chlordécone est un pesticide de la famille des organochlorés, massivement utilisé à partir des années 70. En France, il a notamment été utilisé aux Antilles sur les cultures de bananier, afin de lutter contre le charançon. C'est un polluant organique persistant avec une demi-vie maximale de 46 ans dans l'environnement et comprise entre 120 et 160 jours dans l'organisme (Cohn et al. 1978) selon les conditions. Sa bio-persistence couplée à la révélation de sa toxicité a abouti à son interdiction dans de très nombreux pays dès 1976, et en 1993 en France, mais il est toujours détecté dans les nappes phréatiques, les populations sont donc toujours exposées (DIREN 2001).

Dès le milieu des années 1970 des signes de toxicité du chlordécone ont été mis en évidence aux États-Unis chez les travailleurs exposés au produit. Ces patients présentaient des symptômes cliniques d'un empoisonnement chronique ciblant le système nerveux central (céphalées, troubles oculomoteurs, pertes de mémoire), hypertrophie hépatique et des troubles touchant le système reproducteur (diminution de la fraction de spermatozoïdes motiles) (Cannon et al. 1978; Cohn et al. 1978; Taylor 1982). La plupart des études subséquentes ont mesuré les effets du composé sur le tissu hépatique. Des expérimentations sur des oiseaux et des mammifères ont retrouvé les mêmes symptômes sur le système reproducteur que ceux observés chez les travailleurs exposés (Faroon et al. 1995).

Au niveau épidémiologique, l'effet du chlordécone sur la santé humaine est un des axes de recherche de l'équipe 9 de l'Irset, dirigée par Luc Multigner et Philippe Quénel. Ils ont notamment démontré qu'une exposition au chlordécone était associée avec un risque accru de développement de cancer de la prostate (Multigner et al. 2015). La cohorte mère-enfant Timoun a également été mise en place en Guadeloupe afin d'évaluer l'effet d'une exposition prénatale des fœtus sur leur développement, y compris après la naissance. Au niveau toxicologique, il a été montré que le chlordécone est capable de traverser la barrière placentaire, et de se lier au récepteur des œstrogènes et de stimuler la synthèse *in vivo* du récepteur de la progestérone dans le tissu utérin de ratte (Hammond et al. 1979), traduisant son potentiel effet perturbateur endocrinien. Enfin, l'exposition de testicules fœtaux humains à ce polluant altère leur production de testostérone, ce qui confirme son statut de PE probable (Gaudriault et al., 2017).

Le deuxième projet de toxicogénomique faisait suite au travail de thèse de Pierre Gaudriault au sein de notre équipe (Gaudriault et al., 2017). Il avait mis en évidence l'effet de 8 substances chimiques diverses sur la production de testostérone du testicule fœtal humain, dont 2 polluants industriels, le BPA et le BPS, un produit pharmaceutique, la théophylline (également métabolite de la caféine) et 5 pesticides (le bitertanol, le prochloraz, le propiconazole, le chlordécone et l'imazalil). L'objectif de ce projet était de définir le mécanisme d'action de ces 8 molécules à leurs doses EC₅₀

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

respectives (dose à laquelle une diminution de 50 % de production de testostérone est induite), pour obtenir des signatures toxicogénomiques comparables. Ainsi, le but était de décrypter les différentes voies possibles de perturbation endocrinienne au sein du testicule fœtal humain. Cette étude ambitieuse nécessitait donc des testicules de taille assez importante (10-12 SD) pour pouvoir effectuer au moins 9 explants et ainsi tester toutes les conditions (8 substances, 1 contrôle). De manière similaire à l'étude des antalgiques, les résultats de la filtration statistique ont également été décevants, le chlordécone étant le seul composé induisant un effet assez important pour pouvoir extraire une signature toxigénomique de nos résultats, présentée dans ce chapitre. Bien que chaque composé ait été testé à son EC₅₀, c'est-à-dire à la concentration à laquelle il induit une diminution de 50 % de la production de testostérone par le testicule, aucune signature n'a pu être extraite pour 7 des 8 conditions testées.

3.2. Matériel et Méthodes

3.2.1. La collecte des échantillons, culture et exposition

Tout comme pour le chapitre précédent (« Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines »), les testicules fœtaux humains sont issus d'interruptions volontaires de grossesse (IVG) effectuées à l'hôpital Sud de Rennes. Il s'agit donc de femmes enceintes ne souhaitant pas garder leur enfant pour des raisons non-médicales. Ils ne présentent donc pas un taux d'anomalie supérieur à la moyenne de la population générale. Le projet a été approuvé par le comité d'éthique local et l'agence de la biomédecine (autorisation #PFS09-011, "Etude de différenciation normale et pathologique des gonades humaines" ; Agence de la Biomédecine). Les fœtus sont âgés de 6 à 12 semaines de développement (soit 8 à 14 semaines d'aménorrhée). La procédure nécessite une posologie précise de plusieurs médicaments (Mifégyne; Cytotec; ibuprofène; midazolam, flagyl ou granudoxy), suivi de l'IVG, réalisé par aspiration sous anesthésie locale ou générale. Les testicules sont récupérés au sein des débris d'aspiration.

Chaque testicule est fragmenté en morceaux d'environ 1 mm³. Les explants tissulaires sont ensuite transférés dans des puits stériles contenant le milieu de culture. Pendant les 24 premières heures, le milieu ne contient que de l'hCG. Ensuite, les explants sont exposés au composé d'intérêt pendant 72h. D'une part, trois testicules d'environ 8 SD sont fragmentés (8 SD et 2 jours : 8.2 SD; 8 SD et 4 jours : 8.4 ; 8 SD et 5 jours : 8.5 SD), une partie des explants servant de contrôle, une autre partie étant exposée à l'ibuprofène à 10⁻⁵M. Cette dose a été choisie, car c'est la dose la plus faible entraînant une perturbation de la stéroïdogénèse et n'induit pas de différences histologiques observables sur les coupes de testicules fœtaux (Figure 24).

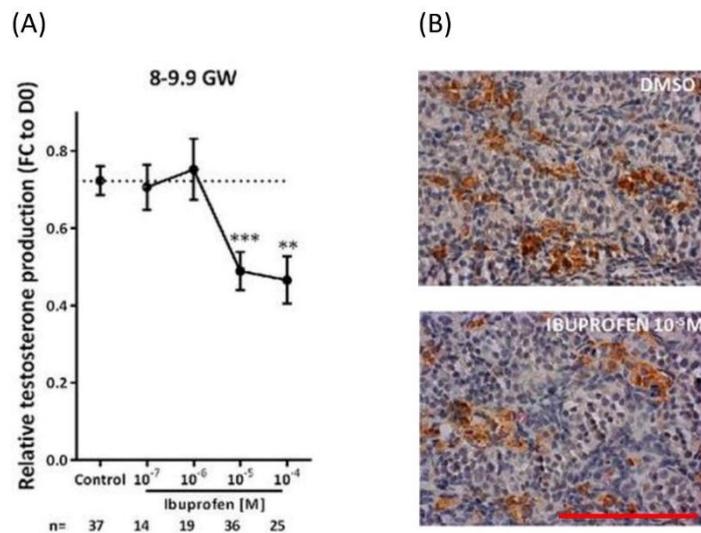


Figure 24. Effet de l'ibuprofène sur le testicule fœtal humain en culture organotypique (issu de Ben Maamar et al., 2017).

(A) Courbe dose-réponse de la production de testostérone par les testicules fœtaux de 8 à 9.9 SD (GW) en fonction de la concentration d'ibuprofène. Les explants sont cultivés en présence hCG pendant 24 h (D0) puis exposés à l'ibuprofène ou au DMSO pendant 72h. Les résultats présentés sont des ratios de production de testostérone après 72h de traitement par rapport au contrôle D0 (%). Les valeurs représentent la moyenne des réplicats (n) +/- l'erreur standard à la moyenne. Un test de Wilcoxon a été appliqué pour évaluer la significativité de la variabilité (**p<0.01 ; ***p<0.001). (B) Coupes d'immunomarquage de CYP11A1 d'explants testiculaires de 7.9 SD cultivés en présence de DMSO (contrôle) ou d'ibuprofène 10⁻⁵M. CYP11A1 est coloré en brun (coloration au 3,3'-diaminobenzidine tetrahydrochloride) et les coupes sont contre-marquées avec de l'hématoxylin. Echelle: 100 μm.

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

D'une manière similaire, trois testicules d'environ 11 SD sont fragmentés (11SD et 2 jours : 11.2 SD; 11 SD et 4 jours : 11.4 ; 11 SD et 6 jours : 11.6 SD), une partie servant de contrôle, une autre partie étant exposée au chlordécone à son EC₅₀ (dose induisant une diminution de 50 % de la production de testostérone), correspondant à 1,66x10⁻⁵M (Figure 25). Cette dose n'est pas cytotoxique.

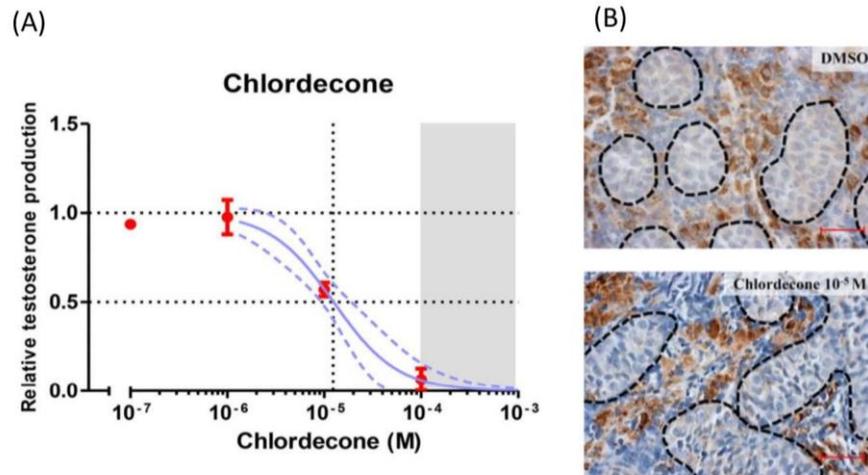


Figure 25. Effet du chlordécone sur le testicule fœtal humain en culture organotypique (issu de Gaudriault et al., 2017).

(A) Courbe dose-réponse de la production de testostérone par les testicules fœtaux de 8 à 9.9 SD (GW) en fonction de la concentration d'ibuprofène. Les explants sont cultivés en présence hCG pendant 24 h (D0) puis exposés à l'ibuprofène ou au DMSO pendant 72h. Les résultats présentés sont des ratios de production de testostérone après 72h de traitement par rapport au contrôle D0 (%). Les valeurs représentent la moyenne des réplicats (n) +/- l'erreur standard à la moyenne. L'aire grise représente les concentrations auxquelles le composé devient cytotoxique. (B) Coupes d'immunomarquage de CYP11A1 d'explants testiculaires de 10-12 SD cultivés en présence de DMSO (contrôle) ou de chlordécone 10⁻⁵M. CYP11A1 est coloré en brun (coloration au 3,3'-diaminobenzidine tetrahydrochloride) et les coupes sont contre-marquées avec de l'hématoxylin. Les cordons testiculaires et le tissu interstitiel sont délimités pour les lignes pointillées. Échelle : 20 μ m.

3.2.2. L'extraction de l'ARN et séquençage

L'ARN total a été extrait des échantillons grâce au kit RNeasy mini de Qiagen, quantifié par le spectrophotomètre NanoDrop™ 8000 (Thermo Scientific) et sa qualité évaluée par le 2100 Electrophoresis Bioanalyzer (Agilent). Les bibliothèques préparées sont adaptées pour conserver l'information du sens de transcription (brin spécifique), et les ARN polyadénylés sont sélectionnés avant la fragmentation. Les fragments d'ARN ont été ensuite retro-transcrit afin d'obtenir les ADNc pour le séquençage. Ceux-ci ont été liés aux adaptateurs grâce à la T4 DNA Ligase, puis ont subi

une amplification par PCR (30 sec à 98°C; [10 sec à 98°C, 30 sec à 60°C, 30 sec à 72°C] x 12 cycles; 5 min à 72°C). Les amorces de PCR en excès ont été éliminées grâce au AMPure XP beads (Agencourt Biosciences Corporation). La qualité des librairies ainsi préparées a été évaluée par le 2100 Electrophoresis Bioanalyzer (Agilent). Elles ont ensuite été appliquées sur le support de séquençage (*flowcell*) à une concentration de 7pM et séquencées par l'Illumina HiSeq 2500 à partir d'une seule extrémité (single-end), générant des lectures de 50 pb. L'analyse d'image et la capture du signal ont été effectuées par RTA 1.17.20 and CASAVA 1.8.2.

3.2.3. Les analyses bio-informatiques

3.2.3.1. L'alignement et quantification

Suite au séquençage, les lectures de chaque échantillon ont été alignées sur le génome humain (hg38) grâce à l'outil STAR (Dobin et al., 2013) en utilisant le transcriptome de référence RefSeq (fichier gtf téléchargé en aout 2017) (Pruitt et al., 2014). Une première étape d'alignement sur le transcriptome de référence a permis de définir un ensemble de nouvelles jonctions exoniques, lesquelles seront fournies à l'outil lors d'un deuxième alignement sur le génome. Cela permet d'optimiser la qualité des résultats. Aucune étape de reconstruction des transcrits n'a été effectuée sur ces données. StringTie a été utilisé pour la quantification des gènes dans chacun des échantillons (Pertea et al., 2015) (version 1.3.3) et fournis des abondances en RPKM (Reads per kilo base per million). La médiane de l'expression au sein des réplicats biologiques a permis de définir l'abondance de chaque gène dans les deux conditions testées (échantillons traités vs échantillons contrôles).

3.2.3.2. La filtration et analyse de l'expression différentielle

Un seuil d'expression de 0,5 FPKM est utilisé pour sélectionner les gènes qui sont exprimés dans au moins un échantillon (contrôle ou traité) pour chacun des projets. Ensuite, les gènes présentant un ratio d'au moins 1,5 entre l'expression dans le le testicule exposé et celle dans le testicule contrôle sont considérés comme différentiellement exprimés. Enfin, les échantillons étant appariés (chaque testicule provenant d'un même individu a été fragmenté en plusieurs explants utilisés dans chaque condition (contrôle et traité), un t-test est appliqué pour évaluer la significativité de la variation de l'expression observée. La p-value est ensuite ajustée par la méthode de Benjamini-Hochberg et un seuil de $\leq 0,1$ est appliqué afin de discriminer les gènes dont l'expression est significativement modifiée.

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

3.2.3.3. Les analyses fonctionnelles

L'enrichissement en termes GeneOntology dans chaque groupe est évalué par une loi hypergéométrique dans AMEN (Chalmel and Primig, 2008), et considéré comme significatif lorsque la p-value (ajustée par la méthode de Benjamini-Hochberg) est inférieure à 5 %. Cela permet d'évaluer si un nombre significatif de gènes impliqués dans un même processus (désigné par un terme GO) sont surreprésentés dans un groupe donné, au regard du nombre de gènes connus participant à ce processus. Le package KEGGprofile (Shilin et al., 2014) implémenté dans R a été utilisé afin d'évaluer l'enrichissement de voies métaboliques au sein des gènes sur- ou sous-exprimés après exposition à l'ibuprofène ou au chlordécone. La loi hypergéométrique est également utilisée, et les termes KEGG avec une p-value ajustée inférieure ou égale à 5 % sont considérés comme significativement enrichis. Enfin, une méthode similaire d'évaluation de l'enrichissement de ces gènes par une loi hypergéométrique est également utilisée pour évaluer la sur-représentation des gènes différentiellement exprimés au sein des groupes de gènes définis dans le projet de l'« Étude transcriptomique de la différenciation et du développement des gonades fœtales humaines ». Pour rappel, ces groupes sont définis par un partitionnement sur la base de leur profil d'expression pendant le développement du testicule et de l'ovaire fœtal humain.

3.3. Résultats

3.3.1. Séquençage et alignement sur le génome humain

Une moyenne de 64 888 917 de lectures ont été séquencées au sein des échantillons de 8 SD contrôles et exposés à l'ibuprofène, dont 79,7 % sont alignées une seule fois sur le génome humain (hg38). Dans l'étude du chlordécone, 60 423 941 lectures sont séquencées en moyenne dans les échantillons contrôles et traités, dont 65,1 % sont alignées une seule fois sur le génome humain (Figure 26).

	FT_Ibu_8.2	FT_Ibu_8.4	FT_Ibu_8.5	FT_Ctrl_8.2	FT_Ctrl_8.4	FT_Ctrl_8.5
Nb lectures	62327080	71048938	55807486	66799931	65098898	68251166
Nb lectures alignées une fois (%)	48716001 (78.16)	57418962 (80.82)	43892745 (78.65)	53739353 (80.45)	52181692 (80.16)	54502632 (79.86)
Nb lectures alignées plusieurs fois (%)	7287734 (11.69)	8205422 (11.55)	6827057 (12.23)	7715063 (11.55)	7768186 (11.93)	8066951 (11.82)
	FT_CHL_11.2	FT_CHL_11.4	FT_CHL_11.6	FT_Ctrl_11.2	FT_Ctrl_11.4	FT_Ctrl_11.6
Nb lectures	58704536	61950014	49493819	81162074	73413980	37819224
Nb lectures alignées une fois (%)	35114998 (59.82)	33154732 (53.52)	31545606 (63.74)	55205796 (68.02)	54156763 (73.77)	27169376 (71.84)
Nb lectures alignées plusieurs fois (%)	6561966 (11.18)	6837081 (11.04)	5599236 (11.31)	9027509 (11.12)	8122639 (11.06)	4216496 (11.15)

Figure 26. Résultats du séquençage et de l'alignement sur le génome humain (hg38)

3.3.2. Quantification et filtration statistique

La quantification des gènes et la filtration statistique ont permis de définir la signature toxicogénomique de chaque composé dans le testicule foetal humain (Figure 27). D'une part, l'exposition à l'ibuprofène induit l'expression différentielle de 1359 gènes, dont 1048 sont sous-exprimés, et 311 sont sur-exprimés. D'autre part, le chlordécone induit une altération de l'expression de 3948 gènes, dont 2599 sont sous-exprimés, et 1349 sont sur-exprimés.

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

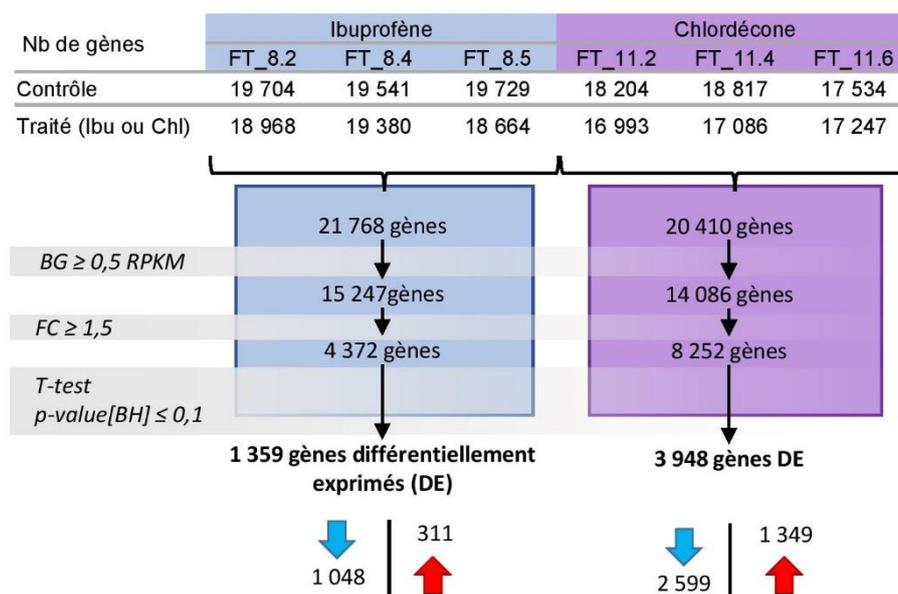


Figure 27. Résultats de la quantification des gènes et de la filtration statistique.

Le tableau présente le nombre de gènes détectés dans chaque échantillon (> 0 RPKM) (FT_8.2 : testicule fœtal âgé de 8SD et 2 jours, voir Matériel et Méthodes), puis le nombre de gènes sélectionnés à chaque étape de la filtration des données issues de l'étude de l'ibuprofène (bleu) et du chlordécone (violet). Les flèches descendantes bleues symbolisent les gènes sous-exprimés, les flèches ascendantes rouges symbolisent les gènes sur-exprimés.

3.3.3. La signature toxicogénomique de l'ibuprofène

La dose de 10^{-5} M est une dose non cytotoxique pour le testicule fœtal humain (Figure 24B), l'exposition de l'organe à cette dose n'entraîne donc pas de variabilité dans la proportion de chaque population cellulaire. De fait, la variation d'expression des gènes résulte bien d'un effet transcriptionnel au sein de cellules et non d'une altération du tissu. Aucun marqueur spécifique du développement des gonades n'est surexprimé, et aucune voie KEGG n'est sur-représentée dans ce groupe.

Au regard des gènes sous-exprimés, l'activité de plusieurs populations cellulaires semble être perturbée (Figure 28B). Tout d'abord, le gène *NR5A1(SFI)*, marqueur des cellules de Leydig différenciées, est sous-exprimé après exposition du testicule fœtal à l'ibuprofène. Ensuite, le gène *DHCR7*, responsable de la dernière étape de la production du cholestérol, présente également une expression altérée. De fait, la production du premier substrat nécessaire à la stéroïdogénèse semble être inhibée.

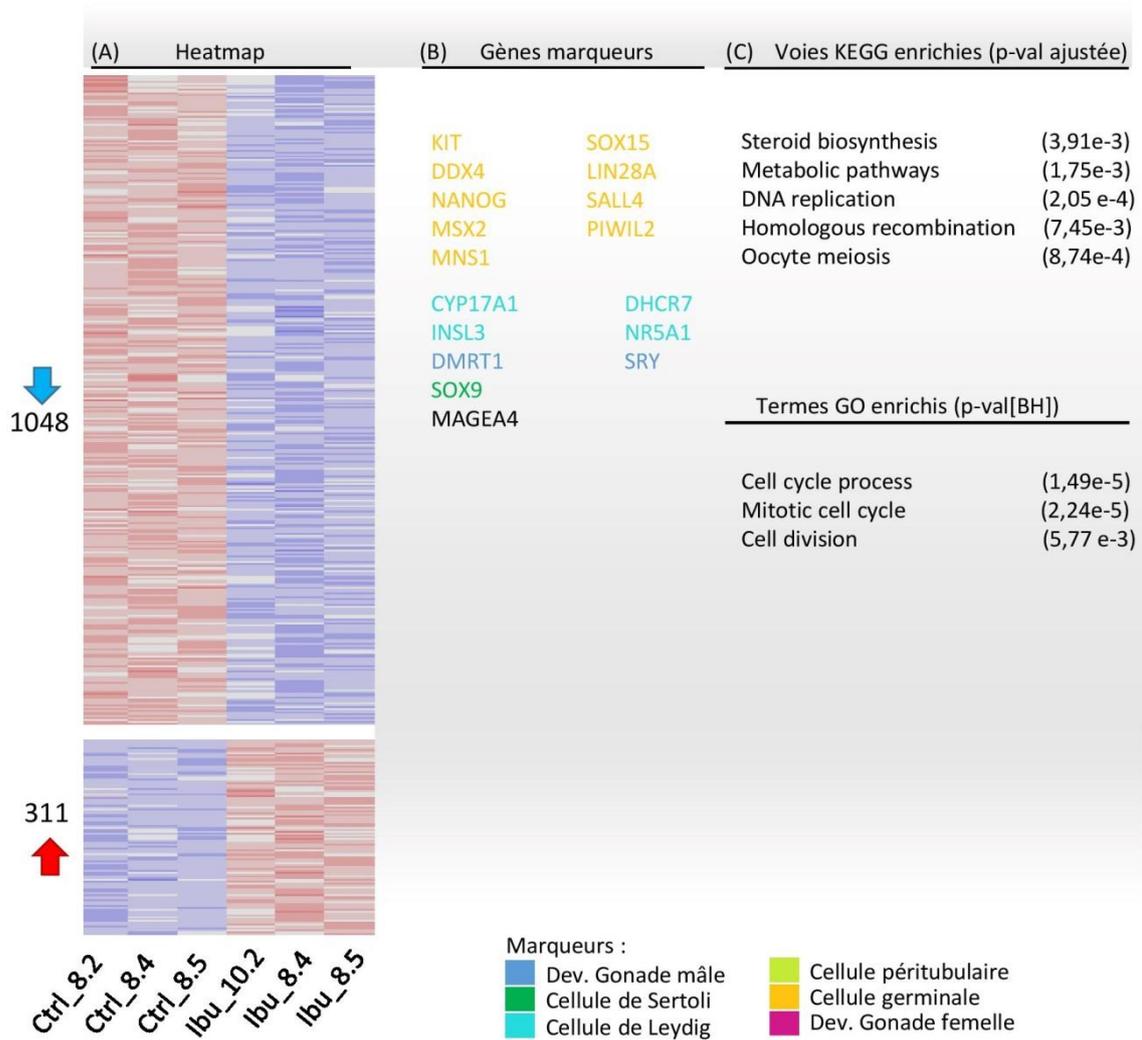


Figure 28. Présentation des groupes de gènes sur- ou sous exprimés dans le testicule foetal après exposition à l'ibuprofène.

(A) Heatmap représentant l'expression de chaque gène dans les six échantillons. Chaque ligne correspond à un gène, chaque colonne à un échantillon. L'expression de chaque gène est standardisée, le rouge représentant une forte expression, le bleu représentant une faible expression. (B) Gènes marqueurs des différentes populations cellulaires du testicule foetal. (C) Résultats des enrichissements en voie KEGG des groupes de gènes sous- et surexprimés.

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

Au niveau de la machinerie responsable de la production de testostérone, le gène *CYP17A1* est sous-exprimé suite à l'exposition à l'ibuprofène, ceci pouvant expliquer la baisse de la production de testostérone. Au sein des cellules de Leydig, on remarque également que l'*INSL3* est sous-exprimée. Cette hormone peptidique étant responsable de la première phase de la descente testiculaire dans l'abdomen, ce qui supporte l'hypothèse que l'exposition à l'ibuprofène pourrait être en lien avec une incidence élevée de la cryptorchidie (Kristensen et al., 2011). On retrouve également une altération de l'expression de *SRY* et de *SOX9*, marqueurs clés des cellules de Sertoli. L'activité des cellules germinales semble également être particulièrement affectée par l'ibuprofène, au vu du nombre de marqueurs dont l'expression est diminuée (Figure 28)

Les calculs d'enrichissement en voies KEGG et en termes GeneOntology permettent d'appréhender les processus biologiques affectés par l'exposition. Les résultats sont cohérents avec ce que les marqueurs affectés laissaient présager (Figure 28C). D'une part, la biosynthèse des stéroïdes (*steroid biosynthesis*) est significativement représentée parmi les gènes sous-exprimés. Bien que l'on mesure les effets de l'ibuprofène sur du testicule fœtal, les gènes associés à la méiose des ovocytes sont également surreprésentés. Cela peut être expliqué par le nombre de marqueurs germinaux affectés. Considérant les processus biologiques, l'activité mitotique des cellules du testicule exposé semble être altérée, au vu du nombre significatif de gènes impliqués et sous-exprimés. Considérant la dynamique de ces gènes pendant le développement des gonades fœtales humaines, les gènes sous-exprimés après exposition à l'ibuprofène sont sur-représentés (p -value : 0.011) dans le groupe P2 SDG, comprenant des gènes présentant un pic d'expression vers la 7^{ème} SD dans le testicule, associé à la différenciation des cellules de Sertoli.

3.3.4. La signature toxico-génomique du chlordécone

Il a été démontré que la dose EC_{50} ($1,66 \times 10^{-5}M$) du chlordécone n'induit pas de cytotoxicité dans les cultures organotypiques des testicules fœtaux humains (Figure 25), donc la variation d'expression des gènes provient bien de modifications transcriptionnelles au sein de cellules et non d'une altération du tissu.

D'une part, de nombreux marqueurs de populations cellulaires des gonades au sein des gènes présentant une expression diminuée. Tout d'abord, seul le gène *HSD17B3*, impliqué dans la stéroïdogénèse, apparaît comme sous-exprimé. De nombreux gènes marqueurs des cellules de Sertoli sont également altérés, comme *SOX9*, *AMH*, *DHH* et *WT1* (Figure 29B). Cela est concordant avec la sur-représentation des gènes sous-exprimés dans le groupe P2 SDG du projet présenté précédemment

(p-value : 4.6e-08), associé à la différenciation des cellules de Sertoli. De nombreux gènes germinaux présentent également une expression altérée. Cependant, contrairement à l'ibuprofène, aucune voie KEGG sur-représenté par les gènes sur- ou sous-exprimés n'est en lien avec un processus particulier du développement testiculaire (Figure 29C).

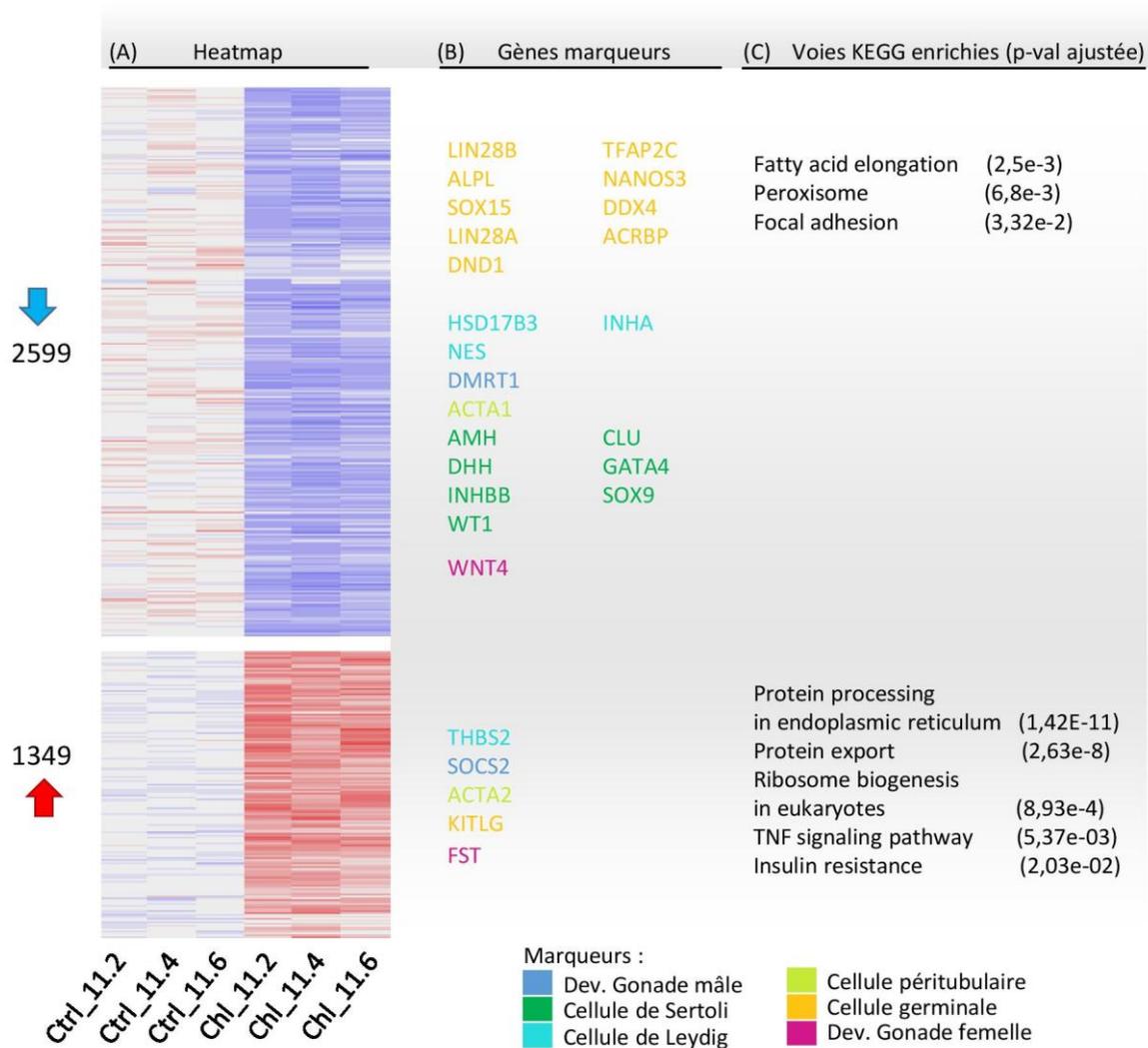


Figure 29. Présentation des groupes de gènes sur- ou sous exprimés dans le testicule foetal après exposition au chlordécone.

(A) Heatmap représentant l'expression de chaque gène dans les six échantillons. Chaque ligne correspond à un gène, chaque colonne à un échantillon. L'expression de chaque gène est standardisée, le rouge représentant une forte expression, le bleu représentant une faible expression. (B) Gènes marqueurs des différentes populations cellulaires du testicule foetal. (C) Résultats des enrichissements en voie KEGG des groupes de gènes sous- et surexprimés.

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

3.3.5. La comparaison des effets transcriptionnels des deux substances

L'ibuprofène et le chlordécone sont deux substances suspectées d'être des perturbateurs endocriniens (PE), et induisent une altération du niveau de production de la testostérone par les testicules fœtaux humains. Les PE ont une large diversité de mécanismes d'action, et la comparaison des altérations transcriptionnelles provoquées par les deux composés peut mettre en évidence les similarités et différences de leurs modes de perturbation.

Les signatures toxicogénomiques des deux substances partagent de nombreux gènes différemment exprimés en commun. Le chevauchement de 377 gènes sous-exprimés communs et de 49 gènes sur-exprimés communs apparaissent comme significatifs par rapport au nombre total de gènes (p -value < 0.05). Au vu des marqueurs dont l'expression est altérée par les deux substances (présentés précédemment), les mêmes types cellulaires semblent être affectés par l'ibuprofène et le chlordécone.

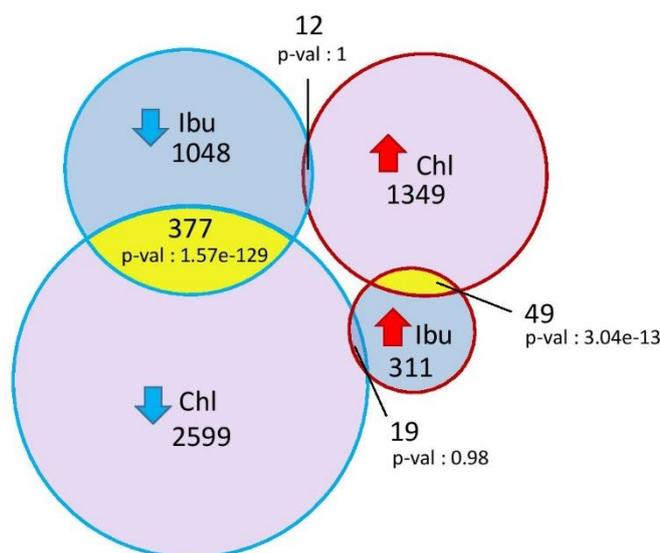


Figure 30. Diagramme de Venn représentant les gènes communs aux deux signatures toxicogénomiques de l'ibuprofène (Ibu) et du chlordécone (Chl) dans le testicule fœtal humain.

Les gènes sur-exprimés sont symbolisés par les flèches ascendantes rouges. Les gènes sous-exprimés sont symbolisés par les flèches descendantes bleues. Les p -values représentent la significativité des chevauchements entre les groupes de gènes différemment exprimés, calculée par une loi hypergéométrique.

3.4. Discussion et perspectives

Les résultats obtenus suite à l'exposition du testicule fœtal à l'ibuprofène sont relativement cohérents avec les résultats obtenus précédemment dans l'équipe (Ben Maamar et al., 2017). Cette étude toxicologique a permis de démontrer la diminution de la production de testostérone suite à l'exposition et l'évaluation de l'expression de certains gènes par qPCR avait permis d'apprécier les mécanismes d'action en jeu.

Au niveau de la machinerie de la stéroïdogénèse, les gènes CYP11A1, CYP17A1, HSD17B3 présentaient une diminution de leur expression, alors que seul CYP17A1 est sous-exprimé dans nos résultats. En ce qui concerne les cellules de Sertoli, cette étude avait démontré que la production d'AMH était réduite. Le gène de l'AMH est connu pour être une des cibles directes de SOX9, mais bien que le facteur de transcription ait été détecté comme sous-exprimé dans nos résultats, son gène cible ne présentait pas d'expression différentielle significative. De la même manière que CYP11A1 et HSD17B3, la figure 31 montre que la valeur d'expression médiane de l'AMH dans les testicules exposés est bien inférieure à celle des contrôles. La variabilité trop importante entre les réplicats biologiques n'a pas permis d'aboutir à une significativité de l'expression différentielle entre contrôle et traité par le t-test. Cela pointe l'un des problèmes principaux posés par l'étude des tissus humains : la variabilité inter-individuelle. De fait, il semble nécessaire d'utiliser davantage de réplicats biologiques pour obtenir une signature toxico-génomique robuste.

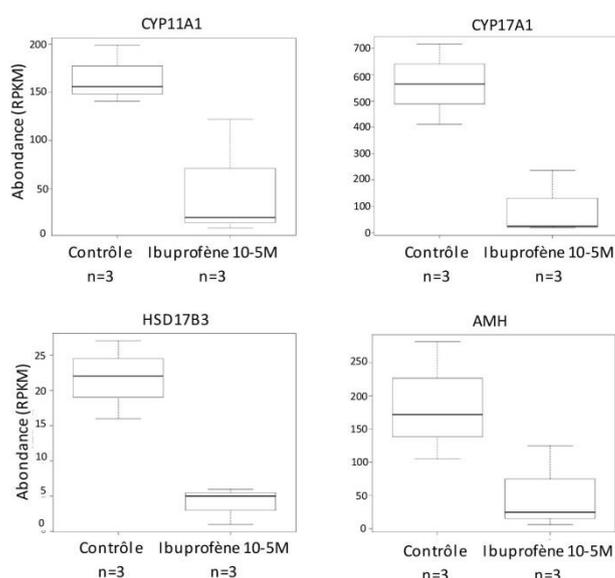


Figure 31. Expression de CYP11A1, CYP17A1, HSD17B3, et de l'AMH.

La représentation en « boîtes à moustaches » permet de visualiser la médiane des valeurs d'expression (ligne en gras) et les valeurs extrêmes. L'abondance est exprimée en RPKM.

Ainsi, pour les deux études de toxicogénomique, il est apparu que les protocoles expérimentaux incluaient des biais importants qui induisaient une variabilité trop importante au sein des échantillons. D'une part, il est possible que la division des testicules en morceaux égaux induise

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

des proportions inégales des populations cellulaires au sein de chaque explant. D'autre part, chaque population cellulaire possède sa propre sensibilité à la toxicité induite par un composé donné, mais il est également possible que les cellules répondent de manière hétérogène au sein d'un même type cellulaire. En effet, nous étudions ici l'altération d'organes humains en développement, ce qui induit donc que des populations sont en cours de différenciation et de développement asynchrone. De fait, il serait nécessaire d'augmenter significativement le nombre de réplicats biologiques pour une étude aussi sensible des mécanismes d'altération induits par des perturbateurs endocriniens suspectés.

Les résultats présentés ici sont préliminaires et l'analyse des voies transcriptionnelles affectées par l'ibuprofène et le chlordécone demande encore à être approfondie. Cependant, les points communs entre les effets des deux substances peuvent soulever une hypothèse sur leur mode d'action. D'une part, ce sont les deux seules substances provoquant une perturbation assez importante sur l'expression des gènes testiculaires pour induire une signature toxicogénomique significative malgré les biais de variabilité inter-individuelle énoncés précédemment. Ensuite, les testicules fœtaux exposés aux deux substances ont environ trois semaines de développement de différences, les processus de différenciation et de développement ne sont pas les mêmes à ces stades de développement. Cependant, le chevauchement entre les gènes sous-exprimés suite à l'exposition apparaît comme significatif et ces deux groupes de gènes sont sur-représentés dans le groupe de gènes associés à la différenciation des cellules de Sertoli. L'effet de l'ibuprofène sur les cellules de Sertoli était apparu comme un mécanisme de perturbation inédit dans l'étude toxicologique effectuée dans l'équipe (Ben Maamar et al., 2017). Cependant, il semble que les deux molécules induisant une forte perturbation du transcriptome du testicule fœtal –indépendamment de son âge- altèrent toutes les deux l'activité des cellules de Sertoli. Considérant le rôle central de cette population cellulaire dans le développement testiculaire et dans le support de la stéroïdogénèse (voir « Cellules de Sertoli »), l'hypothèse peut être émise selon laquelle le mécanisme d'action primaire de l'ibuprofène et du chlordécone réside dans l'altération des cellules de Sertoli, déclenchant en conséquence une cascade de perturbations pour l'ensemble du testicule en développement, cellules de Leydig et cellules germinales comprises. Ces résultats sont préliminaires, et méritent de plus amples investigations et validations.

Bien que ces deux études toxicogénomiques des 11 substances testées n'aient pas eu les résultats escomptés, cela a permis d'adapter les protocoles de transcriptomique aux limitations techniques de la toxicologie et du travail sur tissu humain. Les études toxicologiques menées par notre équipe incluent souvent un nombre conséquent de réplicats biologiques pour tester l'effet de chaque substance sur la production de testostérone. Le matériel biologique étant rare et les procédures

de séquençage étant coûteuses, ces premières études de toxicogénomique ont été élaborées afin de trouver le meilleur compromis entre toutes les limitations rencontrées. Il apparaît maintenant évident que de tels projets requièrent bien plus de réplicats biologiques, afin de surpasser les variabilités inter-individuelles et obtenir des signatures toxicogénomiques robustes. Grâce au développement récent du DGE-seq, il est maintenant possible de quantifier l'expression des gènes aussi sensiblement que le RNA-seq mais à un prix réduit (environ 30 € le séquençage d'un échantillon) (voir « Le DGE-seq »). De fait, il est possible de développer des protocoles plus adaptés à la toxicologie, comprenant davantage de réplicats. Le prochain projet de toxicogénomique mis en place dans l'équipe emploiera cette technologie afin d'observer l'effet du paracétamol à plusieurs doses, comparé aux effets d'autres composés (Ibuprofène, Aspirine, Diclofénac, Naproxène et Méfénamine) sur les ovaires fœtaux humains, comprenant au total 192 échantillons.

Concernant la réponse hétérogène au sein d'une même population cellulaire, il serait opportun d'étudier la réponse transcriptionnelle à l'échelle de la cellule unique, grâce au scRNA-seq (voir « RNA-seq à l'échelle de la cellule unique, scRNA-seq ») et ainsi évaluer l'hétérogénéité de réponse d'une même population cellulaire et identifier celles dont l'activité est altérée.

Impact de l'ibuprofène et du chlordécone sur le programme d'expression du testicule fœtal humain

4. The Reprogenomic Viewer

4.1. Introduction

Grâce aux nouvelles technologies haut débit, les mécanismes moléculaires en jeu dans un tissu ou une cellule peuvent être précisément décryptés, tant au niveau génomique, épigénomique ou transcriptomique. Dans tous les domaines de la recherche, le partage des données permet de faciliter les collaborations, la transparence et la reproductibilité des résultats. Actuellement, les études travaillant à partir de données générées par un tiers peuvent avoir un impact aussi fort sur la communauté scientifique que les études travaillant à partir de données originales (Milham et al., 2017). En pratique, les données -omiques représentent une source indispensable, rapide et fiable pour toute la communauté scientifique, travaillant sur un gène ou une voie particulière. Cependant, une analyse pertinente requiert des connaissances et des compétences bio-informatiques spécifiques, qui ne sont pas toujours disponibles au sein des équipes de recherche.

Comme décrit dans l'introduction, de nombreuses bases de données généralistes existent, agrégeant les connaissances sur la séquence des gènes, leurs annotations, les transcrits et protéines codés, et il existe également des espaces de dépôts de données haut-débit générées par diverses technologies (voir « Les banques de données généralistes »). Certaines bases de données sont spécialisées dans la reproduction, mais aucune n'est adaptée aux données de séquençage (voir « Les banques de données dédiées à la reproduction »). Or, ces informations hétérogènes et complémentaires (transcriptomique, épigénomique, cistromique) ne sont pas facilement accessibles pour les membres de notre communauté. C'est dans ce contexte, mon équipe d'accueil a développé *The ReproGenomics Viewer* (ou RGV) afin de faciliter l'accès à ces données. RGV est une base de données multi- et inter-espèce qui héberge de nombreuses études « omiques », centrées sur la thématique de la reproduction, et son navigateur de génome permet une visualisation rapide des résultats (Darde et al., 2015). La conversion des données entre les espèces utilise les résultats des alignements des génomes entre eux afin de déterminer les régions génomiques conservées, permettant ainsi de visualiser et de comparer les résultats obtenus chez différents organismes.

4.2. Objectifs

Afin d'approcher toujours plus d'exhaustivité dans le contenu de RGV, les objectifs de cette seconde version étaient d'optimiser les protocoles pour une intégration facilitée de nouvelles études, d'uniformiser les analyses des données pour les rendre plus comparables, de développer des approches permettant l'analyse de données issues de nouvelles technologies (telles que celles ayant trait au *single-cell RNA-seq*) et de développer des outils pour une visualisation rapide et adaptée à

chaque type de données. Dans le cadre de ma thèse de doctorat, j'ai significativement contribué à la mise-à-jour de cette base. J'ai notamment pris en charge le recueil et l'annotation manuelle de l'ensemble des études génomiques (plus de 90) actuellement hébergées par le site.

Databases and ontologies

The ReproGenomics Viewer: a multi-omics and cross-species resource compatible with single-cell studies for the reproductive science community

Thomas A. Darde^{1,†}, Estelle Lecluze^{1,†}, Aurélie Lardenois¹, Isabelle Stévant², Nathan Alary¹, Frank Tüttelmann³, Olivier Collin⁴, Serge Nef², Bernard Jégou¹, Antoine D. Rolland¹ and Frédéric Chalmel^{1,*}

¹ Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR_S 1085, F-35000 Rennes, France, ² Department of Genetic Medicine and Development, University of Geneva, 1211 Geneva, Switzerland, ³ Institute of Human Genetics, University of Münster, Münster, Germany, ⁴ Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA/INRIA) - GenOuest platform, Université de Rennes 1, F-35042 Rennes, France,

*To whom correspondence should be addressed.

†These authors contributed equally to this work

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Recent advances in transcriptomics have enabled unprecedented insight into gene expression analysis at a single-cell resolution. While it is anticipated that the number of publications based on such technologies will increase in the next decade, there is currently no public resource to centralize and enable scientists to explore single-cell datasets published in the field of reproductive biology.

Results: Here, we present a major update of the ReproGenomics Viewer (RGV), a cross-species and cross-technology web-based resource of manually-curated sequencing datasets related to reproduction. The redesign of RGV's architecture is accompanied by significant growth of the database content including several landmark single-cell RNA sequencing datasets. The implementation of additional tools enables users to visualize and browse the complex, high-dimensional data now being generated in the reproductive field.

Availability and implementation: The ReproGenomics Viewer resource is freely accessible at <http://rgv.genouest.org>. The website is implemented in Python, JavaScript, and MongoDB, and is compatible with all major browsers. Source codes can be downloaded from <https://github.com/fchalmel/RGV>.

Contact: frederic.chalmel@inserm.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Advances in ultra-high-throughput sequencing technologies, including increased accuracy, diverse applications, and decreased overall cost have significantly contributed to their democratization (Sims *et al.*, 2014). As these techniques became more popular, a myriad of datasets were pub-

TA. Darde *et al.*

lished, and their corresponding raw sequence read data made available from public repositories, such as the Sequence Read Archive (Kodama *et al.*, 2012). While several databases devoted to reproductive biology were developed to provide access to relevant microarray-based transcriptomics data (Lee *et al.*, 2010; Hsueh and Rauch, 2012; Luk *et al.*, 2015; Hua *et al.*, 2015; Lardenois *et al.*, 2010; Schuster *et al.*, 2016; Lee *et al.*, 2009; Zhang *et al.*, 2013), until recently no resources were able to manage ultra-high-throughput sequencing data. To deal with these data, in 2015 we introduced the ReproGenomics Viewer (RGV) (Darde *et al.*, 2015), a web-based genomic resource for researchers of the reproductive science community. Its aim was to centralize and offer easy access to the published sequencing datasets (e.g., RNA-seq, ChIP-seq, and MNase-seq) that have accumulated in this field by overcoming the standard technical issues regarding data format, technology, and cross-species comparison. The system was based on implementation of a 'JBrowse genome browser' (Buels *et al.*, 2016). It also included unique features such as the conversion of genome coordinates between species for the direct comparison of data acquired in different organisms. The ReproGenomics Viewer was thus not only a multi- but also a cross-species resource for comparing genomics data.

More recently, the rapid emergence of novel technologies enabling the study of biological systems at a single-cell resolution (Linnarsson and Teichmann, 2016) has again presented serious concerns for the storage, mining, and visualization of such complex datasets (Raja *et al.*, 2017). Several dedicated databases and webservers have been set up to host (Cao *et al.*, 2017; Abugessaisa *et al.*, 2018) and to explore (Lang *et al.*, 2015; DeTomaso and Yosef, 2016; Zhu *et al.*, 2017; Weinreb *et al.*, 2018) single-cell RNA-sequencing (scRNA-seq) datasets.

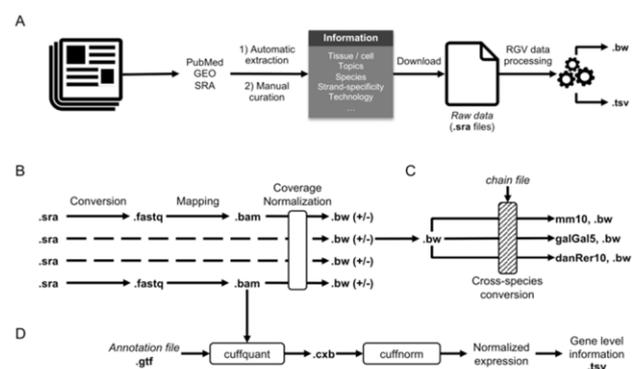
In response to the incoming wave of studies at a single-cell resolution in reproductive biology, we would like to introduce the new release of RGV. Single-cell transcriptome profiling has already contributed substantially to deepening our understanding of reproductive processes in mammals, including gonad development (Stévant *et al.*, 2018; Han *et al.*, 2018), spermatogonial stem cell biology (Li *et al.*, 2017; Guo *et al.*, 2017), and spermatogenesis (Lukassen *et al.*, 2018). The novel architecture of RGV has been totally redesigned and now offers several visualization tools for browsing complex high-dimensional sequencing data including single-cell genomic data. The database content has also grown dramatically and now covers 11 biological topics related to reproduction in 9 species, including 5 landmark scRNA-seq studies. To the best of our knowledge, RGV provides the most comprehensive resource of manually-curated ultra-high-throughput sequencing data in the field of reproduction currently available and thus enables researchers to explore this invaluable source of information.

2 Methods

2.1 Data curation and processing

As previously described (Darde *et al.*, 2015), the backbone of the ReproGenomics Viewer relies on a six-step pipeline to consistently download, curate, organize and process data within the system (Fig. 1, panels A-D). Briefly, *i*) PubMed, SRA, and GEO identifiers assigned to a given study are manually extracted (Fig. 1, panel A). These database entries are then used to mine other information relevant for describing the datasets, such as the species and the technologies. Manual curation is further required to standardize sample names across studies and to gather specific information about, for example, whether or not the RNA sequencing protocol is directional. Next, *ii*) raw data (sra files) are downloaded from

the NCBI SRA repository. The sequencing reads (fastq files) are then *iii*) aligned to the appropriate reference genome with STAR v2.0 (Dobin and Gingeras, 2015) (Fig. 1, panel B). After read mapping *iv*), the deepTools suite (Ramírez *et al.*, 2016) is used to convert the alignment files (bam files) into standard coverage tracks (bigWig files). During this process, coverage values are scaled by using a constant factor corresponding to the maximum number of mapped reads in all samples of a given dataset, divided by the number of mapped reads in the corresponding sample. The next step involves *v*) cross-species conversion, during which the genome coordinates of the scaled coverage tracks are converted to the other reference genomes indexed in the RGV system with CrossMap (Zhao *et al.*, 2014), based on the pairwise whole genome alignment files (chain files) provided by the UCSC genome browser (Kuhn *et al.*, 2013) (Fig. 1, panel C). Finally, and only for transcriptomic datasets (Fig. 1, panel D), *vi*) expression at the gene level is accurately quantified for transcriptome studies by applying the StringTie suite (cuffquant and



cuffnorm tools) (Pertea *et al.*, 2015) on the alignment files (bam files). A principal component analysis is also performed with the FactoMineR package implemented in R (Lê *et al.*, 2008) to project samples onto the first two components.

Fig. 1. The RGV data processing pipeline. **A)** A schematic diagram of the strategy used to organize and process each individual sample in the RGV system from the published datasets. The information used to structure and organize the data – the PubMed, GEO, and SRA IDs – are manually extracted from the publication first before automatic extraction and manual curation (species name, biological topic, technology, samples, etc). Raw data (.sra files) are next downloaded from the NCBI SRA archive and processed (see below, B-D), leading to the generation of two types of files: first, an indexed binary (bigWig or .bw) file for each sample, to enable fast remote access to the data in the genome browser; second, a tabulation-separated value (.tsv) file for each transcriptomic study, used to summarize gene expression levels in the violin and scatter plot visualization tools. **B-D)** A schematic diagram of the RGV data processing workflow. **B)** After download, the SRA Toolkit is used to convert raw data (.sra) files into .fastq files, which are further mapped on their corresponding genome sequences with STAR v2.0. Next, alignment (.bam) files are converted into standard coverage (.bw) files with the deepTools suite, with a scaling factor used to normalize the coverage of all samples within a given study. When strand information is available, two .bw files per sample (+/-) are generated, i.e. one for each strand of the chromosomes. **C)** A cross-species conversion is performed next, with the CrossMap tool and the chain files from the UCSC genome browser. This process allows the visualization of data from any species on a selected reference genome (human, mouse, chicken, and zebrafish in the current version). **D)** Specifically, for transcriptomic data, expression is quantified at the gene level with the StringTie suite on individual .bam files, by providing the reference annotation GTF file and using the cuffquant and cuffnorm tools. This procedure generates the gene expression matrix (.tsv file) suitable for the violin plot and the scatter plot visualization tools.

2.2 Data storage, management, and retrieval

Article short title

The ReproGenomics Viewer database is based on MongoDB (<https://www.mongodb.com/>), a free, open-source, and cross-platform document-oriented database program. This NoSQL database technology provides relevant features for RGV, such as flexible storage of massive and rapidly changing types of data, data replication, and JavaScript compatibility.

2.3 Web interface

The web interface is implemented using two web frameworks, Pyramid (<https://trypyramid.com/>) and AngularJS (<https://angularjs.org/>). Pyramid embeds many features, such as a REST API, a JSON renderer, and compatibility with SMTP servers. AngularJS, on the other hand, is an open JavaScript framework that extends traditional HTML vocabulary; it allows implementation of readable and quickly developable web environments. To handle website traffic and to provide data security, scalability, and deployment, every component of RGV (including website server, MongoDB database, and Elasticsearch server) are hosted on separate individual virtual machines, with Docker (<https://www.docker.com>) as a container system.

2.4 Visualization tools

The ReproGenomics Viewer provides three ways to display datasets, including a genome browser and two gene-level visualization tools: the violin plot and the scatter plot.

The **genome browser** offers a full-featured and highly flexible display for genomic annotations and data via a high-performance, dynamic web interface powered by the implementation of a JBrowse web server (Buels *et al.*, 2016). This system is fully compatible with a large spectrum of data types (fasta, gff, bam, bedGraph, wig, bigWig). The modular design of JBrowse made it possible to implement three plugins to add new features to the system: *i)* *multibigwig* (<https://github.com/elsiklab/multibigwig>), which allows users to plot multiple samples from the same dataset on a single track, thereby facilitating data visualization; *ii)* *Bookmarks-JBrowse* (<https://github.com/awilkey/bookmarks-jbrowse/>), which enables them to save selected tracks and options into a list of bookmarks and to add descriptions and/or comments; *iii)* *Screen Shot* (<https://github.com/bhofmei/jbplugin-screenshot>), which lets users save a high-quality screenshot by using a dedicated dialog box with options including file format, size, quality, and track configuration.

The two gene-level visualization tools were built with Plotly.js, a high-level, declarative charting library based on the famous web graphical library D3.js. Like box plots, the **violin plot** is a method for representing the distribution of quantitative data (here, gene expression) across several groups of samples by using quartiles, but it also features a kernel density estimation (probability density) of the underlying distribution. The **scatter plot** enables exploration of complex datasets and the efficient visualization of a single gene's expression pattern following the projection of all samples in a lower-dimensional space.

3 Results**3.1 Overview and current content**

The ReproGenomics Viewer website has a minimalist and user-friendly web interface. Its responsive design meets all modern web standards and enables users to display RGV on all their devices (desk-

tops, tablets, and smartphones). Briefly, the home page and the top navigation bar offer rapid access to the visualization tools that allow scientists to explore all sequencing datasets listed in the 'Studies' tab. In the past few months, the RGV database content has grown qualitatively and quantitatively through the numerous requests from researchers of the reproductive science community. As of July 2018, the system embeds 2'970 samples from 79 published studies that were carefully curated before data processing and integration (Supplementary Table 1). Altogether these studies cover 11 biological topics related to the biology of reproduction in 9 species and involve 14 "-omics" technologies, including ATAC-seq and single-cell RNaseq (Table 1).

Table 1. RGV content evolution.

	2015	2018
Number of studies	24	79
Number of samples	274	2970
Number of tissues/cells	26	64
Number of biological topics	2	11
Number of technologies	5	14

3.2 The genome browser tool: interactive exploration of large genomic datasets

The genome browser from RGV offers fluid and intuitive genomic navigation among chromosomes of nine species and thus allows multiple users to process data simultaneously. All datasets are formatted and configured for visual display at varying levels of resolution; users can pan and zoom efficiently over a genomic sequence region and simply turn genomics tracks on or off. The background of each track name is color-coded so that samples belonging to the same experimental condition share a similar color.

The full set of available tracks is listed in the faceted track selector tool that enables dynamic queries of RGV's large datasets by the successive application of a series of intersecting filters. This is made possible by the indexation of all study metadata into a CSV file, which facilitates access to information related to biological topics, tissues, age, and technologies (Fig. 2, panel A).

As previously described, one key feature of the RGV data processing workflow is the use of pair-wise alignment files provided by UCSC to perform cross-species conversion of genome coordinates. This allows the direct comparison of experiments in different organisms and thus makes the genome browser not only a multi-species but also a cross-species tool for comparing reproductive genomics data. In addition, the improved processing workflow considers the directionality of the library preparation kits; it can therefore display genomic information associated with the forward and reverse strands on two independent tracks per sample (one for each strand) when applicable (Fig. 2, panel B).

Importantly, the genome browser of RGV also integrates genetic variants known to be associated with reproductive disorders and fertility issues from the ClinVar database (Landrum *et al.*, 2018), as well as the genomic distribution of sequence conservation scores (PhastCons score) provided by the UCSC genome browser (Kuhn *et al.*, 2013).

It is also noteworthy that users can directly upload their individual coverage (bigWig) and annotation (gtf, gff3, and vcf) files into the genome browser by using the option 'Open' in the 'File' tab. These uploaded data are accessible only during the user session and remain totally private.

TA. Darde et al.

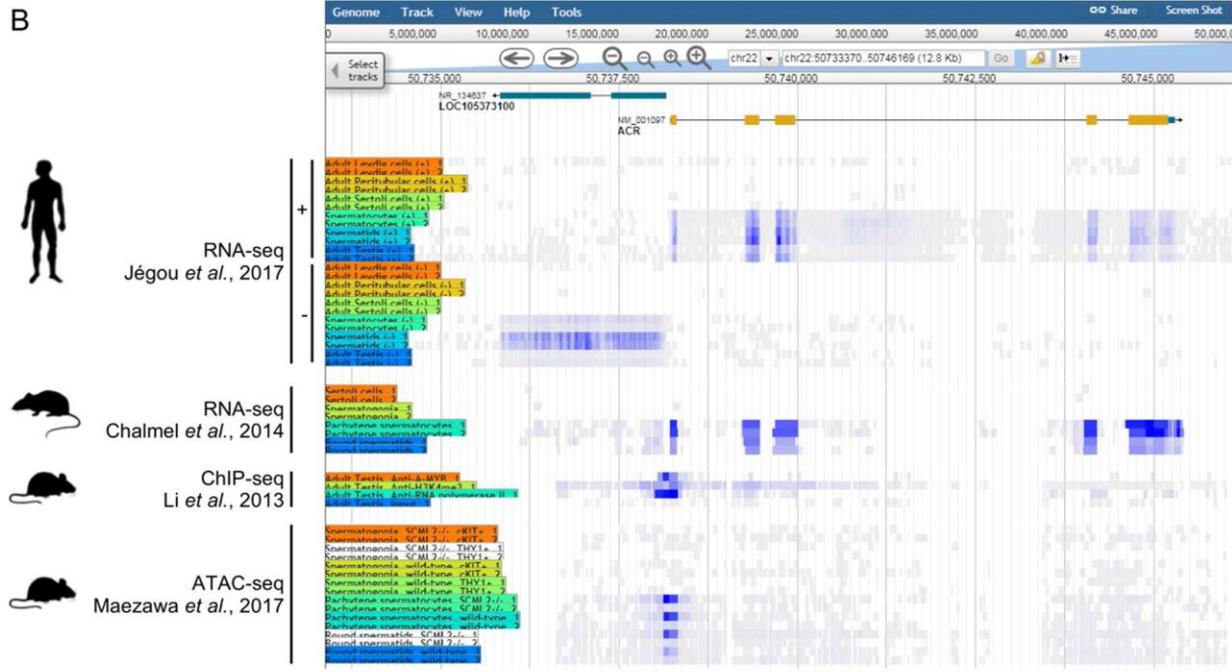
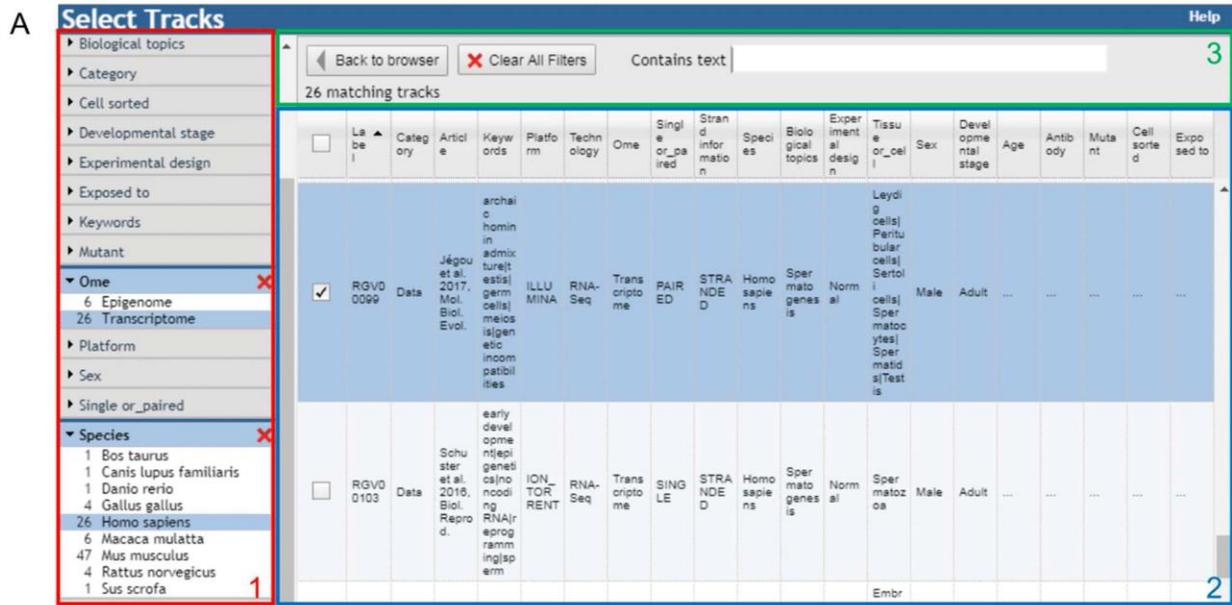


Fig. 2. The RGV genome browser. A) *JBrowse faceted track selector*. The track(s) to be displayed can be selected by applying successive filters to the track metadata. Data can be searched using (1) the topics-related panel and/or can be queried with (2) the text search engine. (3) Studies and data matching the search criteria are displayed in the track list panel for further selection. **B) *Genome browser overview*.** A screen capture from the RGV genome browser, showing the genomic environment of the human acrosin gene (ACR; Gene ID: 49), is presented together with the RefSeq annotation (green and orange boxes correspond to untranslated and coding exons, respectively, and the dark line represents introns) in the human genome (hg38). The RNA-seq analysis of human testicular cells from Jégou *et al.* was selected (Jégou *et al.*, 2017), showing the expression of ACR from the plus strand (“+” tracks) in spermatocytes and spermatids, and that of the LOC105373100 lncRNA (Gene ID: 105373100) from the minus strand (“-” tracks) in spermatids. To illustrate the cross-species and cross-technology capabilities of RGV, three studies in rats and mice were also selected. The RNA-seq data of rat testicular cells demonstrate the conserved expression of ACR in spermatocytes and spermatids (Chalmel *et al.*, 2014), while ChIP-seq data for A-MYB, H3K4me3, and RNA polymerase II in the adult mouse testis (Li *et al.*, 2013) and ATAC-seq data in wild-type and Scml2^{-/-} mouse spermatogonia, spermatocytes, and spermatids (Maezawa *et al.*, 2018) suggest that A-MYB plays a role, independent of SCML2, in regulating ACR expression in meiotic and post-meiotic germ cells by binding to its proximal promoter.

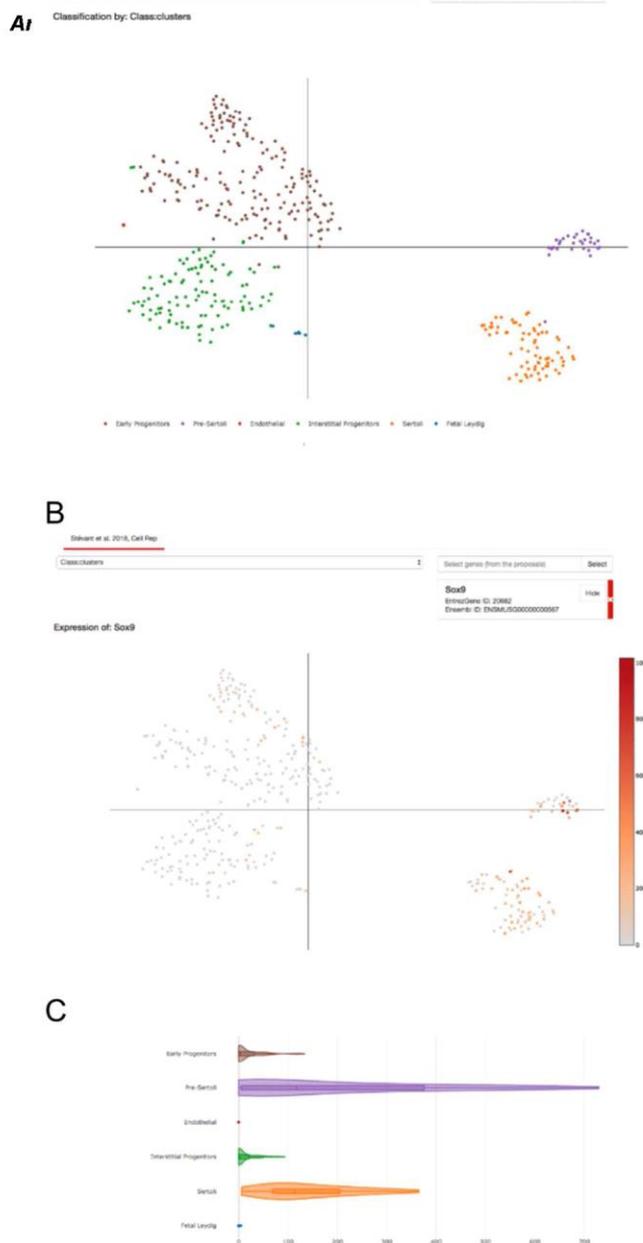


Fig. 3. The RGV visualization tools. Large datasets that may contain hundreds to thousands of cells or samples can be browsed and investigated through both the scatter plot (A and B) and the violin plot (C) visualization tools, which are also compatible with the bulk transcriptomic datasets hosted in RGV. The single-cell RNA sequencing analysis of mouse *sf1*+ fetal testicular cells is taken as an example (Stévant *et al.*, 2018). **A)** Cells are displayed according to the T-distributed Stochastic Neighbor Embedding (t-SNE) coordinates as originally published and can be color-coded according to different classes of cells defined by the authors (cell types, developmental stages, etc). **B)** The expression levels of selected genes (*Sox9* in this example) can then be displayed within each cell, which are color-coded according to the scale bar. **C)** Finally, the expression of candidate genes (*Sox9* again in this example) within classes of cells (cell types, developmental stages, etc) can be summarized with violin plots; kernel density estimation makes it possible to display the full distribution of expression values.

3.3 Gene-level visualization tools: towards shifting paradigms in reproductive biology with single-cell sequencing

The current release of RGV embeds two new visualization tools: the violin plot and the scatter plot. Both are part of the current graphical arsenal used by the “single-cell” community for visualizing the structure of high-dimensional data (Lang *et al.*, 2015; DeTomaso and Yosef, 2016; Cao *et al.*, 2017; Weinreb *et al.*, 2018).

These tools allow users to display quantitative (e.g. gene expression or chromatin accessibility) or qualitative (e.g. copy number variations) information in extremely large datasets that may contain several thousand samples or cells.

For this purpose, the scatter plot (and the violin plot to a lesser extent) relies especially on dimensionality reduction methods such as the t-distributed Stochastic Neighbor Embedding approach (t-SNE) (Van Der Maaten and Hinton, 2008). As of July 2018, five scRNA-seq datasets relevant to male and female gonad development (Han *et al.*, 2018; Stévant *et al.*, 2018), primordial germ and adult spermatogonial stem cells (Li *et al.*, 2017; Guo *et al.*, 2017), and spermatogenesis (Lukassen *et al.*, 2018) have already been indexed in RGV, and other unpublished studies are underway in the process of being submitted (Stévant *et al.*, in preparation; Lardenois *et al.*, in preparation). Of note is that although both the scatter plot and the violin plot were initially designed for single-cell studies, they are also fully compatible with and available for all “bulk” RNA-seq studies hosted in RGV.

Studies of interest can be selected through a dataset selector in the form of a table to which users can apply successive criteria (Fig. 3, panel A). In the scatter plots, the spatial distribution of all cells in a selected study is, by default, color-coded according to the subpopulation classes to which they belong (as published in the corresponding study) (Fig. 3, panel B). Users are then invited to select one or more genes of interest so that each individual cell is color-coded according to its relative expression level (Fig. 3, panel C). This feature could thus help users to identify novel markers specific for particular cell subpopulations, or simply to query their favorite genes.

3.4 RGV documentation and data download

The website provides a ‘Tutorial’ section containing several media types to allow researchers to become rapidly able to use the ReproGenomics Viewer. Moreover, all scripts and data generated or used are freely available through the ‘Download’ page and the GitHub repository (<https://github.com/fchalmel/RGV>). The genome browser tool implemented in RGV lets users download each available individual track by choosing a track of interest and then clicking on ‘Save track data’.

4 Conclusion and perspectives

We report a new version of the ReproGenomics Viewer resource intended to contribute to the reproductive science community by facilitating the centralization and harmonization of the sequencing data we have together accumulated. The system includes a unique workflow for sequencing data management, curation, processing, organization, and comparison across species. With the recent democratization of single-cell genomics, the number of these studies will increase rapidly in the near future. To deal efficiently with this new information, we have complemented existing features with two novel visualization tools, the scatter plot and the violin plot, which enable a more in-depth investigation of these highly complex datasets. This logical transition from “bulk” to single-cell approaches is also accompanied by additional major changes to the system, including a new user-friendly and powerful project selector that makes dataset retrieval and mining easier. Since the

original publication (Darde *et al.*, 2015), the database content has also increased dramatically – there are now 12 times as many samples as in 2015, covering four times as many biological topics related to reproductive biology. To the best of our knowledge, the ReproGenomics Viewer is the most comprehensive and the only cross-species resource of sequencing data in this field.

We plan to keep improving the features and possibilities already offered by the ReproGenomics Viewer. For instance, in the short term, a lasso select tool will be added to the scatter plot visualization tool to generate additional plots based on manually-defined cell subpopulations. A false-color heatmap visualization tool will also be implemented to display the expression pattern of several selected genes within the same graphical representation. In the mid-term, we will allow the creation of private sessions so that users can upload and display their own processed data in the RGV system. Finally, in the long term, we intend to develop a social community toolbox in RGV that we hope will facilitate and stimulate collaborative work in our research field.

It goes without saying that we also intend to gather even more datasets from a wide range of species to cover additional biological topics relevant to reproductive biology. In our continuous effort to help researchers make the most of the technological breakthroughs in the life sciences, we also intend to make RGV compatible with other technologies, such as proteomics and bisulfite sequencing experiments. Other genetic information related to reproductive disorders (such as GWAS and Quantitative trait loci information from diverse sources and diverse model organisms) will also be added to the genome browser tool, as we already plan to integrate data from the GWAS catalog database (Welter *et al.*, 2014).

To keep the resource up-to-date, we strongly encourage scientists to request new datasets and types of data to be added to the system. Indeed, in view of the community's demand for direct and easy access to processed data, we think that making raw data available to the community may be insufficient. We also encourage scientists to upload their own reproductomic datasets to RGV before submission for publication. This could create a win-win situation for all parties, since submitted studies would benefit from enhanced visibility, while the community would take advantage of a continually updated resource. In this context we would like to draw attention to an important issue, which does not concern only RGV, but all (omics) databases in life sciences. Researchers often cite the studies hosted by the databases, but they forget to cite the databases themselves. These citations, however, are essential for the teams that develop and maintain such resources.

The ReproGenomics Viewer is a valuable resource for and supported by researchers in the field of reproductive biology. This system may help scientists and clinicians to overcome the data format and species issues so that they can compare their own datasets with relevant published studies. The very positive feedback of scientists, at international conferences notably, encourages us to maintain our effort and to improve RGV further in terms of both its content and its functionalities. The modular design of the RGV could be applied to other biological domains studied in several model organisms. In that sense, we believe that the ReproGenomics Viewer resource can serve as a reference database in the life sciences.

Acknowledgements

The ReproGenomics Viewer database is supported, built, and maintained by the Research Institute for Environmental and Occupational Health (IRSET), the French School of Public Health (EHESP), and the GenOuest Bioinformatics core facility.

We thank the Institut national de la santé et de la recherche médicale (Inserm) and the Université de Rennes 1 for supporting this work.

Funding

This work was supported by the French agency for food and safety [ANSES n° EST-13-081 to F.C.]; the French National Research Agency [ANR n° 16-CE14-0017-02 to F.C.]; the Fondation pour la recherche médicale [FRM n° DBI20131228558 to F.C.], the Swiss National Science Foundation [SNF n° CRS115_171007], the French National Institute of Health and Medical Research (Inserm), the University of Rennes 1 and the French School of Public Health (EHESP).

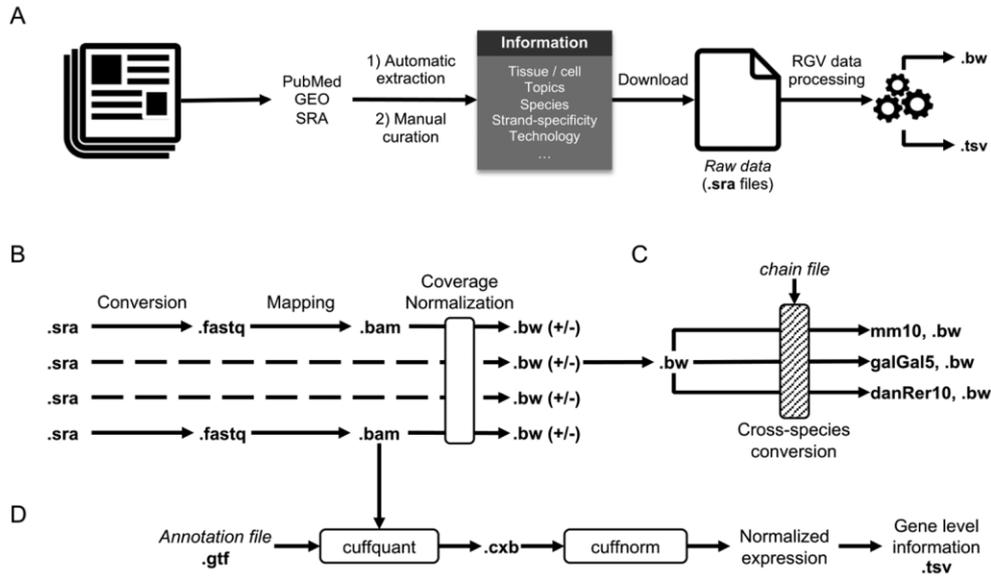
Conflict of interest: none declared.

References

- Abugessaisa, I. *et al.* (2018) SCPortalen: human and mouse single-cell centric database. *Nucleic Acids Res.*, **46**, D781–D787.
- Buels, R. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
- Cao, Y. *et al.* (2017) scRNASeqDB: A Database for RNA-Seq Based Gene Expression Profiles in Human Single Cells. *Genes (Basel)*, **8**.
- Chalmel, F. *et al.* (2014) High-Resolution Profiling of Novel Transcribed Regions During Rat Spermatogenesis1. *Biol. Reprod.*, **91**, 5.
- Darde, T. *et al.* (2015) The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Res.*, **43**, 1–8.
- DeTomaso, D. and Yosef, N. (2016) FastProject: a tool for low-dimensional analysis of single-cell RNA-Seq data. *BMC Bioinformatics*, **17**, 315.
- Dobin, A. and Gingeras, T.R. (2015) Mapping RNA-seq Reads with STAR. In: *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc., Hoboken, NJ, USA, p. 11.14.1–11.14.19.
- Guo, H. *et al.* (2017) DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res.*, **27**, 165–183.
- Han, X. *et al.* (2018) Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*, **172**, 1091–1107.e17.
- Hsueh, A.J. and Rauch, R. (2012) Ovarian Kaleidoscope database: ten years and beyond. *Biol. Reprod.*, **86**, 192.
- Hua, J. *et al.* (2015) Follicle Online: an integrated database of follicle assembly, development and ovulation. *Database*, **2015**, bav036–bav036.
- Jégou, B. *et al.* (2017) Meiotic Genes Are Enriched in Regions of Reduced Archaic Ancestry. *Mol. Biol. Evol.*, **34**, 1974–1980.
- Kodama, Y. *et al.* (2012) The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Res.*, **40**, D54–6.
- Kuhn, R.M. *et al.* (2013) The UCSC genome browser and associated tools. *Brief. Bioinform.*, **14**, 144–161.
- Landrum, M.J. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.
- Lang, S. *et al.* (2015) SCEXV: a webtool for the analysis and visualisation of single cell qRT-PCR data. *BMC Bioinformatics*, **16**, 320.
- Lardenois, A. *et al.* (2010) GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database (Oxford)*, **2010**, baq030.
- Lê, S. *et al.* (2008) FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.*, **25**, 1–18.
- Lee, T.-L. *et al.* (2009) GermsAGE: a comprehensive SAGE database for transcript

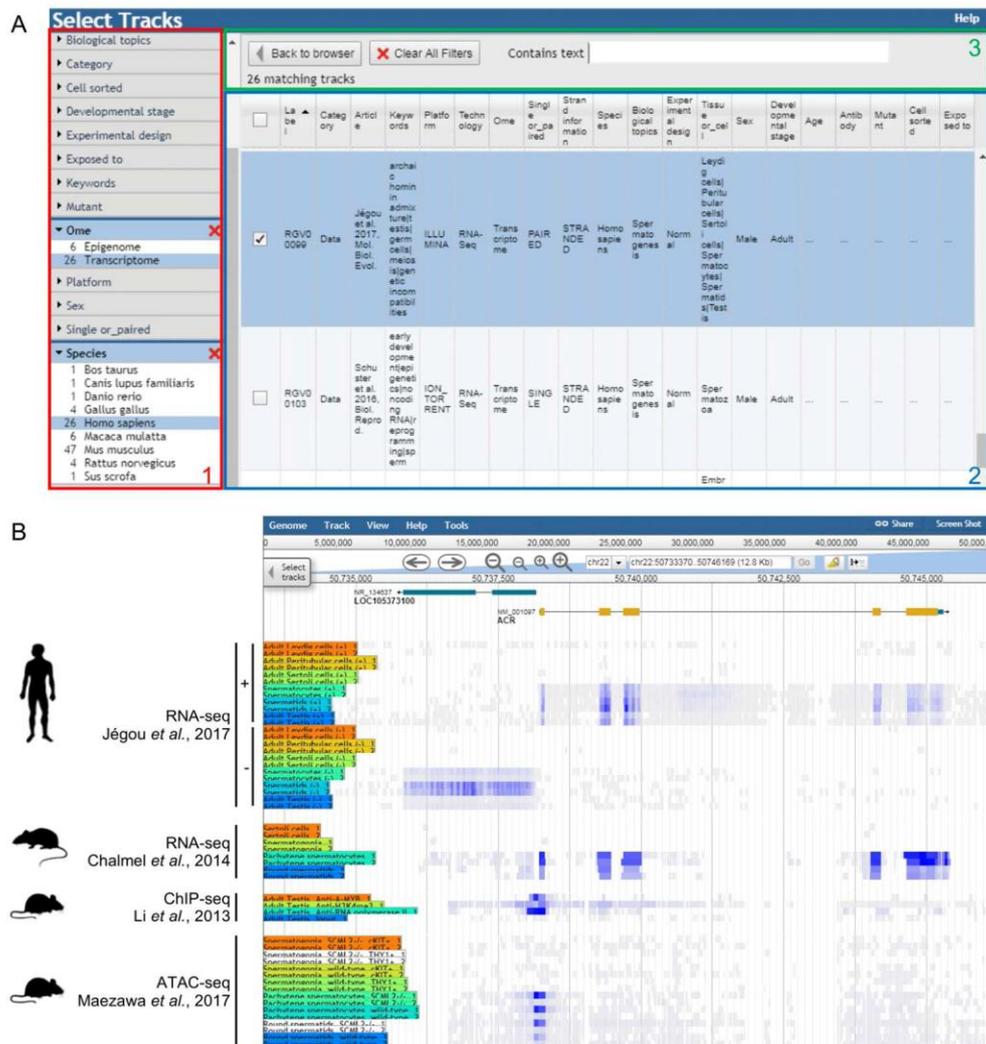
Article short title

- 1
2
3 discovery on male germ cell development. *Nucleic Acids Res.*, **37**, D891-7.
- 4 Lee,T.-L. *et al.* (2010) GonadsAGE: a comprehensive SAGE database for
5 transcript discovery on male embryonic gonad development.
6 *Bioinformatics*, **26**, 585–6.
- 7 Li,L. *et al.* (2017) Single-Cell RNA-Seq Analysis Maps Development of Human
8 Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell*, **20**, 858–
9 873.e4.
- 10 Li,X.Z. *et al.* (2013) An ancient transcription factor initiates the burst of piRNA
11 production during early meiosis in mouse testes. *Mol. Cell*, **50**, 67–81.
- 12 Linnarsson,S. and Teichmann,S.A. (2016) Single-cell genomics: coming of age.
13 *Genome Biol.*, **17**, 97.
- 14 Luk,A.C.-S. *et al.* (2015) GermlncRNA: a unique catalogue of long non-coding
15 RNAs and associated regulations in male germ cell development.
16 *Database (Oxford)*, **2015**, bav044.
- 17 Lukassen,S. *et al.* (2018) Characterization of germ cell differentiation in the male
18 mouse through single-cell RNA sequencing. *Sci. Rep.*, **8**, 6521.
- 19 Van Der Maaten,L. and Hinton,G. (2008) Visualizing Data using t-SNE. *J. Mach.*
20 *Learn. Res.*, **9**, 2579–2605.
- 21 Maezawa,S. *et al.* (2018) Dynamic reorganization of open chromatin underlies
22 diverse transcriptomes during spermatogenesis. *Nucleic Acids Res.*, **46**,
23 593–608.
- 24 Pertea,M. *et al.* (2015) StringTie enables improved reconstruction of a
25 transcriptome from RNA-seq reads. *Nat. Biotechnol.*, **33**, 290–295.
- 26 Raja,K. *et al.* (2017) A Review of Recent Advancement in Integrating Omics Data
27 with Literature Mining towards Biomedical Discoveries. *Int. J. Genomics*,
28 **2017**, 1–10.
- 29 Ramirez,F. *et al.* (2016) deepTools2: a next generation web server for deep-
30 sequencing data analysis. *Nucleic Acids Res.*, **44**, W160-5.
- 31 Schuster,A. *et al.* (2016) SpermBase: A Database for Sperm-Borne RNA Contents.
32 *Biol. Reprod.*, **95**, 99.
- 33 Sims,D. *et al.* (2014) Sequencing depth and coverage: key considerations in
34 genomic analyses. *Nat. Rev. Genet.*, **15**, 121–32.
- 35 Stévant,I. *et al.* (2018) Deciphering Cell Lineage Specification during Male Sex
36 Determination with Single-Cell RNA Sequencing. *Cell Rep.*, **22**, 1589–
37 1599.
- 38 Weinreb,C. *et al.* (2018) SPRING: a kinetic interface for visualizing high
39 dimensional single-cell expression data. *Bioinformatics*, **34**, 1246–1248.
- 40 Welter,D. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-
41 trait associations. *Nucleic Acids Res.*, **42**, D1001-6.
- 42 Zhang,Y. *et al.* (2013) SpermatogenesisOnline 1.0: a resource for spermatogenesis
43 based on manual literature curation and genome-wide data mining. *Nucleic*
44 *Acids Res.*, **41**, D1055-62.
- 45 Zhao,H. *et al.* (2014) CrossMap: a versatile tool for coordinate conversion between
46 genome assemblies. *Bioinformatics*, **30**, 1006–7.
- 47 Zhu,X. *et al.* (2017) Granatum: a graphical single-cell RNA-Seq analysis pipeline
48 for genomics scientists. *Genome Med.*, **9**, 108.
- 49
50
51
52
53
54
55
56
57
58
59
60



The RGV data processing pipeline. A) A schematic diagram of the strategy used to organize and process each individual sample in the RGV system from the published datasets. The information used to structure and organize the data – the Pubmed, GEO, and SRA IDs – are manually extracted from the publication first before automatic extraction and manual curation (species name, biological topic, technology, samples, etc). Raw data (.sra files) are next downloaded from the NCBI SRA archive and processed (see below, B-D), leading to the generation of two types of files: first, an indexed binary (bigWig or .bw) file for each sample, to enable fast remote access to the data in the genome browser; second, a tabulation-separated value (.tsv) file for each transcriptomic study, used to summarize gene expression levels in the violin and scatter plot visualization tools. B-D) A schematic diagram of the RGV data processing workflow. B) After download, the SRA Toolkit is used to convert raw data (.sra) files into .fastq files, which are further mapped on their corresponding genome sequences with STAR v2.0. Next, alignment (.bam) files are converted into standard coverage (.bw) files with the deepTools suite, with a scaling factor used to normalize the coverage of all sam-ples within a given study. When strand information is available, two .bw files per sample (+/-) are generated, i.e. one for each strand of the chromosomes. C) A cross-species conversion is performed next, with the CrossMap tool and the chain files from the UCSC genome browser. This process allows the visualization of data from any species on a selected reference genome (human, mouse, chicken, and zebrafish in the current version). D) Specifically, for transcriptomic data, expres-sion is quantified at the gene level with the StringTie suite on individual .bam files, by providing the reference annotation GTF file and using the cuffquant and cuffnorm tools. This procedure generates the gene expression matrix (.tsv file) suitable for the violin plot and the scatter plot visualization tools.

493x296mm (72 x 72 DPI)



The RGV genome browser. A) JBrowse faceted track selector. The track(s) to be displayed can be selected by applying successive filters to the track metadata. Data can be searched using (1) the topics-related panel and/or can be queried with (2) the text search engine. (3) Studies and data matching the search criteria are displayed in the track list panel for further selection. B) Genome browser overview. A screen capture from the RGV genome browser, showing the genomic environment of the human acrosin gene (ACR; Gene ID: 49), is presented together with the RefSeq annotation (green and orange boxes correspond to untranslated and coding exons, respectively, and the dark line represents introns) in the human genome (hg38). The RNA-seq analysis of human testicular cells from Jégou et al. was selected (Jégou et al., 2017), showing the expression of ACR from the plus strand ("+" tracks) in spermatocytes and spermatids, and that of the LOC105373100 lncRNA (Gene ID: 105373100) from the minus strand ("-" tracks) in spermatids. To illustrate the cross-species and cross-technology capabilities of RGV, three studies in rats and mice were also selected. The RNA-seq data of rat testicular cells demonstrate the conserved expression of ACR in spermatocytes and spermatids (Chaimel et al., 2014), while ChIP-seq data for A-MYB, H3K4me3, and RNA polymerase II in the adult mouse testis (Li et al., 2013) and ATAC-seq data in wild-type and Scml2^{-/-} mouse spermatogonia, spermatocytes, and spermatids (Maezawa et al., 2018) suggest that A-MYB plays a role, independent of SCML2, in regulating ACR expression in meiotic and post-meiotic germ cells by binding

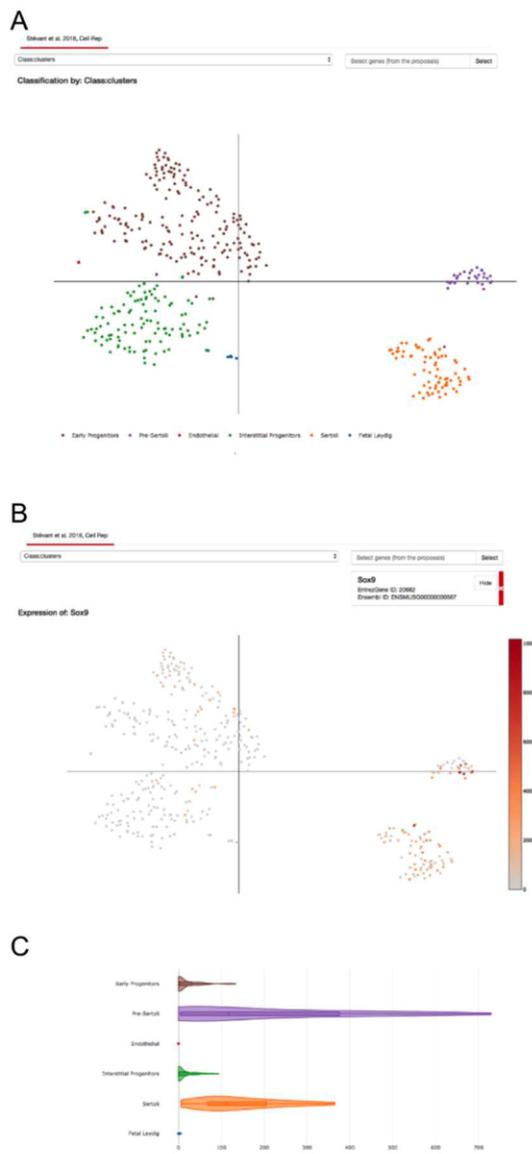
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

to its proximal promoter.

494x555mm (72 x 72 DPI)

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



The RGV visualization tools. Large datasets that may contain hundreds to thousands of cells or samples can be browsed and investigated through both the scatter plot (A and B) and the violin plot (C) visualization tools, which are also compatible with the bulk transcriptomic datasets hosted in RGV. The single-cell RNA sequencing analysis of mouse sf1+ fetal testicular cells is taken as an example (Stévant et al., 2018). A) Cells are displayed according to the T-distributed Stochastic Neighbor Embedding (t-SNE) coordinates as originally published and can be color-coded according to different classes of cells defined by the authors (cell types, developmental stages, etc). B) The expression levels of selected genes (Sox9 in this example) can then be displayed within each cell, which are color-coded according to the scale bar. C) Finally, the expression of candidate genes (Sox9 again in this example) within classes of cells (cell types, developmental stages, etc) can be summarized with violin plots; kernel density estimation makes it possible to display the full distribution of expression values.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

	2015	2018
n° Studies	24	79
n° Samples	274	2970
n° Tissues/Cells	26	64
n° Biological topics	2	11
n° Technologies	5	14

For Peer Review

4.3. Résultats

4.3.1. Description de la première version de RGV

La première version de RGV intégrait une trentaine de projets principalement centrés sur l'étude la spermatogenèse et de la lignée germinale mâle par des approches de transcriptomique ou d'épigénomique. Les données correspondaient aux résultats des analyses fournies par les auteurs (format .bed), induisant donc un biais pour la comparaison des résultats entre les projets, notamment par les outils et les références utilisés pour l'identification et la quantification des différents phénomènes observés. De plus, l'information du sens des transcrits était perdue lors de l'intégration, ce qui empêchait de distinguer les gènes situés en antisens d'un gène connu. La procédure de conversion des données reposait sur les fichiers d'alignements entre les différentes versions d'un même génome et entre les génomes d'espèces différentes (fichiers .chain fournis par UCSC), permettant de convertir les coordonnées des régions génomiques entre les espèces.

4.3.2. La nouvelle stratégie d'intégration

Dans cette nouvelle version de l'outil, les données brutes de séquençage ont été téléchargées à partir de la banque Sequence Read Archive (SRA, Leinonen et al., 2011). J'ai notamment contribué au développement d'une procédure automatisée permettant de télécharger automatiquement les données, de convertir les fichiers téléchargés, d'aligner les lectures sur le génome enfin de quantifier les gènes sur la base de ces alignements. Celle-ci s'adapte en fonction du type de données (RNA-seq, ChIP-seq, scRNA-seq ...) mais également en fonction du protocole de séquençage (single-end ou paired-end, brin spécifique ou non...). L'uniformisation totale du traitement des données de séquençage permet de standardiser la procédure pour toutes les études et ainsi de faciliter la comparaison entre les études. Par exemple, les études effectuées chez une même espèce seront alignées sur la même version de référence, ce qui homogénéise les résultats, indépendamment de l'année de publication. Contrairement à la stratégie initiale développée dans RGV, qui incluait la conversion des données vers les 13 autres génomes disponibles, coûteuses en temps et en ressources informatiques, les organismes modèles sont maintenant privilégiés. Ainsi, indépendamment de l'organisme étudié, toutes les données sont converties sur le génome humain, souris, poisson-zèbre et rat.

4.3.3. Les nouvelles études intégrées

Aujourd'hui, plus de 79 études sont disponibles dans RGV et couvrent un large spectre de données « omiques » (transcriptomique, génomique, cistromique et épigénomique) relatives à la reproduction (développement des gonades, spermatogenèse et l'ovogenèse) (voir Tableau II). Cela

représente environ 2500 échantillons de 9 organismes différents, répertoriés dans le tableau II. Parmi ces nouvelles études, cinq utilisent notamment du scRNA-seq afin de caractériser différents aspects de la reproduction à l'échelle de la cellule unique (voir «

RNA-seq à l'échelle de la cellule unique, scRNA-seq »).

4.3.4. Les nouveaux outils de visualisation

La visualisation grâce au navigateur de génome est toujours disponible et a été améliorée afin de fournir un accès toujours plus intuitif aux données. Deux nouveaux outils ont vu le jour permettant une visualisation quantitative de l'expression des gènes au sein des projets de transcriptomique. D'une part, le diagramme en violon (*violin plot*) permet de visualiser l'expression d'un gène à travers les échantillons d'une étude de manière rapide et intuitive ainsi que d'estimer la variabilité entre les réplicats. D'autre part, les projets de scRNA-seq ont également un outil dédié à la visualisation de leurs résultats. La représentation de données aussi complexes nécessite la réduction de leur dimensionnalité, par une approche communément utilisée par la communauté, le t-distributed Stochastic Neighbor Embedding approach (t-SNE) (Maaten and Hinton, 2008). Cela permet la représentation des échantillons (les cellules) en nuage de point (*scatter plot*). Plusieurs paramètres peuvent caractériser chaque cellule : son échantillon d'origine (duquel elle a été capturée), la partition à laquelle elle appartient (fournie par les auteurs suite à leurs propres analyses), ou le niveau d'expression d'un gène sélectionné par l'utilisateur.

4.4. Discussion et Perspectives

RGV est un outil de visualisation et de comparaison des études haut débit, mais n'a pas vocation à interpréter les résultats obtenus ou de faire des conjectures sur la conservation fonctionnelle des gènes entre les espèces. La procédure de conversion des données entre les espèces repose sur le principe de synténie des régions génomiques entre les génomes, mais ne présume pas de leur conservation de fonction.

Il a été démontré que la source de l'annotation des génomes et transcriptomes, RefSeq, Ensembl ou UCSC, pouvait grandement influencer les résultats de l'analyse bio-informatique de RNA-seq (Frankish et al., 2015). Pour surpasser ce biais, l'approche utilisée dans notre équipe depuis l'initiation des projets de RNA-seq est de fusionner les annotations de différentes banques pour créer une référence non redondante d'annotation des transcrits. Cela permet de prendre en compte les résultats des différentes méthodes de *curation* des banques, et ainsi utiliser la majorité des connaissances actuelles en génomique. Une procédure automatisée a été mise en place pour télécharger les génomes, les indexer par les différents outils communément utilisés dans nos analyses

(STAR, Dobin et al., 2013; Cufflinks, Trapnell et al., 2012; StringTie, Pertea et al., 2015), télécharger les annotations de Ensembl, RefSeq et UCSC, et créer les fichiers non redondants qui seront utilisés dans RGV et autres projets de l'équipe (dont les deux projets présentés dans ce manuscrit). De plus, l'uniformisation des références utilisées pour l'analyse permet une comparaison plus pertinente des résultats obtenus dans une même espèce. En effet, même si les technologies évoluent, que de nouveaux jeux de données sont régulièrement générés, les données déjà disponibles peuvent toujours révéler de nouveaux phénomènes grâce à l'évolution de l'annotation des génomes.

Les nouveaux outils de visualisation permettent un accès toujours plus facilité des utilisateurs aux données d'intérêt, que ce soit les échantillons ou la quantification des gènes qui les intéressent. En particulier, le *scatter plot* est particulièrement adapté aux nouvelles études de scRNA-seq, et produit une visualisation pertinente de tels jeux de données. Au vu de l'engouement de la communauté scientifique pour ces nouvelles technologies de séquençage, il est opportun d'adapter RGV et d'anticiper les besoins des chercheurs pour la visualisation pertinente de ces jeux de données complexes. L'intégration et l'analyse automatisée des futurs projets haut-débit permettront une mise à jour facilitée du contenu et ainsi rendre rapidement disponible les résultats des études les plus récentes à la communauté. À l'avenir, l'implémentation d'autres outils de visualisation viendra étoffer ceux déjà disponibles dans RGV, comme le *heatmap*, afin de faciliter toujours plus l'accès des données aux utilisateurs en fonction de leurs habitudes.

Certaines technologies d'épigénomique utilisent le principe de la conversion sélective des cytosines méthylées en uracil pour déterminer le profil de méthylation du génome entier, comme le Bisulfite-seq ou le RRBS (Li and Tollefsbol, 2011; Meissner et al., 2005). De fait, la mise en place de l'analyse automatisée de ces données requiert une attention particulière au niveau de l'optimisation du traitement et de l'analyse bio-informatique. En transcriptomique, les études de smallRNA-seq n'ont pas encore été intégrées. Cette technologie permet de séquencer les petits ARN (<200 nt) qui peuvent être impliqués dans de nombreux processus de régulation au sein d'une cellule (voir « Les petits ARN non codants »). En reproduction, il a déjà été observé que les piRNA et miRNA ont un rôle important dans la spermatogenèse (Chen et al., 2017b; Yadav and Kotaja, 2014) et plus globalement lors de la différenciation de la lignée germinale (García-López et al., 2015; Girard et al., 2006; Watanabe et al., 2006). Une procédure particulière devra être développée pour l'analyse pertinente de telles données (Conesa et al., 2016).

De même, l'une des perspectives pour l'intégration toujours plus exhaustive des études haut-débit nécessitera une approche adaptée aux puces à ADN. Bien que ces technologies soient de moins en moins plébiscitées de nos jours, elles sont encore très utilisées en raison de leur fiabilité, leur robustesse et leur coût (humain) réduit. De fait, les puces sont particulièrement indiquées et utilisées en toxicogénomique, permettant l'analyse de nombreux échantillons, et donc des protocoles plus

larges et complexes (Bourdon-Lacombe et al., 2015). Elles sont utilisées dans des projets de grande envergure tels que DrugMatrix (Ganter et al., 2006), Open TG-GATES (Igarashi et al., 2015) et Connectivity Map (Lamb et al., 2006). Ces bases de données contiennent les réponses transcriptionnelles de milliers d'échantillons biologiques (en général, des organes de rat ou des lignées cellulaires), exposés à des milliers de composés chimiques à différentes doses et différents temps d'exposition. On peut notamment retrouver des tests effectués sur ovaire et testicule, soumis à de nombreux composés, y compris des perturbateurs endocriniens. D'une part, il faudra donc adapter les protocoles d'analyses et d'intégration de RGV à ces technologies, mais également les outils de visualisation. D'autre part, la *curation* manuelle qui permet d'homogénéiser les metadata (description des protocoles expérimentaux, i.e. nom des échantillons), au niveau des tissus, âges ou technologies, est un travail chronophage, mais est indispensable pour une navigation intuitive et facilitée des utilisateurs à travers la base de données RGV. Concernant la toxicogénomique, il faudra développer un autre vocabulaire contrôlé permettant de décrire de la manière la plus uniforme et pertinente les protocoles toxicologiques. Des metadata homogènes permettent une utilisation simplifiée et une recherche rapide et complète au sein des différents échantillons présents dans RGV, ce qui permet une meilleure utilisation au quotidien.

Au même titre que la toxicogénomique, les susceptibilités génétiques et les mutations associées aux pathologies liées à la reproduction seront intégrées à RGV. Aujourd'hui, très peu de bases de données sont disponibles quant aux variations génétiques et nucléotidiques associées à des phénotypes ou pathologies. On recense actuellement des initiatives telles que ClinVar (Landrum et al., 2016) et GWAS Catalog (MacArthur et al., 2017), qui repose sur la soumission des données par les chercheurs, et qui restent donc limitées. Bien que le partage des données soit communément accepté dans de nombreux domaines de recherche, la recherche clinique reste l'une des dernières disciplines réticentes. À cause de la réglementation stricte sur les données de patients, les règles éthiques et le climat de grande compétitivité qui règne dans ce domaine, il est difficile d'accéder à de tels résultats, qui pourraient être également utiles à la recherche fondamentale et appliquée sur la reproduction.

Article	PubMedID	Sujet	Technology	Espèce	Echantillons	Nb d'éch
The ENCODE consortium et al. 2004, Science	15499007	Analyse de tissu	ChIP-Seq	Souris	Testicule	9
			RNA-Seq	Souris	Ovaire, Testicule	11
Brykczynska et al. 2010, Nat. Struct. Mol. Biol.	20473313	Spermatogenèse	MNase-Seq	Homme	Spermatozoïdes	7
Bernstein et al. 2010, Nat. Biotechnol.	20944595	Analyse de tissu	ChIP-Seq	Homme	Ovaire	1
			DNase-Hypersensitivity	Homme	Ovaire, Testicule	4
			MeDIP-Seq	Rat	Spermatozoïdes	1
			MRE-Seq	Rat	Spermatozoïdes	1
			RNA-Seq	Homme	Ovaire, Testicule	3
Smagulova et al. 2011, Nature	21460839	Spermatogenèse	ChIP-Seq	Souris	Foie, Testicule	18
			MNase-Seq	Souris	Testicule	2
Cabili et al. 2011, Genes Dev.	21890647	Analyse de tissu	RNA-Seq	Homme	Cerveau, Foie, Placenta, Testicule, Foreskin fibroblast, Cellules HeLa, Fibroblaste primaire de poumon	16
Tan et al. 2011, Cell	21925322	Spermatogenèse	ChIP-Seq	Homme	Fibroblaste primaire de poumon	3
			ChIP-Seq	Souris	Spermatocytes, Spermatides ronds	6
Brawand et al. 2011, Nature	22012392	Analyse de tissu	RNA-Seq	Poulet	Cerveau, Cervelet, Cœur, Rein, Foie, Testicule	12
				Homme	Cervelet, Cœur, Rein, Cortex frontal, Foie, Cortex préfrontal, Lobe temporal, Testicule	18
				Macaque	Cerveau, Cervelet, Cœur, Rein, Foie, Testicule	13
				Souris	Cerveau, Cervelet, Cœur, Rein, Foie, Testicule	17
Khil et al. 2012, Genome Res.	22367190	Spermatogenèse	ChIP-Seq	Souris	Testicule	11
	22454233	Analyse de tissu	PolyA-Seq	Chien	Cerveau, Rein, Testicule	3

Derti et al. 2012, Genome Res.				Homme	Cerveau, Rein, Foie, Muscle, Testicule, Universal Human Reference	11
				Macaque	Cerveau, Iléon, Rein, Foie, Testicule	5
				Souris	Cerveau, Rein, Foie, Muscle, Testicule	5
				Rat	Cerveau, Testicule	2
Brick et al. 2012, Nature	22660327	Spermatogenèse	ChIP-Seq	Souris	Foie, Testicule	17
Shen et al. 2012, Nature	22763441	Analyse de tissu	ChIP-Seq	Souris	Moelle osseuse, Cervelet, Cortex, Cœur, Intestin, Rein, Foie, Poumon, Bulbe Olfactif, Placenta, Rate, Testicule, Thymus, Fibroblastes embryonnaires, Cellules souches embryonnaires, Cerveau, Membres	122
			RNA-Seq	Souris	Moelle osseuse, Cervelet, Cortex, Cœur, Intestin, Rein, Foie, Poumon, Bulbe Olfactif, Placenta, Rate, Testicule, Thymus, Fibroblastes embryonnaires, Cellules souches embryonnaires, Cerveau, Membres	19
Vourekas et al. 2012, Nat. Struct. Mol. Biol.	22842725	Spermatogenèse	HITS-CLIP	Souris	Testicule	8
			RNA-Seq	Souris	Testicule, Spermatogonia, Spermatoocytes pachytènes, Spermatoïdes ronds, Spermatoïdes allongés	6
Seisenberger et al. 2012, Mol. Cell	23219530	CGP	RNA-Seq	Souris	Cellules germinales primordiales	5
Gkountela et al. 2012, Nat. Cell Biol.	23242216	CGP	RNA-Seq	Homme	Cellules souches embryonnaires, Cellules germinales primordiales	7
Ng et al. 2013, Dev. Cell	23352811	CGP	ChIP-Seq	Souris	Cellules souches embryonnaires, Cellules germinales primordiales	31
Yokobayashi et al. 2013, Nature	23486062	CGP	RNA-Seq	Souris	Cellules germinales primordiales	3
Li et al. 2013, Mol. Cell	23523368	Spermatogenèse	CAGE	Souris	Testicule	2
			ChIP-Seq	Poulet	Testicule	3

				Souris	Testicule	4
			PAS-Seq	Souris	Testicule	1
			RNA-Seq	Souris	Testicule	14
Jiang et al. 2013, Cell	23663777	Développement	MeDIP-Seq	Poisson-zèbre	Embryon, Spermatozoïdes	5
			RNA-Seq	Poisson-zèbre	Spermatozoïdes, Œuf, Embryon	4
Gan et al. 2013, Nat Commun	23759713	Spermatogenèse	hMeDIP-Seq	Souris	Cellules de Sertoli, Spermatogonies type A, Spermatogonies type B, Spermatoocytes leptotènes, Spermatoocytes pachytènes, Spermatoïdes ronds, Spermatoïdes allongés, Spermatozoïdes	10
			RNA-Seq	Souris	Cellules de Sertoli, Spermatogonies type A, Spermatogonies type B, Spermatoocytes leptotènes, Spermatoocytes pachytènes, Spermatoïdes ronds, Spermatoïdes allongés	8
Erkek et al. 2013, Nat. Struct. Mol. Biol.	23770822	Spermatogenèse	ChIP-Seq	Souris	Spermatoïdes ronds, Spermatozoïdes	15
			MNase-Seq	Souris	Spermatozoïdes	3
			RNA-Seq	Souris	Spermatoïdes ronds	2
Soumillon et al. 2013, Cell Rep	23791531	Spermatogenèse, Analyse de tissu	ChIP-Seq	Souris	Cerveau, Foie, Testicule, Spermatoïdes, Spermatoocytes	5
			RNA-Seq	Poulet	Cerveau, Ovaire, Testicule	4
				Homme	Cortex frontal, Ovaire, Testicule, Placenta	11
				Macaque	Cerveau, Testicule	2
				Souris	Cerveau, Foie, Ovaire, Testicule, Placenta, Cellules de Sertoli, Spermatogonies, Spermatoocytes, Spermatoïdes, Spermatozoïdes	17
Smagulova et al. 2013, BMC Genomics	23870400	Spermatogenèse	ChIP-Seq	Souris	Testicule	9
	24043772	CGP, Spermatogenèse	ChIP-Seq	Souris	Cellules germinales primordiales, Spermatoocytes pachytènes, Spermatoïdes ronds	42

Lesch et al. 2013, Proc. Natl. Acad. Sci. U.S.A.			RNA-Seq	Souris	Cellules germinales primordiales, Spermatoocytes pachytènes, Spermatoïdes ronds	20
Nielsen et al. 2014, RNA	24344320	Analyse de tissu	RNA-Seq	Homme	Vessie, Cerveau, Seins, Colon, Cœur, Rein, Foie, Poumon, Muscle, Ovaire, Prostate, Peau	12
Margolin et al. 2014, BMC Genomics	24438502	Spermatogenèse	ChIP-Seq	Souris	Testicule	4
			RNA-Seq	Souris	Testicule	8
Necsulea et al. 2014, Nature	24463510	Analyse de tissu	RNA-Seq	Poulet	Cerveau, Ovaire, Testicule	4
				Homme	Cortex frontal, Ovaire, Testicule, Placenta	11
				Macaque	Cerveau, Testicule	2
				Souris	Cerveau, Ovaire, Testicule, Placenta	9
Meikar et al. 2014, RNA	24554440	Spermatogenèse	RNA-Seq	Souris	Testicule, Spermatoïdes ronds	14
Djureinovic et al. 2014, Mol. Hum. Reprod.	24598113	Analyse de tissu	RNA-Seq	Homme	Tissu adipeux, Glandes surrénales, Appendice, Vessie, Moelle osseuse, Colon, Cortex, Duodenum, Endometrium, Œsophage, Cœur, Rein, Foie, Poumon, Ganglions lymphatiques, Ovaire, Pancreas, Placenta, Prostate, Glandes salivaires, Peau, petit intestin, Rate, Stomach, Testicule, thyroïde	95
Chalmel et al. 2014, Biol. Reprod.	24740603	Spermatogenèse	RNA-Seq	Rat	Cellules de Sertoli, Spermatoïdes, Spermatoocytes pachytènes, Spermatoïdes ronds	8
Wang et al. 2014, Cell	24813617	Développement	RNA-Seq	Souris	Ovocytes (métaphase II)	1
Hammoud et al. 2014, Cell Stem Cell	24835570	Spermatogenèse, Spermatoïdes	ChIP-Seq	Homme	Spermatoïdes	5
				Souris	AGSC Kit, Spermatoïdes, Spermatoocytes, Spermatoïdes	21
			hMeDIP-Seq	Souris	AGSC Kit, Spermatoocytes, Spermatoïdes, Spermatoïdes	4
			MeDIP-Seq	Souris	Sang	1

				Rat	Sang, Cerveau, Spermatozoïdes	3
			MRE-Seq	Souris	Sang	1
				Rat	Sang, Cerveau, Spermatozoïdes	3
			RNA-Seq	Homme	Spermatozoïdes	2
				Souris	Spermatogonies, Spermatoocytes, Spermatides, Spermatozoïdes	13
Molaro et al. 2014, Genes Dev.	25030694	Spermatoocytes, Développement	RNA-Seq	Souris	Crête génitale, Embryon, Spermatoocytes	5
Chocu et al. 2014, Biol. Reprod.	25210130	Spermatogenèse	RNA-Seq	Rat	Cellules de Sertoli, Spermatogonies, Spermatoocytes pachytènes, Spermatides ronds	8
Mu et al. 2014, Genes Dev.	25228648	Spermatogenèse	ChIP-Seq	Souris	Spermatoocytes	7
Yue et al. 2014, Nature	25409824	Analyse de tissu	ChIP-Seq	Souris	Testicule	9
			RNA-Seq	Souris	Ovaire, Testicule	11
Korfanty et al. 2014, Int. J. Biochem. Cell Biol.	25450459	Spermatogenèse	ChIP-Seq	Souris	Spermatoocytes	9
Bao et al. 2015, Development	25503407	Cellules somatiques de la gonade	RNA-Seq	Souris	Testicule	6
Irie et al. 2015, Cell	25543152	CGP	RNA-Seq	Homme	Cellules souches embryonnaires, Cellules somatiques de la gonade, CPG-like, Cellules germinales primordiales, TCam-2 (lignée)	16
Reyes et al. 2015, Mol. Reprod. Dev.	25560149	Ovogenèse	RNA-Seq	Taureau	Vésicules germinales d'ovocytes, Ovocytes (metaphase II)	6
Hasegawa et al. 2015, Dev. Cell	25703348		ChIP-Seq	Souris	Cellules souches de spermatogonies, Spermatoocytes pachytènes	7

			MNase-Seq	Souris	Cellules souches de spermatogonies, Spermatocytes pachytènes	7
			RNA-Seq	Souris	Cellules souches de spermatogonies, Spermatocytes pachytènes, Spermatides ronds, Spermatogonies	14
Zimmermann et al. 2015, Mol. Endocrinol.	25710594	Spermatogenèse, Développement gonadique, Cellules germinales de la gonade	RNA-Seq	Souris	Cellules de Sertoli	16
McClelland et al. 2015, Biol. Reprod.	25855264	Cellules germinales de la gonade	RNA-Seq	Souris	Cellules de Leydig, Cellules de Sertoli, Cellules interstitielles	9
Goh et al. 2015, Genes Dev.	25995188	Spermatogenèse	RNA-Seq	Souris	Spermatocytes pachytènes, Spermatocytes secondaires, Spermatides ronds	23
Gkountela et al. 2015, Cell	26004067	CGP, Ovogenèse	RNA-Seq	Homme	Cellules souches embryonnaires, Cellules germinales primordiales	19
Guo et al. 2015, Cell	26046443	CGP	RNA-Seq	Homme	Cellules germinales primordiales, Cellules somatiques de la gonade	328
Tang et al. 2015, Cell	26046444	CGP	RNA-Seq	Homme	Cellules souches embryonnaires, Cellules germinales primordiales, Cellules somatiques de la gonade, PGC-like	1
Sin et al. 2015, BMC Biol.	26198001		ChIP-Seq	Souris	Cellules souches de spermatogonies, Spermatocytes pachytènes, Spermatides ronds	29
Kubo et al. 2015, BMC Genomics	26290333	Spermatogonies	RNA-Seq	Souris	Spermatogonies primaires, Pro-spermatogonies, Spermatogonies	4
Ayers et al. 2015, BMC Genomics	26377738	Développement	RNA-Seq	Poulet	Embryon, Ovaire, Testicule	12
Soh et al. 2015, PLoS Genet.	26378784	Développement gonadique, Ovogenèse	RNA-Seq	Souris	Tiers antérieur de l'ovaire, Ovaire, Tiers postérieur de l'ovaire, Testicule	20

Gely-Pernot et al. 2015, BMC Genomics	26518232	Spermatogenèse	ChIP-Seq	Souris	Testicule	6
Hammoud et al. 2015, Genes Dev.	26545815	Spermatogonies	ChIP-Seq	Souris	Spermatogonies	6
			RNA-Seq	Souris	Spermatogonies	16
Zhu et al. 2016, Sci Rep	26753906	Spermatogenèse	RNA-Seq	Homme	Cellules de Sertoli, Spermatogonies, Primary spermatocytes, Spermatides	1
da Cruz et al. 2016, BMC Genomics	27094866	Spermatogenèse	RNA-Seq	Souris	Diploid testicular cells, Leptotene and zygotene spermatocytes, Spermatocytes pachytènes, Spermatides ronds	4
Goudarzi et al. 2016, Mol. Cell	27105113	Spermatogenèse	ChIP-Seq	Souris	Spermatocytes, Spermatides ronds	10
Li et al. 2016, Sci Rep	27229484	Spermatogenèse, Développement gonadique	RNA-Seq	Cochon	Testicule	12
Lin et al. 2016, RNA Biol	27560004	Spermatogenèse	RNA-Seq	Souris	Spermatogonial stem cells, Spermatogonies type A, Spermatocytes preleptotène, Spermatocytes pachytènes, Spermatides ronds	12
Schuster et al. 2016, Biol. Reprod.	27628216	Spermatogenèse	RNA-Seq	Homme	Spermatozoïdes	3
				Souris	Sperm head, Spermatozoïdes	4
				Rat	Spermatozoïdes	3
von Meyenn et al. 2016, Dev. Cell	27728778	CGP	RNA-Seq	Homme	Cellules souches embryonnaires, Epiblast-like, PGC-like	11
				Souris	Cellules souches embryonnaires, Epiblast-like, PGC-like	12
Mutoji et al. 2016, Biol. Reprod.	27733379	Spermatogonies	ChIP-Seq	Souris	Spermatogonies	8
			RNA-Seq	Souris	Spermatogonies	9
Guo et al. 2016, Cell Res.	27824029	CGP	RNA-Seq	Homme	Cœur, Cellules germinales primordiales	6
				Souris	Cellules de l'épiblaste, Cellules germinales primordiales, Cellules somatiques de la gonade	18
				scRNA-Seq	Homme	Cellules somatiques de la gonade, Cellules germinales primordiales

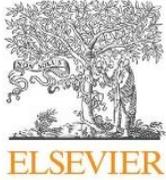
Helsel et al. 2017, Development	28087628	Spermatogonies	RNA-Seq	Souris	Spermatogonial stem cells, Testicule	12
Neuhaus et al. 2017, Mol. Hum. Reprod.	28093458	Spermatogonies	RNA-Seq	Homme	Spermatogonies	2
Zhang et al. 2017, PLoS ONE	28291811	Spermatogenèse	RNA-Seq	Souris	Testicule, Spermatozoïdes	6
Naro et al. 2017, Dev. Cell	28366282	Spermatogenèse	RNA-Seq	Souris	Spermatocytes, Spermatides	5
Li et al. 2017, Oncotarget	28404936	Spermatogonies	RNA-Seq	Souris	Ovogonal stem cells, Spermatogonial stem cells	6
JÃ©gou et al. 2017, Mol. Biol. Evol.	28444387	Spermatogenèse	RNA-Seq	Homme	Cellules de Leydig, Peritubular cells, Cellules de Sertoli, Spermatocytes, Spermatides, Testicule	12
Li et al. 2017, Cell Stem Cell	28457750	CGP	scRNA-Seq	Homme	Cellules germinales primordiales	43
Rahmoun et al. 2017, Nucleic Acids Res.	28472341	Cellules somatiques de la gonade	ChIP-Seq	Taureau	Testicule	3
				Souris	Testicule	3
			RNA-Seq	Souris	Ovaire, Testicule	9
Iwasaki et al. 2017, Methods	28552266	Spermatogenèse	RNA-Seq	Souris	Testicule	2
Hannigan et al. 2017, Cell Rep	28636946	Spermatogenèse	HITS-CLIP	Souris	Testicule	3
			RNA-Seq	Souris	Testicule	4
Jan et al. 2017, Development	28935708	Spermatogenèse	RNA-Seq	Homme	Spermatocytes pachytènes, Leptotene and zygotene spermatocytes, Spermatides ronds, Type A dark spermatogonia, Type A pale spermatogonia	6
Ramaswamy et al. 2017, Hum. Reprod.	28938749	Spermatogonies, Spermatogenèse	RNA-Seq	Macacque	Testicule	10
Guo et al. 2017, Cell Stem Cell	28985528	Spermatogonies	ATAC-Seq	Homme	Spermatogonial stem cells, Cellules souches embryonnaires	6
			RNA-Seq	Homme	Spermatogonial stem cells	18

			scRNA-Seq	Homme	Spermatogonial stem cells	175
Reyes et al. 2017, Hum. Reprod.	29025019	Ovogenèse	RNA-Seq	Homme	Vésicules germinales d'ovocytes, Ovocytes (metaphase II)	20
Chitwood et al. 2017, Biol. Reprod.	29025079	Ovogenèse, Développement, Embryogenèse	RNA-Seq	Macaque	Vesicules germinales d'ovocytes, Metaphase I oocytes, Ovocytes (metaphase II), Embryo, Morula, Blastocyst	22
Wichman et al. 2017, Biol. Reprod.	29044429	Spermatogenèse	RNA-Seq	Souris	Spermatogonies, Spermatocytes pachytènes, Spermatides ronds	9
Maezawa et al. 2017, Nucleic Acids Res.	29126117	Spermatogenèse	ATAC-Seq	Souris	Spermatogonies, Spermatocytes pachytènes, Spermatides ronds	16
Stévant et al. 2018, Cell Rep	29425512	Développement gonadique, Cellules somatiques de la gonade, Determination sexuelle	scRNA-Seq	Souris	Cellules somatiques de la gonade	435
Han et al. 2018, Cell	29474909	Analyse de tissu	Microwell-Seq	Souris	Ovaire, Testicule, Uterus, Placenta	10
Skinner et al. 2018, Epigenetics Chromatin	29482626		ChIP-Seq	Rat	Spermatozoïdes	12
			MeDIP-Seq	Rat	Spermatozoïdes	18
Ben Maamar et al. 2018, Sci Rep	29593303		ChIP-Seq	Rat	Spermatozoïdes	23
Thibault-Sennett et al. 2018, Genetics	29674518		ChIP-Seq	Souris	Spermatogonies GC-1, Testicule	19
			RNA-Seq	Souris	Testicule	5
Lukassen et al. 2018, Sci Rep	29695820	Spermatogenèse	scRNA-Seq	Souris	Testicule	2
	29732173		ChIP-Seq	Rat	Spermatozoïdes	12

Ben Maamar et al. 2018, Environ Epigenet		MeDIP-Seq	Rat	Spermatozoïdes	18
Winge et al. 2018, Cell Death Dis	29789566	RNA-Seq	Homme	Testicule	22
Zhang et al. 2018, Development	29848638	RNA-Seq	Souris	Spermatides ronds	4

Tableau II. Ensemble des études intégrées dans la nouvelle version de RGV.

5. Revue de la littérature



New transcriptomic tools to understand testis development and functions

Estelle Lecluze, Bernard Jégou, Antoine D. Rolland¹, Frédéric Chalmel^{*,1}

Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, Environnement et travail) - UMR_S1085, F-35000 Rennes, France

ARTICLE INFO

Article history:

Received 10 November 2017
Received in revised form
26 February 2018
Accepted 27 February 2018
Available online 1 March 2018

Keywords:

Testis
Development
Spermatogenesis
Germ cells
RNA-sequencing
Transcriptomics

ABSTRACT

The testis plays a central role in the male reproductive system - secreting several hormones including male steroids and producing male gametes. A complex and coordinated molecular program is required for the proper differentiation of testicular cell types and maintenance of their functions in adulthood. The testicular transcriptome displays the highest levels of complexity and specificity across all tissues in a wide range of species. Many studies have used high-throughput sequencing technologies to define the molecular dynamics and regulatory networks in the testis as well as to identify novel genes or gene isoforms expressed in this organ. This review intends to highlight the complementarity of these transcriptomic studies and to show how the use of different sequencing protocols contribute to improve our global understanding of testicular biology.

© 2018 Elsevier B.V. All rights reserved.

Contents

1. Introduction	47
2. Tissue profiling analyses and first evidences of a testicular specificity	49
3. Gonad development and somatic cell differentiation	50
4. Primordial germ cells and gonocytes	51
5. Spermatogonial stem cells	51
6. Gene expression program at play during spermatogenesis	52
7. Splicing factors and regulatory RNA at play during male germ cell differentiation	53
8. Resources	54
9. New technologies and perspectives	54
Funding	55
Acknowledgements	55
Supplementary data	55
References	55

1. Introduction

The transcriptome can be defined as the entire set of transcripts present in a given biological sample. For a long time, the

central dogma in molecular biology assumed that one gene is transcribed into one messenger RNA (mRNA) which in turn is translated into one protein that fulfills various structural, biological and/or regulatory functions within the cell. However, it is now clear that the transcriptome is a far more complex machinery, composed by several RNA classes performing a myriad of functions. About two thirds of human genes contain more than one alternatively spliced exon (Johnson et al., 2003) and are thus

* Corresponding author.

E-mail address: frederic.chalmel@inserm.fr (F. Chalmel).

¹ These authors contributed equally to this work.

Abbreviations			
BS-seq	bisulfite sequencing	PGCs	primordial germ cells
CB	chromatoid body	PGCLCs	PGC like cells
CeRNAs	competing endogenous RNA	piRNA	Piwi interacting RNA
ChIP-seq	chromatin immunoprecipitation followed by sequencing	RNA-seq	RNA sequencing
CircRNA	circular RNA	rRNA	ribosomal RNA
Dpc	days post coitum	SAGE	serial analysis of gene expression
Dpp	days post partum	scRNA-seq	Single-cell RNA sequencing
EST	Expressed Sequenced Tag	siRNA	small interfering RNA
GCs	germ cells	sncRNA	small noncoding RNA
lncRNA	Long noncoding RNA	Spc	spermatocytes
miRNA	micro RNA	Spg	spermatogonia
NOME-seq	nucleosome occupancy and DNA methylation profiling followed by sequencing	Spt	spermatids
		Spz	spermatozoa
		SSCs	spermatogonial stem cells
		UHTS	ultra-high-throughput sequencing

able to produce numerous RNA and protein isoforms. A large part of the transcriptome is also composed of RNAs that do not encode proteins - noncoding RNAs (ncRNAs). These are involved in a wide variety of biological processes (for review, see: [Cech and Steitz, 2014](#)), including testis differentiation/development ([Rastetter et al., 2015](#); [Taylor et al., 2015](#)) and spermatogenesis ([Luk et al., 2014](#)). Long noncoding RNAs (lncRNAs) constitute the first class of noncoding transcripts, longer than 200bp, and are involved in many regulatory mechanisms, such as transcriptional, post-transcriptional and direct protein activity regulation ([Wang and Chang, 2011](#)). Small noncoding RNAs (sncRNAs) refer to transcripts shorter than 200bp that can be further divided into micro RNAs (miRNAs), small interfering RNAs (siRNAs) or Piwi interacting RNAs (piRNAs), most of which regulate gene expression at a transcriptional and/or post-transcriptional level (for review, see: [Luo et al., 2016](#)). Capturing all types of RNA species is necessary to fully understand the transcriptional network at play within a given cell. Owing to the intrinsic structural differences between mRNAs, lncRNAs and sncRNAs, transcriptomic studies typically focus on a single type of RNA. These intrinsic differences also warrant specific extraction methods, sequencing protocols and distinctive analytical strategies.

The first transcriptomic technology, called Expressed Sequenced Tag (EST), was based on the Sanger method and consisted of partial sequencing of cDNAs cloned into bacterial plasmids. The serial analysis of gene expression (SAGE) ([Velculescu et al., 1995](#)) was an improved version of the EST technique that allowed increasing the sequencing throughput as well as quantifying the corresponding transcripts. Both EST and SAGE techniques already introduced the notion of transcript assembly (tags concatenated into contigs) prior to association of corresponding genes in a reference genome. While these methodologies paved the way to decipher transcriptomes, they remained resource, time and money consuming. The technology that allowed transcriptomics to truly become a high-throughput discipline was DNA microarrays. These consist of solid surfaces on which DNA fragments, called probes, are spotted or synthesized *in situ* ([Heller, 2002](#); [Nelson, 2001](#); [Pozhitkov et al., 2007](#); [Schena et al., 1995](#)). The complementary sequence between probes and transcripts (or corresponding cDNAs) allows their specific hybridization, while radioactivity or fluorescence intensity emitted following hybridization is used for quantification purposes. The design of probes is critical and requires prior knowledge of genes and transcripts. Aside from the whole-genome tiling arrays designed to interrogate an entire genome ([Mockler et al., 2005](#)), this technology does not allow the characterization of new

transcriptional events and, ultimately, the discovery of new genes. Nonspecific hybridization can also result in biased transcript quantification ([Pozhitkov et al., 2007](#)). Even though the number of PubMed records using the term “microarray” is declining since 2015, most probably because of the democratization of high-throughput RNA-sequencing (RNA-seq), the overall fast assessment of gene expression at a reasonable price explains why microarrays remain a widely-used technology.

Over the last decade, ultra-high-throughput sequencing (UHTS) technologies have revolutionized transcriptomics ([Morozova et al., 2009](#); [Wang et al., 2009](#)), and all fields under the umbrella of genomics. Briefly, total RNAs or a subset of these are reverse-transcribed, amplified by PCR and fragmented. Subsequently, millions of short DNA fragments are sequenced in parallel, leading to generation of many sequences called reads. The mapping of these reads onto a reference genome (genome-based assembly) or of reads to each other (*de novo* assembly) is then required to reconstruct the initial transcripts. Finally, the expression level is assessed by counting reads associated to each transcript or gene ([Fig. 1](#)). Such quantification methods have been shown to be accurate and well-correlated to other technologies such as microarrays, and to eventually outperform these in terms of accuracy for low-abundance transcripts ([Chen et al., 2017a](#); [Izadi et al., 2016](#); [Wang et al., 2014](#)). Since the emergence of RNA-seq, the accuracy and sensitivity of related methodologies have greatly improved. In addition, these techniques have diversified to respond to scientists' specific biological questions ([Head et al., 2014](#)). For example, the sequencing depth can be improved by selecting RNAs of specific interest, such as polyA-RNAs, or by depleting undesired ones (e.g. rRNAs). Depending on the study's focus, the sequencing protocol can also be optimized. Single-end sequencing is usually sufficient for gene expression analysis, while paired-end sequencing is highly recommended for the purpose of transcript assembly - an absolute pre-requisite to the identification of new isoforms or novel genes ([Conesa et al., 2016](#)). Preserving strand information is also crucial to determine the direction in which a given locus is transcribed, allowing a better gene prediction in non-model organisms and the discovery of antisense lncRNAs. Over the past years, many bio-informatic tools have also been developed to improve all steps of UHTS analyses, including mapping of reads to a reference genome, transcript reconstruction and quantification, and differential expression analysis. Rather than describing the best practices in experimental design ([Hardwick et al., 2017](#)), RNA-seq protocol ([Conesa et al., 2016](#); [Han et al., 2015](#); [Hrdlickova et al., 2017](#)) or differential expression analysis ([Fang et al., 2012](#); [Huang et al.,](#)

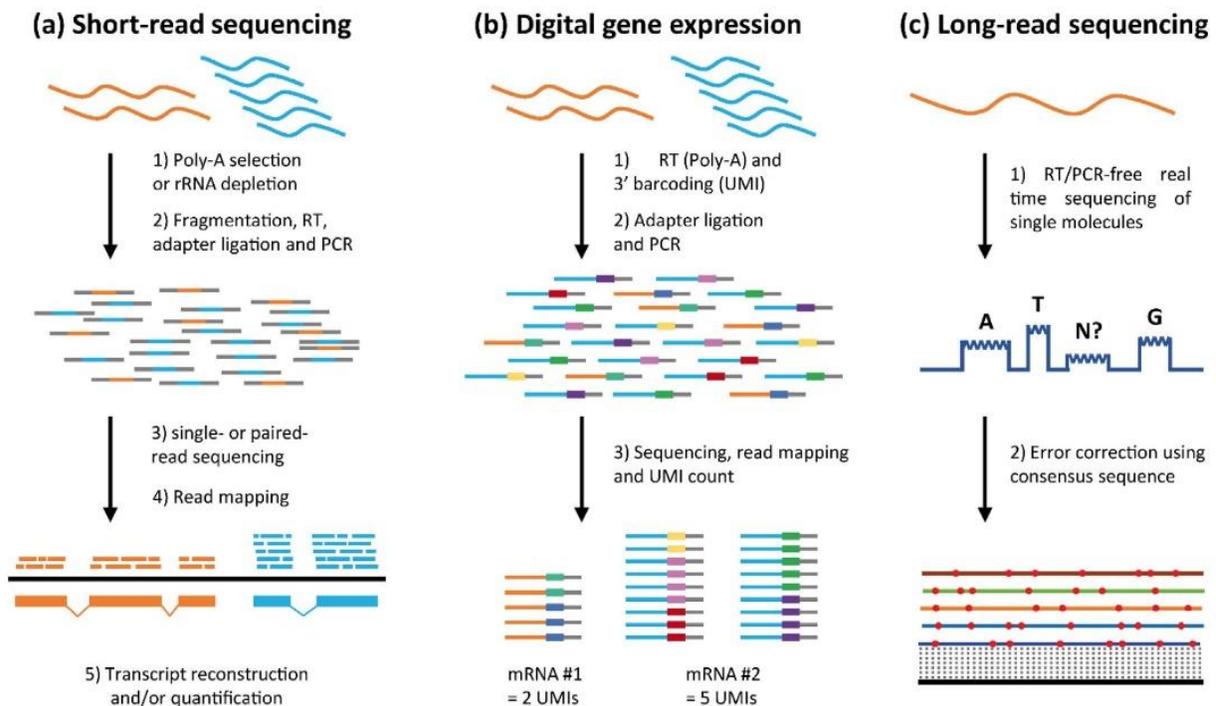


Fig. 1. High-throughput sequencing technologies supporting transcriptomic analyses (2-column fitting, colors should be used).

(a) Until recently, most sequencing technologies have relied on the generation of relatively short reads (*i.e.* sequences of few tens to hundreds of nucleotides in length). When applied to the sequencing of RNA, following reverse transcription (RT) of either poly-A enriched or of ribosomal RNA (rRNA)-depleted transcripts, the bioinformatic analysis then consists in mapping reads to a reference sequence (typically the genome if available) prior to transcript assembly (based on reads spanning exonic junctions) and/or quantification (based on density of reads mapping each individual transcript/gene). (b) Alternatively, modified library construction protocols involve the barcoding of each cDNA by a Unique Molecular Index (UMI) during reverse transcription, followed by sequencing of the 3' extremity of the resulting cDNA. The counting of UMIs, rather than measuring read density, then allows more precise quantification, by correcting for eventual PCR biases. Such protocols were designed for differential gene expression analysis and are particularly well-suited for single-cell RNA-seq analyses as well as low-input samples. (c) Currently, third-generation sequencing technologies are revolutionizing many fields of genomics, including transcriptomics. These emerging technologies indeed differ from previous one since they allow the direct sequencing of RNAs (*i.e.* with neither reverse transcription nor PCR amplification) and the generation of long-reads (*i.e.* sequences of several tens of kilobases in length). Consequently, they make it possible to sequence full length transcripts, therefore avoiding the tedious and equivocal step of transcript assembly. The relatively low throughput (in terms of number of sequences delivered per run) at a still relatively high cost, however, do not make yet such technologies an appropriate solution for quantification purposes.

2015; Kvam et al., 2012; Wu and Wu, 2016), this review intends to illustrate the wide range of biological questions that can be answered, thanks to these sequencing methods, using the testis as a complex organ model.

The establishment and maintenance of testicular functions, *i.e.* the secretion of several hormones including androgens and the continuous production of male gametes, require a permanent communication network between several cell types. The testicular transcriptome therefore represents the sum of transcripts expressed in distinct cell populations, present in various proportions within the testis, including Leydig cells (that produce androgens), peritubular cells (that surround seminiferous tubules), Sertoli cells (that support germ cells) and germ cells (GCs) at various differentiation stages (mitotic spermatogonia, meiotic spermatocytes and haploid spermatids). Despite its extreme cellular complexity, a commonly-used strategy to study the testis consists of using the whole organ at different ages: fetal or post-natal. The expression changes highlighted with such approaches can be used to identify: i) expression dynamics within a given cell population, and/or ii) preferential expression in cell types whose proportions fluctuate during development. Additionally, many techniques allow the isolation of enriched populations of each testicular cell type (Fig. 2). These two different strategies have recently been combined with the UHTS methods, notably RNA-seq, to gain insights into testis biology.

2. Tissue profiling analyses and first evidences of a testicular specificity

RNA-seq has been used in a wide range of species to perform tissue profiling analyses, *i.e.* studies in which the transcriptomes of various tissues or organs, including the testis, are analyzed and compared. A pioneer study using RNA-seq was performed on several tissues in humans and rodents (Ramsköld et al., 2009). This study demonstrated that the most complex tissues in terms of number of expressed genes and RNA classes were the brain, kidney and testis. With the aim of describing the evolutionary dynamics of the mammalian transcriptomics, Brawand and colleagues investigated six organs from several species representing mammalian lineages and avians (Brawand et al., 2011). This analysis pointed towards the testis being the most rapidly evolving organ with regards to gene expression, thereby confirming previous phenotypic observations (Harcourt et al., 1981). UHTS has also been combined with proteomics analysis in order to define the specific proteome of several organs (Djureinovic et al., 2014). The authors used RNA-seq as a clue to determine the protein-coding genes specifically expressed in all the tissues analyzed and concluded that the testis was by far the organ with the highest number of tissue-specific genes. Additionally, a landmark RNA-seq study comparing transcriptomes of 24 tissues and cell types in Humans was performed to gain insight into the role of lncRNAs in testis specificity (Cabili et al., 2011). The authors found that lncRNAs are transcribed

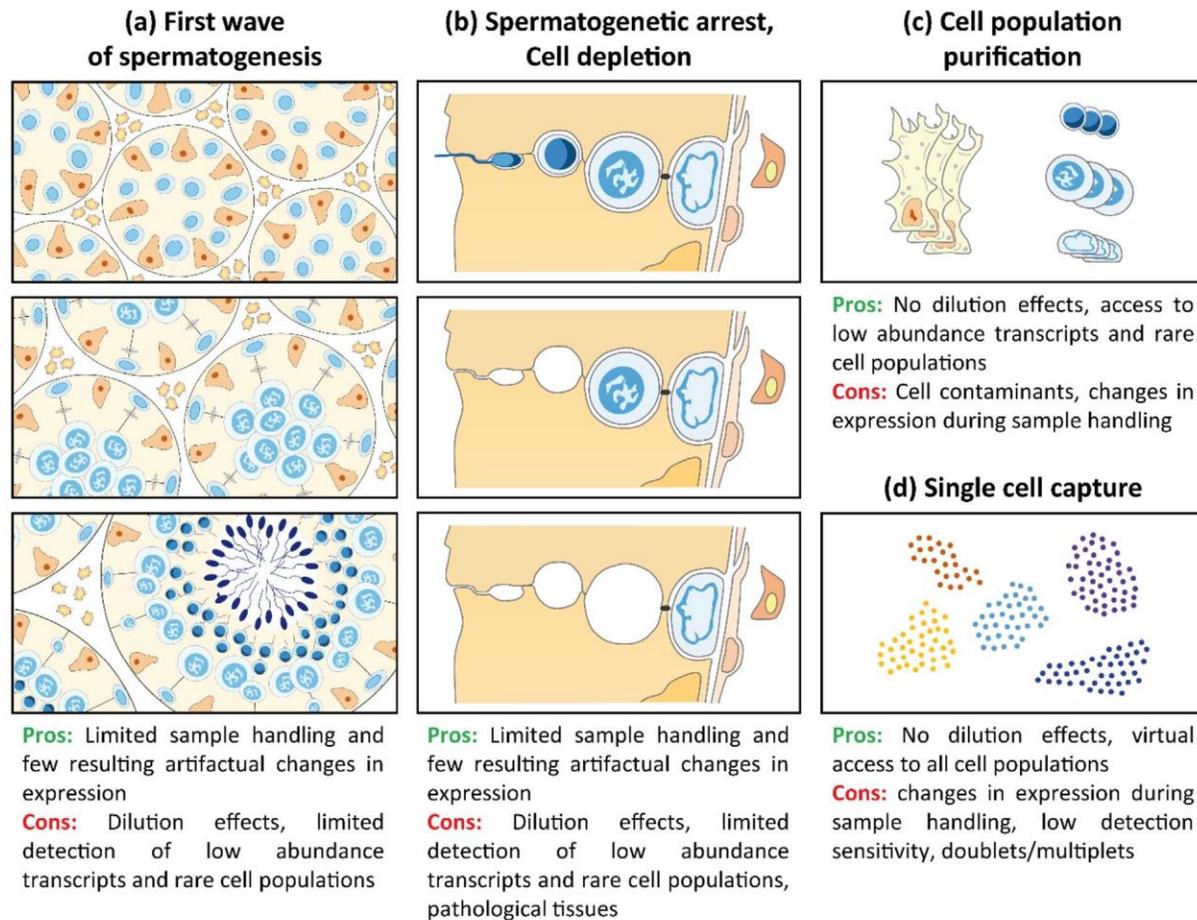


Fig. 2. Strategies to study gene expression during testis development and functions (2-column fitting, colors should be used).

Several strategies are typically used to investigate the transcriptomic profiles of testicular cells, notably during germ cell development. (a) These include the use of whole testes at various developmental stages during the first wave of spermatogenesis in various species. In such approaches, the specific or preferential expression program of a given cell type is inferred from the changes in expression observed between two stages during which this cell type appears. (b) Based on the same rationale, the use of testicular biopsies from patients with spermatogenesis arrested at distinct developmental stages has been commonly used in humans. (c) Alternatively, many isolation procedures enable to obtain enriched cell populations in order to investigate more directly the transcript content of various cell types present within the testis. (d) Finally, the relatively recent development of single-cell approaches combined with high-throughput technologies holds great promises for capturing and studying any cell type within a sample, including rare cell populations or cell subtypes for which absence of appropriate markers prevents purification. Main advantages (Pros) and disadvantages (Cons) are indicated for each strategy.

in a highly tissue-specific manner and that a third of these are predominantly expressed in the testis. These results were further confirmed by a tissue profiling analysis across 11 species, including mammals and tetrapods (Necsulea et al., 2014), which also showed that the high divergence rate of genes expressed within the testis does not exclusively apply to protein-coding genes, but is also inclusive of lncRNAs. This testis-specific expression of lncRNAs was further investigated by Zhang and collaborators, who performed RNA-seq on the whole testis and mature sperm. They compared their results with six other mouse tissues by integrating available data from NONCODE database, and they found that almost 7000 sequenced lncRNAs were exclusively expressed in the testis or sperm. Unlike other tissues, the average level of lncRNAs expression appeared higher than that of mRNAs in mature sperm.

These RNA-seq based studies have confirmed previous findings from microarray experiments (for reviews, see: Calvel et al., 2010; Chalmel and Rolland, 2015; Zhu et al., 2015) and have provided additional insight into testis-specificity of lncRNAs. The high tissue-specificity of the adult testis is mostly due to the germline expression program. During differentiation (i.e. meiosis and spermiogenesis), male GCs undergo unique biological processes that require molecular factors expressed exclusively in the testis.

Additionally, strong evolutionary constraints appear to act on the male gonad, shaping its genes with exaggerated traits, at both transcripts' sequences and expression levels.

3. Gonad development and somatic cell differentiation

Following sexual differentiation, the bipotent gonad differentiates to either a testis or an ovary. In the testis, this involves the differentiation of Sertoli and Leydig cells, the formation of testis cords as well as the early establishment of a relatively prominent vasculature. These developmental processes occur during the entire embryonic period and continue after birth. Many microarray studies have investigated these critical processes in various species and paved the way for a better understanding of the underlying molecular mechanisms (Beverdam and Koopman, 2005; Bouma et al., 2007; Combes et al., 2011; del Valle et al., 2017; Houmard et al., 2009; Jameson et al., 2012; Munger et al., 2013; Small et al., 2005). To date, only a few studies performed in non-mammalian species, have analyzed this process at the whole organ level with the use of RNA-seq.

For instance, the transcriptomes of chicken ovaries and testes at two developmental time-points surrounding sex differentiation

(4.5 and 6 dpc) were investigated using a poly(A)-enrichment protocol (Ayers et al., 2015). Additionally, the transcriptional dynamics throughout gonadal development were studied in *Portunus trituberculatus* by sequencing whole testis and ovaries at various stages of development (Meng et al., 2015). The authors undertook *de novo* transcript assembly using the Trinity tool (O'Neil and Emrich, 2013) and used protein databases (Nr, Pfam, Swiss-Prot) as well as Blast2go and WEGO softwares for annotation purposes. Owing to the non-reliance of RNA-seq on genome annotation, it offered a great potential of discovery in non-model species studies.

Till date, only a single study has focused on somatic cell differentiation within the fetal mammalian testis using RNA-seq (McClelland et al., 2015). The authors used FACS to isolate sf1-eGFP cells from mouse testes at different stages. By comparing Sertoli cells (high-GFP cells), Leydig cells (low-GFP cells) and GFP-negative cells, they were able to identify 61 genes, previously unknown to be involved in fetal Leydig cells differentiation, prior to the onset of steroidogenic genes. In this context, it is also worth mentioning the RNA-seq analysis of Sertoli cells isolated from Sox9-EGFP mouse testes at different postnatal ages (Zimmermann et al., 2015). This study also integrated data from purified GCs (Soumillon et al., 2013) to filter out transcripts arising from potential residual GCs. By doing so, the authors successfully described the expression dynamics of known transcripts as well as of newly discovered isoforms and unannotated transcripts (29 novel unknown intronic or intergenic transcripts) in Sertoli cells as they differentiate from immature proliferating cells to mature quiescent cells (in charge of sustaining spermatogenesis).

4. Primordial germ cells and gonocytes

Studies on sex differentiation and testicular somatic cells using RNA-seq are less numerous than microarray-based studies (see above). However, there have been many studies investigating primordial germ cells (PGCs) and/or gonocytes by means of UHTS over the recent years. These cells indeed undergo a deep epigenetic reprogramming involving removal of cytosine methylation from imprinted genes and restoration of totipotency (Hajkova et al., 2010; Popp et al., 2010; Seki et al., 2005) (reviewed in this issue). In this context, many studies have characterized the chromatin status of immature GCs from Oct4-GFP mice (Hammoud et al., 2015; Kubo et al., 2015; Lesch et al., 2013; Seisenberger et al., 2012) by combining RNA-seq together with epigenomics technologies (for review, see: Mensaert et al., 2014), including bisulfite sequencing (BS-seq), (hydroxyl-)methylated DNA immunoprecipitation followed by sequencing ((h)MeDIP-seq) (Jacinto et al., 2008; Jin et al., 2011); chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Park, 2009), or nucleosome occupancy and DNA methylation profiling followed by sequencing (NOME-seq) (Kelly et al., 2012). A first study compared gene expression and DNA methylation (using BS-seq) and showed that global erasure of methylation does not lead to promiscuous transcription, including retrotransposons, suggestive of other transcriptional repression mechanisms in these cells (Seisenberger et al., 2012). A second study correlated transcriptional activity with poised chromatin domains, i.e. which harbor both activation (H3K4me3) and repression (H3K27me3) histone marks, using ChIP-seq and highlighted a set of essential developmental regulators which poised chromatin state was conserved from fetal life (in PGCs) to adulthood (in meiotic and post-meiotic germ cells) (Lesch et al., 2013). The authors further proposed that the flexible transcriptional state of these genes from fertilization onwards is essential to regenerate totipotent cells from differentiated germ cells. Finally, one study combined BS-seq and RNA-seq to compare mouse gonocytes (16.5dpc), prospermatogonia (0.5 dpp), undifferentiated (KIT-) and

differentiating (KIT+) spermatogonia (7.5 dpp), allowing to correlate methylation status together with were changes in expression of important genes (Kubo et al., 2015).

In humans, the *in vitro* differentiation of the PGC like cells (hPGCLCs) from embryonic stem cells (hESCs) could represent a great mean for the study of PGCs but such models also require a better understanding of fetal GCs. For instance, a first transcriptional landscape of purified gonocytes (cKIT+) from human fetal testes and ovaries, between 8 and 20 gestational weeks (GW), was generated to evaluate and compare five protocols for *in vitro* generation of hPGCLCs, all of which proved unsatisfactory (Gkoutela et al., 2013). In 2015, four independent studies investigated human PGCs and gonocytes between 4 and 19 GW using similar isolation procedures (Gkoutela et al., 2015; Guo et al., 2015; Irie et al., 2015; Tang et al., 2015). Combination of RNA-seq and BS-seq produced comprehensive datasets involving transcriptional and single-base-resolution methylome dynamics during GC development. Except for SINE-variable number of tandem repeats-Alu elements (SVAs) (Tang et al., 2015), global DNA demethylation was not found to correlate with gene upregulation, notably for transposons, implicating the presence of another regulatory process (Gkoutela et al., 2015; Tang et al., 2015). Another study combined RNA-seq and NOME-seq with previously published ChIP-seq data (H. Guo et al., 2017a,b), showing that promoters' accessibility is strongly correlated to the corresponding gene expression level. The SVAs elements also specifically displayed an open state in fetal GCs. In terms of transcriptional and epigenomic dynamics during GC development, two studies have successfully highlighted similarities as well as significant differences between human and mouse (H. Guo et al., 2017a,b; Tang et al., 2015). Other labs also developed protocols for the *in vitro* differentiation of hPGCLCs from hESCs or human induced pluripotent stem cells (hiSPCs). A first group evaluated such *in vitro* differentiation protocol by comparing hPGCLCs to hPGCs, cultured hESCs, pre-induced cells, the seminoma-derived Tcam-2 cell line as well as somatic cells using RNA-seq (Irie et al., 2015). They first showed a significant correlation between the transcriptomes of hPGCLCs, hPGCs and Tcam-2 cells. They also identified hPGCLCs as potential "differentiating" hPGCs as they do not express the typical late germ cell markers DAZL, VASA and MAEL, further demonstrating that hPGCLCs constitute an interesting model to investigate the mechanisms of human germ line establishment. Most importantly, they identified the requirement of SOX17 for specification of human PGCs (Irie et al., 2015). A second group compared the transcriptome of hPGCLCs with other previously published RNA-seq data and evidenced significant differences in terms of gene expression program and epigenetic reprogramming between human and mouse PGCLCs as well as regarding the respective timing of specification and reprogramming between these species (von Meyenn et al., 2016).

5. Spermatogonial stem cells

Spermatogonia (Spg) constitute a heterogeneous population of male GCs that include spermatogonial stem cells (SSC), proliferative progenitor Spg and differentiating Spg committed to spermatogenesis. In mouse, undifferentiated Spg consist of type A single (As), paired (Apr) and aligned (Aal4-16) Spg while differentiating Spg consist of A1-4 Spg, intermediate Spg and type B Spg. Whether only As Spg encompass the pool of SSC is still a matter of debate (for review, see Lord and Oatley, 2017). In humans, there are only three types of Spg: undifferentiated Adark and Apale Spg, which are thought to be reserve and active SSC, respectively, and type B Spg which are differentiating progenitors (Boitani et al., 2016).

An extensive characterization (RNA-seq, smallRNA-seq, ChIP-seq and BS-Seq) of mouse GCs at different stages, including self-

renewing Spg (THY1+), differentiating Spg (KIT+), spermatocytes (Spc), spermatids (Spt) and mature sperm allowed for the unraveling of key pathways regulating the balance between self-renewal and differentiation of SSCs (Hammoud et al., 2014). This study also showed epigenetic specificities concerning these cells, which includes the poising of key transcription factors as well as the activity of DNA-methylated promoters during gametogenesis. By including additional markers (THY1, KIT, OCT4, ID4, and GFRa1), the same group further characterized the transcriptome and epigenome of adult mice SSCs (Hammoud et al., 2015) and proposed a classification of spermatogonial subpopulations, i.e. epithelial-like Spg (THY1+; highOCT4, ID4, and GFRa1), mesenchymal-like Spg (THY1+; moderate OCT4 and ID4; high mesenchymal markers), and Spg committed to gametogenesis (high KIT+), that differ in terms of methylation status and imprinting. The recent identification of ID4 as being preferentially-expressed in a subset of mouse As Spg and playing an important role in the maintenance of SSC pool (Oatley et al., 2011), allowed further characterization of the SSC transcriptome (Chan et al., 2014; Helsel et al., 2017; Hermann et al., 2015; Mutoji et al., 2016). Taking advantage of previously-published transcriptomic datasets (Hammoud et al., 2015; Mutoji et al., 2016), a list of 123 genes representing the core expression program of SSC was identified (Helsel et al., 2017).

A great work has been recently conducted, using SSEA4 as a marker of human SSCs, and KIT as a marker of Spg committed into gametogenesis (J. Guo et al., 2017a,b). The authors performed whole-genome bisulfite sequencing (WGBS), Assay for Transposase-Accessible Chromatin with highthroughput sequencing (ATAC-seq) and RNA-seq on these two purified cell populations, aiming to describe the complete DNA methylation, chromatin and transcriptional states which may explain their stemness differences, and focused their analysis on pluripotency factors. Single-cell RNA-seq technologies were also used to profile the transition between SSEA4+ and KIT+ cells. The authors described four distinct cellular states, which highlight more precisely the differentiation pathway of human SSCs. This work represents a great example of the complementarity and power of different high-throughput technologies (J. Guo et al., 2017a,b).

A number of studies have also been conducted in order to characterize the repertoire of ncRNAs, including sncRNAs, expressed in immature male GCs. The pioneer work in this field compared mouse SSC (THY1+) to somatic cells (THY1-) and identified a set of preferentially-expressed miRNAs in self-renewing SSC (Niu et al., 2011). The use of unbiased sequencing also confirmed that piRNAs represent the preponderant class of smallRNA in differentiating mouse GCs, i.e. Spc and Spt, whereas, other cell types such as ESCs, SSCs, Sertoli cells, and mesenchymal stem cells preferentially express miRNAs (Tan et al., 2014). Recently, RNA-seq was also used to characterize not only classical mRNAs and lncRNAs but also circular RNAs (circRNAs) expressed in SSCs and ovarian germline stem cells (X. Li et al., 2017a,b). circRNA is a novel class of ncRNA which specificity lies in its covalent bond linking the 3' and 5' ends generated by backsplicing (for review, see: Huang et al., 2017). Thanks to correlation analysis, a competing endogenous RNA (ceRNAs) network, comprising lncRNA-miRNA-circRNA-mRNA, identified competitive RNAs - capable of binding and inhibiting the regulatory activities of other RNAs.

6. Gene expression program at play during spermatogenesis

While at the first glance, working with a whole organ might seem irrelevant for studying GC differentiation, it has been successfully demonstrated that in several mammalian species (including rodents), the first wave of spermatogenesis is synchronous in all seminiferous tubules (Bellvé et al., 1977; Oakberg, 1957,

1956). Therefore, following the evolution of the testicular transcriptome at representative post-natal ages allows extrapolation of the gene expression program of each GC population (Fig. 2). This hypothesis is based on the assumption that observed transcriptional variations are due to germline differentiation. This strategy has the critical advantage of conserving the testis's integrity by preserving cell junctions and avoiding transcriptome modifications inherent to cell isolation procedures. In 2013, Gong and collaborators were among the first to profile the transcriptome of the developing mouse testis at three postnatal stages (infant, 6 dpp; juvenile, 4 weeks after birth; and, adult, 10 weeks after birth) using UHTS and a polyA-enrichment protocol (Gong et al., 2013). At the same time, Laiho and collaborators published about the transcriptome evolution during the first wave of spermatogenesis using postnatal mouse testes at 7, 14, 17, 21 and 28 dpp (Laiho et al., 2013). Instead of using a classical polyA-enriched approach, the authors used a rRNA depletion strategy to characterize the transcriptional dynamics of a wide range of protein-coding and noncoding transcripts. They observed for the first time that lncRNAs accumulate during meiotic and postmeiotic stages.

Kaessmann's lab was among the first to determine the cellular source of the testicular transcriptome complexity during mouse spermatogenesis (Soumillon et al., 2013) using RNA-seq on five enriched populations of testicular cells including Sertoli cells, type A Spg, pachytene Spc, round Spt and spermatozoa (Spz). Their results confirmed that meiotic and postmeiotic GCs have the highest transcriptome complexity (considering protein-coding genes, lncRNAs, pseudogenes or transposable elements), in comparison to testicular somatic cells (Sertoli cells) or other organs. At the same time, Gan and colleagues also published a landmark paper describing global 5-hydroxymethylcytosine distribution in GCs and its impact on the mouse germline transcriptional landscape (Gan et al., 2013). The authors sequenced the transcriptome of seven isolated cell populations (primitive SpgA, SpgA, SpgB, preleptotene Spc, pachytene Spc, round Spt and elongated Spt) and found that the highest transcriptional transitions during spermatogenesis occurred between preleptotene Spc and pachytene Spc, and also between pachytene Spc to round Spt. The following year, Chalmel and collaborators used RNA-seq to study the rat testicular non-coding expression program and to discover novel genes (Chalmel et al., 2014). Four distinct testicular cell populations (Sertoli cells, Spg, Spc and Spt) were investigated using a rRNA depletion strategy. This allowed them to discover 1419 novel genes, called TUTs for testis-expressed unannotated transcripts, most of which exhibited characteristic genomic features of lncRNAs. This study confirmed that lncRNAs and TUTs accumulate in the meiotic and postmeiotic GCs (Laiho et al., 2013; Soumillon et al., 2013) and identified a distinct class of meiosis-related lncRNAs exhibiting exons twice as long as those of other transcripts. By using a "proteomics informed by transcriptomics" (PIT) strategy combining RNA sequencing data with shotgun proteomics analyses, the authors further identified 44 novel protein-coding transcripts initially thought to be lncRNAs or TUTs (Chocu et al., 2014; Evans et al., 2012). Recently, Da Cruz and colleagues took advantage of improved FACS-based methods (da Cruz et al., 2016) to isolate leptotene/zygotene Spc, pachytene Spc, secondary Spc, and round Spt and subsequently performed RNA-seq analysis. Their meiosis-centric study revealed a large number of genes showing a burst of expression at the leptotene/zygotene stage. The authors also confirmed that the transcription of some spermiogenesis-related genes is initiated as early as pachytene Spc. They also found that some X-linked genes are overexpressed during meiosis, thus escaping the meiotic sex chromosome inactivation (MSCI) (Ichijima et al., 2012; van der Heijden et al., 2011). Another RNA-seq study on five enriched mouse GCs populations focused on lncRNAs and also identified candidate genes

that escape MSCI (Wichman et al., 2017). Moreover, another study analyzed the differential expression of not only mRNAs, lncRNAs but also included circRNAs in mouse primitive SpgA, preleptotene Spc, pachytene Spc and rSpt (Lin et al., 2016). Interestingly, the authors minimized the limitations of their study by taking advantage of several previous transcriptomic studies to define the set of testis-specific genes expressed in their samples (Brawand et al., 2011; Cabili et al., 2011), assess the conservation of the spermatogenic expressed lncRNAs (Necsulea et al., 2014), compare the circRNAs production across mouse tissues (Rybak-Wolf et al., 2015) and investigate the expression of potential precursors of piRNAs (Li et al., 2013).

In humans, microarray experiments have been conducted to decipher the specific transcriptome of testicular cell populations and to identify key genes involved in spermatogenesis by using biopsies from patients with distinct infertility phenotypes (Chalmel et al., 2012; Ellis et al., 2007; Fox et al., 2003; Gatta et al., 2010; Nguyen et al., 2009; Okada et al., 2008; Spiess et al., 2007; von Kopylow et al., 2010). Similar approaches were also used in rodents with GC-deficient models (Ellis et al., 2004), and chemical induced infertility (Orwig et al., 2008; Rockett et al., 2001). Recently, studies have relied on testicular samples from patients who underwent orchiectomy (Djureinovic et al., 2014). Zhu and colleagues used testicular biopsies of patients with obstructive azoospermia (i.e. with normal spermatogenesis) and combined FACS and MACS methods to obtain purified Spg, Spc and Spt (Zhu et al., 2016). The RNA-seq analysis allowed them to describe stage specific expression of known protein-coding and lncRNAs, as well as to focus their analysis on potential key transcriptional regulators of each GC stage such as HOXs, JUN, SP1, and TCF3. Recently, Jégou and collaborators published a RNA-seq analysis on five testicular cell populations including Leydig cells, Sertoli cells, peritubular cells, Spc and Spt (Jégou et al., 2017). By classifying differentially expressed genes into three broad co-expression clusters associated with somatic, meiotic or postmeiotic cells, the authors found that genes expressed during meiosis are significantly depleted in archaic hominin alleles (Sankararaman et al., 2016, 2014). Jan and collaborators used a microdissection approach to obtain enriched cell populations from adult human testis including A_{dark} and A_{pale} Spg, leptotene/zygotene Spc, early pachytene Spc, late pachytene Spc and round Spt (Jan et al., 2017). The dynamic changes of transcriptome have been characterized during human germline differentiation and enlightened that the Spg express transcripts, at protein level, which are required in later stages of spermatogenesis. The authors also compared their sequencing results with previously published datasets, in order to assess the efficiency of their cell isolation protocol (Zhu et al., 2016), and to define the conserved expression program between human and mouse spermatogenesis in GCs populations (da Cruz et al., 2016).

7. Splicing factors and regulatory RNA at play during male germ cell differentiation

RNA-binding proteins and splicing factors play a critical role during male gamete development (for reviews, see Bettgowda and Wilkinson, 2010; Elliott, 2004; Idler and Yan, 2012; Paronetto and Sette, 2010; Venables, 2002; Walker et al., 1999). Several studies also took great advantage of RNA-sequencing to study splicing events occurring during spermatogenesis and particularly in the meiosis phase. Schmid and collaborators compared the transcriptome between whole mouse testis before GCs enter meiosis (6dpp) and at the end of the process (21dpp) (Schmid et al., 2013). They used a dedicated pipeline, MISO (Katz et al., 2010), to identify differentially regulated isoforms or exons across samples, revealing switches in splicing patterns during the meiotic process. They

pointed out several splicing factors that may play a key role in the evolution of meiotic protein isoforms, which can be possibly crucial for the germline mitotic-meiotic transition. Margolin and colleagues investigated whole testes at eight developmental stages of between 6 and 38 dpp (Margolin et al., 2014). They identified more than 13000 novel isoforms, including 159 predicted transcripts mapped in intergenic regions, further demonstrating that alternative spliced isoforms are abundant through spermatogenesis. They found transcripts mapping on unknown splice junctions for which an open reading frame was maintained. In 2016, the dynamics of the transcriptional landscape governing testis maturation (20 dpp, 75 dpp, 270 dpp) has also been characterized in two closely related species, Meishan and Duroc boars (Ding et al., 2016). Using *Sus scrofa* as the reference genome to map reads, the authors enlightened a set of potential critical genes for the onset of spermatogenesis. They used a dedicated strategy based on the SOAPsplice tool, allowing them to identify alternative splicing events (Huang et al., 2011). Importantly, transcriptomic differences in such closely related species illustrates the high divergence of transcription within the testis across mammalian species. Alternative splicing has also been explored in enriched mouse populations of Spc and Spt, to decipher more precisely the splicing changes occurring through transmeiotic differentiation of GC (Naro et al., 2017). Following a polyA-enrichment protocol, the authors used the FAST-DB splicing annotation tool and identified intron retention (IRT) as being the most represented pattern. They showed that IRTs are highly stable transcripts that accumulate in the meiotic nuclei and are enriched in genes involved in mature gamete functions (i.e. Spt development and sperm-egg recognition). The authors therefore suggested that IRT allows storage and stabilization of key transcripts after the transcriptional burst in meiotic cells until their translation in transcriptionally-silenced post-meiotic cells.

Besides the widely described mRNAs and lncRNAs, sncRNAs are involved in many expression regulation pathways and can be divided in different classes according to their genomic features and sequence modifications. Specific RNA-seq protocols had to be developed to allow smallRNA purification, library preparation and sequencing, as well as mapping and annotation of corresponding reads. The role of sncRNAs in testis maturation and spermatogenesis have been recently reviewed (for reviews, see Luo et al., 2016; Meikar et al., 2011; Yadav and Kotaja, 2014), and will not be extensively described here. Briefly, sncRNA profiling and differential expression analyses were performed on testes at different developmental stages (Gebert et al., 2015; Li et al., 2016; Lian et al., 2012; Liu et al., 2012), by comparing the testis with other tissues (Luo et al., 2015), including the ovary (Kowalczykiewicz et al., 2014; Li et al., 2011). Several works used whole mouse testes at different ages to study the specific biogenesis of piRNAs during spermatogenesis (Beyret et al., 2012; Li et al., 2013). In humans, the small noncoding transcriptome has also been characterized in the adult testis (Ha et al., 2014; Yang et al., 2013) and several other studies have been performed on purified GCs to understand the implication of such sncRNAs during spermatogenesis (Gan et al., 2011; García-López et al., 2015; Goh et al., 2015; Vourekas et al., 2012; Zhang et al., 2015).

The Noora Kotaya's laboratory is particularly known in the field of reproductive biology for its work on the chromatoid body (CB). The CB is a GC-specific organelle composed of RNAs and RNA-binding proteins which appears in the cytoplasm of GCs during the transition between meiotic and post-meiotic phase of spermatogenesis (Kotaja and Sassone-Corsi, 2007; Meikar et al., 2011; Parvinen, 2005; Peruquetti, 2015). Meikar and collaborators isolated CBs from juvenile mouse testes at 22 and 26 dpp using several steps of cell lysis, filtration and immunoprecipitation (Meikar et al.,

2014). They combined sncRNA-seq and RNA-seq to compare purified CB to round Spt and identified sncRNAs, lncRNAs and mRNAs that accumulate within this organelle. This analysis showed that lncRNAs identified as genomic clusters generating piRNAs, accumulate in the CB, suggesting their important role in piRNAs biogenesis. However, a proteomic analysis failed at identifying the core primary piRNA processing components within the CB, therefore challenging the role of this organelle in the biogenesis of piRNAs.

8. Resources

While microarrays remain useful and accurate tools for measuring expression levels, RNA-seq provides a more complete transcriptomic solution by allowing transcript isoform determination and eventually new gene discovery, on top of being more accurate when it comes to quantification (Chen et al., 2017b; Izadi et al., 2016; Malone and Oliver, 2011; Wang et al., 2014). With the continuous decrease of sequencing costs, the number of RNA-seq studies therefore drastically increased over recent years, and the corresponding datasets now present a great potential for reanalysis or integration within one's own study. Similar to microarrays, the MIAME compliance (Brazma et al., 2001) applies to UHTS data and generalist databases such as the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) (Barrett et al., 2012), or the ArrayExpress (Kolesnikov et al., 2015, <https://www.ebi.ac.uk/arrayexpress/>) resources, also serve as repositories of high-throughput sequencing data. Most of the datasets arising from studies described in this review are therefore publicly available in these public repositories (see Supplemental Table). Despite that several groups have already taken advantage of these available resources, differences in sequencing protocols, such as transcript type and size selection, strand-specificity, length or pairing of reads, may somehow prevent full integration of, or direct comparison between different datasets.

Despite the wide use of UHTS, the handling of such data requires specific skills and is both time and resource consuming. Dedicated databases have therefore started to emerge in order to extract, summarize and/or visualize the corresponding information, including the field of reproductive biology. For instance, the ReproGenomics Viewer (RGV) is a useful toolbox dedicated to genomic data for the reproductive science community (Darde et al., 2015). It consists of a genome browser that allows the visualization of selected datasets from the literature in the genomic context of their respective species. Moreover, cross-species comparisons are now possible, allowing simultaneous visualization of several types of data (e.g. RNA-seq, ChIP-seq) performed in different species. Currently, studies implemented in RGV mainly focus on testis biology and spermatogenesis, but this tool is likely to be regularly updated in order to expand its focus to other areas of reproductive biology. As described earlier in this review, lncRNAs are expressed in a highly tissue-specific manner and especially in male GCs. Germ lncRNA constitutes a dedicated catalogue of known and novel lncRNAs (Luk et al., 2015), by integrating several published transcriptomic studies covering three GC populations, i.e. type A Spg, pachytene Spc and round Spt (Gan et al., 2013; Lee et al., 2012, 2009; Soumillon et al., 2013; Sun et al., 2013). Last but not the least, SpermBase represents a great resource dedicated to sperm transcriptome in four mammalian species, i.e. human, rat, mouse and rabbit (Schuster et al., 2016). Using a standardized sperm RNA isolation and sequencing protocol, this work allows a comparative RNA-seq analysis, covering long as well as small RNAs, and identifying the conserved counterpart of the sperm transcriptome across mammalian species.

9. New technologies and perspectives

The application fields of UHTS technologies have expanded at an amazing pace since its emergence, enabling investigation of RNA as well as DNA, both quantitatively and qualitatively, at the sequence, structural, and conformation levels. In parallel, the accuracy and throughput of sequencing methods are also improving to overcome technical limitations. For instance, most current UHTS approaches involve amplification steps by PCR and rely on the sequencing of relatively short sequences in order to maintain reasonable sequencing error rates. The future of UHTS particularly relies on the use of PCR-free protocols, to prevent amplification biases, and on increasing the length of reads, to avoid or, at least, reduce and make easier assembly steps for both transcripts and chromosomes/genomes (Chu et al., 2017). In this context, it is worth mentioning the emergence of third generation sequencing technologies which allow the PCR-free or even the direct sequencing of single RNA molecules and generate long reads of up to several kilobases (Garalde et al., 2018; Laver et al., 2015; Lu et al., 2016; Rhoads and Au, 2015) (Fig. 1). Such technologies remain very expensive and the sequence error rates are still very high (up to 10%), but they are likely to improve on both aspects in the near future. Very importantly, and also quite unexpectedly, both technologies are able to capture information on bases' modifications such as methylation during the sequencing process.

The outcome of transcriptomic studies also greatly depends on the biological material. While whole organs preserve cell-cell interactions and reduce gene expression alterations due to extensive sample handling, they prevent access to discrete cell populations as well as to low-copy transcripts, and they do not allow to ascertain their cellular origins. On the other hand, cell enrichment protocols provide cell-type information and improve sensitivity, but potential contaminations and/or changes in expression cannot be ruled out. Moreover, such a bulk approach, even with enriched cell populations, is likely to mask differences between distinct subpopulations. Over the past years, single cell approaches have been developed and adapted to RNA sequencing (Figs. 1 and 2). Single-cell RNA-seq (scRNA-seq) includes cell isolation and library preparation with molecular barcodes. Thanks to microfluidic (Klein et al., 2015; Zilionis et al., 2016), droplet-based (Macosko et al., 2015) or FACS-based methods (Jaitin et al., 2014), scRNA-seq is able to capture the transcriptome of a very large number of cells. These high-throughput technologies (Ziegenhain et al., 2017) have proved very efficient for inferring and describing the temporal dynamic of cell differentiation, or the spatial description of different cell populations within an organ (Kalisky et al., 2017; Macaulay et al., 2017; Tanay and Regev, 2017; Wagner et al., 2016). In the field of reproductive biology, several single-cell transcriptomic studies have already been performed, on human PGCs/gonocytes and fetal somatic cells (Guo et al., 2015; L. Li et al., 2017a,b) as well as on mouse fetal somatic cells (Stévant et al., 2018) (reviewed in this issue). scRNAseq is a promising technology that represents a great potential of discovery, but as with all other technologies, some experimental limitations are still in place. For instance, the dissociation of complex tissues into single-cell suspension can be the first limiting step. Single nucleus RNAseq (snuc-RNAseq) has initially been developed to overcome this issue in the brain (Habib et al., 2016; Lake et al., 2016), and then combined with microfluidic technology, leading to DroNc-seq (Habib et al., 2017). Furthermore, the different available technologies still show relatively low sensitivity, allowing to detect medium to highly expressed transcripts exclusively. Therefore, for now at least, these methods cannot replace classical RNA-seq approaches in terms of quantification accuracy and completeness.

The future of these single cell high-throughput technologies is

evolving (Yuan et al., 2017), and many perspectives remain to be explored, such as *in situ* transcriptomic analysis (Coskun and Cai, 2016; Satija et al., 2015), live imaging transcriptomic analysis (Skylaki et al., 2016), lineage tracing (McKenna et al., 2016; Perli et al., 2016; Schmidt et al., 2017) or single-cell multi-omics (Cheow et al., 2016; Genshaft et al., 2016; Hou et al., 2016).

Funding

This work was supported by the Swiss National Foundation (CRS115_171007), the French National Institute of Health and Medical Research (Inserm), the University of Rennes 1 and the French School of Public Health (EHESP).

Acknowledgements

We thank Dr. Julien Dagher and Dr. Sadia Khan for critical reading of the manuscript. We also thank the Swiss National Foundation (SNF), Research Institute for Environmental and Occupational Health (Irset), the National Institute of Health and Medical Research (Inserm), the University of Rennes 1 and the French School of Public Health (EHESP) for supporting this work.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.mce.2018.02.019>.

References

- Ayers, K.L., Lambeth, L.S., Davidson, N.M., Sinclair, A.H., Oshlack, A., Smith, C.A., 2015. Identification of candidate gonadal sex differentiation genes in the chicken embryo using RNA-seq. *BMC Genom.* 16, 704. <https://doi.org/10.1186/s12864-015-1886-5>.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2012. NCBI GEO: archive for high-throughput genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. <https://doi.org/10.1093/nar/gks1193>.
- Bellvé, A.R., Cavicchia, J.C., Millette, C.F., O'Brien, D.A., Bhatnagar, Y.M., Dym, M., 1977. Spermatogenic cells of the prepubertal mouse. isolation and morphological characterization. *J. Cell Biol.* 74, 68–85.
- Bettgowda, A., Wilkinson, M.F., 2010. Transcription and post-transcriptional regulation of spermatogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1637–1651. <https://doi.org/10.1098/rstb.2009.0196>.
- Beverdam, A., Koopman, P., 2005. Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. *Hum. Mol. Genet.* 15, 417–431. <https://doi.org/10.1093/hmg/ddi463>.
- Beyret, E., Liu, N., Lin, H., 2012. piRNA biogenesis during adult spermatogenesis in mice is independent of the ping-pong mechanism. *Cell Res.* 22, 1429–1439. <https://doi.org/10.1038/cr.2012.120>.
- Boitani, C., Di Persio, S., Esposito, V., Vicini, E., 2016. Spermatogonial cells: mouse, monkey and man comparison. *Semin. Cell Dev. Biol.* 59, 79–88. <https://doi.org/10.1016/j.semcdb.2016.03.002>.
- Bouma, G.J., Affourtit, J., Bult, C.J., Eicher, E.M., 2007. Transcriptional profile of mouse pre-granulosa and Sertoli cells isolated from early-differentiated fetal gonads. *Gene Expr. Patterns.* <https://doi.org/10.1016/j.modgep.2006.05.011>.
- Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., Albert, F.W., Zeller, U., Khaitovich, P., Grützner, F., Bergmann, S., Nielsen, R., Pääbo, S., Kaessmann, H., 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478. <https://doi.org/10.1038/nature10532>.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C.P., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., Vingron, M., 2001. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29, 365–371. <https://doi.org/10.1038/ng1201-365>.
- Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. <https://doi.org/10.1101/gad.17446611>.
- Calvel, P., Rolland, A.D., Jégou, B., Pineau, C., 2010. Testicular postgenomics: targeting the regulation of spermatogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1481–1500. <https://doi.org/10.1098/rstb.2009.0294>.
- Cech, T.R., Steitz, J.A., 2014. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 157, 77–94. <https://doi.org/10.1016/j.cell.2014.03.008>.
- Chalmel, F., Lardenois, A., Evrard, B., Mathieu, R., Feig, C., Demougin, P., Gattiker, A., Schulze, W., Jégou, B., Kirchoff, C., Primig, M., 2012. Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum. Reprod.* 27, 3233–3248. <https://doi.org/10.1093/humrep/des301>.
- Chalmel, F., Lardenois, A., Evrard, B., Rolland, A.D., Sallou, O., Dumargne, M.-C., Coiffec, I., Collin, O., Primig, M., Jégou, B., 2014. High-resolution profiling of novel transcribed regions during rat spermatogenesis. *Biol. Reprod.* 91 <https://doi.org/10.1095/biolreprod.114.118166>, 5–5.
- Chalmel, F., Rolland, A.D., 2015. Linking transcriptomics and proteomics in spermatogenesis. *Reproduction* 150, R149–R157. <https://doi.org/10.1530/REP-15-0073>.
- Chan, F., Oatley, M.J., Kaucher, A.V., Yang, Q.-E., Bieberich, C.J., Shashikant, C.S., Oatley, J.M., 2014. Functional and molecular features of the Id4+ germline stem cell population in mouse testes. *Genes Dev.* 28, 1351–1362. <https://doi.org/10.1101/gad.240465.114>.
- Chen, L., Sun, F., Yang, X., Jin, Y., Shi, M., Wang, L., Shi, Y., Zhan, C., Wang, Q., 2017a. Correlation between RNA-Seq and microarrays results using TCGA data. *Gene* 628, 200–204. <https://doi.org/10.1016/j.gene.2017.07.056>.
- Chen, L., Sun, F., Yang, X., Jin, Y., Shi, M., Wang, L., Shi, Y., Zhan, C., Wang, Q., 2017b. Correlation between RNA-seq and Microarrays Results Using TCGA Data. <https://doi.org/10.1016/j.gene.2017.07.056>.
- Cheow, L.F., Courtois, E.T., Tan, Y., Viswanathan, R., Xing, Q., Tan, R.Z., Tan, D.S.W., Robson, P., Loh, Y.-H., Quake, S.R., Burkholder, W.F., 2016. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat. Methods* 13, 833–836. <https://doi.org/10.1038/nmeth.3961>.
- Chocu, S., Evrard, B., Lavigne, R., Rolland, A.D., Aubry, F., Jégou, B., Chalmel, F., Pineau, C., 2014. Forty-Four novel protein-coding loci discovered using a Proteomics Informed by Transcriptomics (PIT) approach in rat male germ cells. *Biol. Reprod.* 91 <https://doi.org/10.1095/biolreprod.114.122416>, 123–123.
- Chu, J., Mohamadi, H., Warren, R.L., Yang, C., Birol, I., 2017. Innovations and challenges in detecting long read overlaps: an evaluation of the state-of-the-art. *Bioinformatics* 33, 1261–1270. <https://doi.org/10.1093/bioinformatics/btw811>.
- Combes, A.N., Bowles, J., Feng, C.-W., Chiu, H.S., Khoo, P.-L., Jackson, A., Little, M.H., Tam, P.P.L., Koopman, P., 2011. Expression and functional analysis of Dkk1 during early gonadal development. *Sex Dev.* 5, 124–130. <https://doi.org/10.1159/000327709>.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.V., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13. <https://doi.org/10.1186/s13059-016-0881-8>.
- Coskun, A.F., Cai, L., 2016. Dense transcript profiling in single cells by image correlation decoding. *Nat. Methods* 13, 657–660. <https://doi.org/10.1038/nmeth.3895>.
- da Cruz, I., Rodríguez-Casuriaga, R., Santinaque, F.F., Fariás, J., Curti, G., Capoano, C.A., Folle, G.A., Benavente, R., Sotelo-Silveira, J.R., Geisinger, A., 2016. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genom.* 17, 294. <https://doi.org/10.1186/s12864-016-2618-1>.
- Darde, T.A., Sallou, O., Becker, E., Evrard, B., Monjeaud, C., Le Bras, Y., Jégou, B., Collin, O., Rolland, A.D., Chalmel, F., 2015. The reproductomics viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Res.* 43, W109–W116. <https://doi.org/10.1093/nar/gkv345>.
- del Valle, I., Buonocore, F., Duncan, A.J., Lin, L., Barenco, M., Parnaik, R., Shah, S., Hubank, M., Gerrelli, D., Achermann, J.C., 2017. A genomic atlas of human adrenal and gonad development. *Wellcome Open Res.* 2, 25. <https://doi.org/10.12688/wellcomeopenres.11253.1>.
- Ding, H., Luo, Y., Liu, M., Huang, J., Xu, D., 2016. Histological and transcriptome analyses of testes from Duroc and Meishan boars. *Nat. Publ. Gr.* <https://doi.org/10.1038/srep20758>.
- Djureinovic, D., Fagerberg, L., Danielsson, A., Lindskog, C., 2014. The Human Testis Specific Proteome Defined by Transcriptomics and, vol. 20, pp. 1–39. <https://doi.org/10.1093/molehr/gau018>.
- Elliott, D.J., 2004. The role of potential splicing factors including RBMY, RBMX, hnRNP-G and STAR proteins in spermatogenesis. *Int. J. Androl.* 27, 328–334. <https://doi.org/10.1111/j.1365-2605.2004.00496.x>.
- Ellis, P.J., Furlong, R.A., Conner, S.J., Kirkman-Brown, J., Afnan, M., Barratt, C., Griffin, D.K., Affara, N.A., 2007. Coordinated transcriptional regulation patterns associated with infertility phenotypes in men. *J. Med. Genet.* 44, 498–508. <https://doi.org/10.1136/jmg.2007.049650>.
- Ellis, P.J., Furlong, R.A., Wilson, A., Morris, S., Carter, D., Oliver, G., Print, C., Burgoyne, P.S., Loveland, K.L., Affara, N.A., 2004. Modulation of the mouse testis transcriptome during postnatal development and in selected models of male infertility. *Mol. Hum. Reprod.* 10, 271–281. <https://doi.org/10.1093/molehr/gah043>.
- Evans, V.C., Barker, G., Heesom, K.J., Fan, J., Bessant, C., Matthews, D.A., 2012. De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat. Methods* 9, 1207–1211. <https://doi.org/10.1038/nmeth.2227>.
- Fang, Z., Martin, J., Wang, Z., 2012. Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. *Cell Biosci.* 2, 26. <https://doi.org/>

- 10.1186/2045-3701-2-26.
- Fox, M.S., Ares, V.X., Turek, P.J., Haqq, C., Reijo Pera, R.A., 2003. Feasibility of global gene expression analysis in testicular biopsies from infertile men. *Mol. Reprod. Dev.* 66, 403–421. <https://doi.org/10.1002/mrd.10364>.
- Gan, H., Lin, X., Zhang, Z., Zhang, W., Liao, S., Wang, L., Han, C., 2011. piRNA profiling during specific stages of mouse spermatogenesis. *Bioinformatics*. <https://doi.org/10.1261/rna.2648411>.
- Gan, H., Wen, L., Liao, S., Lin, X., Ma, T., Liu, J., Song, C.-X., Wang, M., He, C., Han, C., Tang, F., 2013. Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat. Commun.* 4 <https://doi.org/10.1038/ncomms2995>.
- Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E.J., Jayasinghe, L., Wright, C., Blasco, J., Young, S., Brocklebank, D., Juul, S., Clarke, J., Heron, A.J., Turner, D.J., 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods*. <https://doi.org/10.1038/nmeth.4577>.
- García-López, J., Alonso, L., Cárdenas, D.B., Artaza-Alvarez, H., Hourcade, J., de, D., Martínez, S., Briño-Enríquez, M.A., Del Mazo, J., 2015. Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization. *RNA* 21, 946–962. <https://doi.org/10.1261/rna.048215.114>.
- Gatta, V., Raicu, F., Ferlin, A., Antonucci, I., Scioletti, A., Garolla, A., Palka, G., Foresta, C., Stuppia, L., 2010. Testis transcriptome analysis in male infertility: new insight on the pathogenesis of oligo-azoospermia in cases with and without AZFc microdeletion. *BMC Genom.* 11, 401. <https://doi.org/10.1186/1471-2164-11-401>.
- Gebert, D., Ketting, R.F., Zischler, H., Rosenkranz, D., 2015. piRNAs from pig testis provide evidence for a conserved role of the piwi pathway in post-transcriptional gene regulation in mammals. *PLoS One* 10, e0124860. <https://doi.org/10.1371/journal.pone.0124860>.
- Genshaft, A.S., Li, S., Gallant, C.J., Darmanis, S., Prakadan, S.M., Ziegler, C.G.K., Lundberg, M., Fredriksson, S., Hong, J., Regev, A., Livak, K.J., Landegren, U., Shalek, A.K., 2016. Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction. *Genome Biol.* 17, 188. <https://doi.org/10.1186/s13059-016-1045-6>.
- Gkoutela, S., Li, Z., Vincent, J.J., Zhang, K.X., Chen, A., Pellegrini, M., Clark, A.T., 2013. The ontogeny of cKIT⁺ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat. Cell Biol.* 15, 113–122. <https://doi.org/10.1038/ncb2638>.
- Gkoutela, S., Zhang, K.X., Shafiq, T.A.A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y., Clark, A.T.T., Hargan-Calvopi, a., Chen, P.-Y., Clark, A.T.T., 2015. DNA demethylation dynamics in the human prenatal germline. *Cell* 161, 1425–1436. <https://doi.org/10.1016/j.cell.2015.05.012>.
- Goh, W.S.S., Falcatori, I., Tam, O.H., Burgess, R., Meikar, O., Kotaja, N., Hammell, M., Hannon, G.J., 2015. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev.* 29, 1032–1044. <https://doi.org/10.1101/gad.260455.115>.
- Gong, W., Linlin, P., Qiang, L., Yuanyuan, Z., Chengqi, X., Xiaomin, Y., Peng, C., Songnian, H., Jun, Y., 2013. Transcriptome Profiling of the Developing Postnatal Mouse Testis Using Next-generation Sequencing, vol. 56, pp. 1–12. <https://doi.org/10.1007/s11427-012-4411-y>.
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., Wang, W., Li, R., Yan, J., Zhi, X., Zhang, Y., Jin, H., Zhang, W., Hou, Y., Zhu, P., Li, J., Zhang, L., Liu, S., Ren, Y., Zhu, X., Wen, L., Gao, Y.Q., Tang, F., Qiao, J., 2015. The transcriptome and DNA methylation landscapes of human primordial germ cells. *Cell* 161, 1437–1452. <https://doi.org/10.1016/j.cell.2015.05.015>.
- Guo, H., Hu, B., Yan, L., Yong, J., Wu, Y., Gao, Y., Guo, F., Hou, Y., Fan, X., Dong, J., Wang, X., Zhu, X., Yan, J., Wei, Y., Jin, H., Zhang, W., Wen, L., Tang, F., Qiao, J., 2017a. DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res.* 27, 165–183. <https://doi.org/10.1038/cr.2016.128>.
- Guo, J., Grow, E.J., Yi, C., Mlcochova, H., Maher, G.J., Lindskog, C., Murphy, P.J., Wike, C.L., Carrell, D.T., Gorieli, A., Hotaling, J.M., Cairns, B.R., 2017b. Chromatin and single-cell RNA-seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. *Cell Stem Cell* 21, 533–546. <https://doi.org/10.1016/j.stem.2017.09.003>.
- Ha, H., Song, J., Wang, S., Kapusta, A., Feschotte, C., Chen, K.C., Xing, J., 2014. A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genom.* <https://doi.org/10.1186/1471-2164-15-545>.
- Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., Weitz, D.A., Rozenblatt-Rosen, O., Zhang, F., Regev, A., 2017. Massively parallel single-nucleus RNA-seq with DroNC-seq. *Nat. Methods* 14, 955–958. <https://doi.org/10.1038/nmeth.4407>.
- Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., Regev, A., 2016. Div-Seq: single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* 353, 925–928. <https://doi.org/10.1126/science.aad7038>.
- Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., Surani, M.A., 2010. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329, 78–82. <https://doi.org/10.1126/science.1187945>.
- Hammoud, S.S., Low, D.H.P., Yi, C., Carrell, D.T., Guccione, E., Cairns, B.R., Schübeler, D., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al., 2014. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* 15, 239–253. <https://doi.org/10.1016/j.stem.2014.04.006>.
- Hammoud, S.S., Low, D.H.P., Yi, C., Lee, C.L., Oatley, J.M., Payne, C.J., Carrell, D.T., Guccione, E., Cairns, B.R., 2015. Transcription and imprinting dynamics in developing postnatal male germline stem cells. *Genes Dev.* 29, 2312–2324. <https://doi.org/10.1101/gad.261925.115>.
- Han, Y., Guo, S., Muegge, K., Zhang, W., Zhou, B., 2015. Advanced applications of RNA sequencing and challenges. *Bioinform. Biol. Insights* 9, 29. <https://doi.org/10.4137/BBI.S28991>.
- Harcourt, A.H., Harvey, P.H., Larson, S.G., Short, R.V., 1981. Testis weight, body weight and breeding system in primates. *Nature* 293, 55–57.
- Hardwick, S.A., Deveson, I.W., Mercer, T.R., 2017. Reference standards for next-generation sequencing. *Nat. Rev. Genet.* 18, 473–484. <https://doi.org/10.1038/nrg.2017.44>.
- Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., Ordoukhanian, P., 2014. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* 56. <https://doi.org/10.2144/000114133>, 61–4, 66, 68, passim.
- Heller, M.J., 2002. DNA microarray technology: devices, systems, and applications. *Annu. Rev. Biomed. Eng.* 4, 129–153. <https://doi.org/10.1146/annurev.bioeng.4.020702.153438>.
- Helsel, A.R., Yang, Q.-E., Oatley, M.J., Lord, T., Sablitzky, F., Oatley, J.M., 2017. ID4 levels dictate the stem cell state in mouse spermatogonia. *Development* 144, 624–634. <https://doi.org/10.1242/dev.146928>.
- Hermann, B.P., Mutoji, K.N., Velte, E.K., Ko, D., Oatley, J.M., Geyer, C.B., McCarrey, J.R., 2015. Transcriptional and translational heterogeneity among neonatal mouse spermatogonia. *Biol. Reprod.* 92, 54. <https://doi.org/10.1095/biolreprod.114.125757>.
- Hou, Y., Guo, H., Cao, C., Li, X., Hu, B., Zhu, P., Wu, X., Wen, L., Tang, F., Huang, Y., Peng, J., 2016. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 26, 304–319. <https://doi.org/10.1038/cr.2016.23>.
- Houmard, B., Small, C., Yang, L., Naluai-Cecchini, T., Cheng, E., Hassold, T., Griswold, M., 2009. Global gene expression in the human fetal testis and ovary. *Biol. Reprod.* 443. <https://doi.org/10.1095/biolreprod.108.075747>.
- Hrdlickova, R., Toloue, M., Tian, B., 2017. RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* 8, e1364. <https://doi.org/10.1002/wrna.1364>.
- Huang, H.-C., Niu, Y., Qin, L.-X., 2015. Differential expression analysis for RNA-Seq: an overview of statistical methods and computational software. *Cancer Inf.* 14, 57–67. <https://doi.org/10.4137/CIN.S21631>.
- Huang, S., Yang, B., Chen, B.J., Bliim, N., Ueberham, U., Arendt, T., Janitz, M., 2017. The emerging role of circular RNAs in transcriptome regulation. *Genomics* 109, 401–407. <https://doi.org/10.1016/j.ygeno.2017.06.005>.
- Huang, S., Zhang, J., Li, R., Zhang, W., He, Z., Lam, T.-W., Peng, Z., Yiu, S.-M., 2011. SOAPSsplice: genome-wide ab initio detection of splice junctions from RNA-seq data. *Front. Genet.* 2, 46. <https://doi.org/10.3389/fgene.2011.00046>.
- Ichijima, Y., Sin, H.-S., Namekawa, S.H., 2012. Sex chromosome inactivation in germ cells: emerging roles of DNA damage response pathways. *Cell. Mol. Life Sci.* 69, 2559–2572. <https://doi.org/10.1007/s00018-012-0941-5>.
- Idler, R.K., Yan, W., 2012. Control of messenger RNA fate by RNA-binding proteins: an emphasis on mammalian spermatogenesis. *J. Androl.* 33, 309–337. <https://doi.org/10.2164/jandrol.111.014167>.
- Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., Surani, M.A., 2015. SOX17 is a critical specifier of human primordial germ cell fate. *Cell* 160, 253–268. <https://doi.org/10.1016/j.cell.2014.12.013>.
- Izadi, F., Zarrini, H.N., Kiani, G., Jelodar, N.B., 2016. A comparative analytical assay of gene regulatory networks inferred using microarray and RNA-seq datasets. *Bioinformatics* 12, 340–346. <https://doi.org/10.6026/97320630012340>.
- Jacinto, F.V., Ballestar, E., Esteller, M., 2008. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *Biotechniques* 44, 35, 37, 39 passim.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., Amit, I., 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776–779. <https://doi.org/10.1126/science.1247651>.
- Jameson, S.A., Natarajan, A., Cool, J., DeFalco, T., Maatouk, D.M., Mork, L., Munger, S.C., Capel, B., 2012. Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS Genet.* 8, e1002575. <https://doi.org/10.1371/journal.pgen.1002575>.
- Jan, S.Z., Vormer, T.L., Jongejan, A., Röling, M., Silber, S.J., de Rooij, D.G., Hamer, G., Repping, S., van Pelt, A.M.M., 2017. Unraveling transcriptome dynamics in human spermatogenesis. *Develop. dev.* <https://doi.org/10.1242/dev.152413>, 152413.
- Jégou, B., Sankararaman, S., Rolland, A.D., Reich, D., Chalmel, F., 2017. Meiotic genes are enriched in regions of reduced archaic ancestry. *Mol. Biol. Evol.* 34, 1974–1980. <https://doi.org/10.1093/molbev/msx141>.
- Jin, S.-G., Wu, X., Li, A.X., Pfeifer, G.P., 2011. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res.* 39, 5015–5024. <https://doi.org/10.1093/nar/gkr120>.
- Johnson, J.M., Castle, J., Garrett-Engle, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., Shoemaker, D.D., 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144. <https://doi.org/10.1126/science.1090100>.

- Kalisky, T., Oriol, S., Bar-Lev, T.H., Ben-Haim, N., Trink, A., Wineberg, Y., Kanter, I., Gilad, S., Pyne, S., 2017. A brief review of single-cell transcriptomic technologies. *Brief. Funct. Genomics*. <https://doi.org/10.1093/bfpg/ekx019>.
- Katz, Y., Wang, E.T., Airoldi, E.M., Burge, C.B., 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* 7, 1009–1015. <https://doi.org/10.1038/nmeth.1528>.
- Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Beriman, B.P., Jones, P.A., 2012. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* 22, 2497–2506. <https://doi.org/10.1101/gr.143008.112>.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., Kirschner, M.W., 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201. <https://doi.org/10.1016/j.cell.2015.04.044>.
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., Megy, K., Piličeva, E., Rustici, G., Tikhonov, A., Parkinson, H., Petryszak, R., Sarkans, U., Brazdas, A., 2015. Arrayexpress update—simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116. <https://doi.org/10.1093/nar/gku1057>.
- Kotaja, N., Sassone-Corsi, P., 2007. The chromatoid body: a germ-cell-specific RNA-processing centre. *Nat. Rev. Mol. Cell Biol.* 8, 85–90. <https://doi.org/10.1038/nrm2081>.
- Kowalczykiewicz, D., Świercz, A., Handschuh, L., Leśniak, K., Figlerowicz, M., Wrzesinski, J., 2014. Characterization of *Sus scrofa* small non-coding RNAs present in both female and male gonads. *PLoS One* 9, e113249. <https://doi.org/10.1371/journal.pone.0113249>.
- Kubo, N., Toh, H., Shirane, K., Shirakawa, T., Kobayashi, H., Sato, T., Sone, H., Sato, Y., Tomizawa, S., Tsurusaki, Y., Shibata, H., Saitsu, H., Suzuki, Y., Matsumoto, N., Suyama, M., Kono, T., Ohbo, K., Sasaki, H., 2015. DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. *BMC Genom.* 16, 624. <https://doi.org/10.1186/s12864-015-1833-5>.
- Kvam, V.M., Liu, P., Si, Y., 2012. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.* 99, 248–256. <https://doi.org/10.3732/ajb.1100340>.
- Laiho, A., Kotaja, N., Gyenesei, A., Sironen, A., 2013. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *PLoS One* 8, e61558. <https://doi.org/10.1371/journal.pone.0061558>.
- Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.-L., Chen, S., Vijayaraghavan, R., Wong, J., Chen, A., Sheng, X., Kaper, F., Shen, R., Ronaghi, M., Fan, J.-B., Wang, W., Chun, J., Zhang, K., 2016. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 352, 1586–1590. <https://doi.org/10.1126/science.aaf1204>.
- Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K., Studholme, D.J., 2015. Assessing the performance of the Oxford nanopore technologies MinION. *Biomol. Detect. Quantif.* 3, 1–8. <https://doi.org/10.1016/j.bdq.2015.02.001>.
- Lee, T.-L., Cheung, H.-H., Claus, J., Sastry, C., Singh, S., Vu, L., Rennert, O., Chan, W.-Y., 2009. GermsAGE: a comprehensive SAGE database for transcript discovery on male germ cell development. *Nucleic Acids Res.* 37, D891–D897. <https://doi.org/10.1093/nar/gkn644>.
- Lee, T.-L., Rennert, O.M., Chan, W.-Y., 2012. Revealing the transcriptome landscape of mouse spermatogonial cells by tiling microarray. *Methods Mol. Biol.* 825, 75–92. https://doi.org/10.1007/978-1-61779-436-0_7.
- Lesch, B.J., Dokshin, G.A., Young, R.A., McCarrey, J.R., Page, D.C., 2013. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proc. Natl. Acad. Sci. U. S. A* 110, 16061–16066. <https://doi.org/10.1073/pnas.1315204110>.
- Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X., Zhu, X., Li, R., Yan, J., Wei, Y., Zhao, Y., Wang, W., Ren, Y., Yuan, P., Yan, Z., Hu, B., Guo, F., Wen, L., Tang, F., Qiao, J., 2017a. Single-cell RNA-seq analysis maps development of human germline cells and gonadal niche interactions. *Cell Stem Cell* 20, 891–892. <https://doi.org/10.1016/j.stem.2017.05.009>.
- Li, M., Liu, Y., Wang, T., Guan, J., Luo, Z., Chen, H., Wang, X., Chen, L., Ma, J., Mu, Z., Jiang, A., Zhu, L., Lang, Q., Zhou, X., Wang, J., Zeng, W., Li, N., Li, K., Gao, X., Li, X., 2011. Repertoire of porcine microRNAs in adult ovary and testis by deep sequencing. *Int. J. Biol. Sci.* 7, 1045–1055.
- Li, X., Ao, J., Wu, J., 2017b. Systematic identification and comparison of expressed profiles of lncRNAs and circRNAs with associated co-expression and ceRNA networks in mouse germline stem cells. *Oncotarget* 8, 26573–26590. <https://doi.org/10.18632/oncotarget.15719>.
- Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., Zamore, P.D., 2013. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol. Cell* 50, 67–81. <https://doi.org/10.1016/j.molcel.2013.02.016>.
- Li, Y., Li, J., Fang, C., Shi, L., Tan, J., Xiong, Y., Bin Fan, B., Li, C., 2016. Genome-wide differential expression of genes and small RNAs in testis of two different porcine breeds and at two different ages. *Sci. Rep.* 6, 26852. <https://doi.org/10.1038/srep26852>.
- Lian, C., Sun, B., Niu, S., Yang, R., Liu, B., Lu, C., Meng, J., Qiu, Z., Zhang, L., Zhao, Z., 2012. A comparative profile of the microRNA transcriptome in immature and mature porcine testes using Solexa deep sequencing. *FEBS J.* 279, 964–975. <https://doi.org/10.1111/j.1742-4658.2012.08480.x>.
- Lin, X., Han, M., Cheng, L., Chen, J., Zhang, Z., Shen, T., Wang, M., Wen, B., Ni, T., Han, C., 2016. Expression dynamics, relationships, and transcriptional regulations of diverse transcripts in mouse spermatogenic cells. *RNA Biol.* 13, 1011–1024. <https://doi.org/10.1080/15476286.2016.1218588>.
- Liu, G., Lei, B., Li, Y., Tong, K., Ding, Y., Luo, L., Xia, X., Jiang, S., Deng, C., Xiong, Y., Li, F., 2012. Discovery of potential piRNAs from next generation sequences of the sexually mature porcine testes. *PLoS One* 7, e34770. <https://doi.org/10.1371/journal.pone.0034770>.
- Lord, T., Oatley, J.M., 2017. A revised Asingle model to explain stem cell dynamics in the mouse male germline. *Reproduction* 154, R55–R64. <https://doi.org/10.1530/REP-17-0034>.
- Lu, H., Giordano, F., Ning, Z., 2016. Oxford nanopore MinION sequencing and genome assembly. *Dev. Reprod. Biol.* 14, 265–279. <https://doi.org/10.1016/j.jgp.2016.05.004>.
- Luk, A.C.-S., Chan, W.-Y., Rennert, O.M., Lee, T.-L., 2014. Long noncoding RNAs in spermatogenesis: insights from recent high-throughput transcriptome studies. *Reproduction* 147, R131–R141. <https://doi.org/10.1530/REP-13-0594>.
- Luk, A.C.-S., Gao, H., Xiao, S., Liao, J., Wang, D., Tu, J., Rennert, O.M., Chan, W.-Y., Lee, T.-L., 2015. GermlincRNA: a Unique Catalogue of Long Non-coding RNAs and Associated Regulations in Male Germ Cell Development. *Database (Oxford)*. 2015. <https://doi.org/10.1093/database/bav044> bav044.
- Luo, L.-F., Hou, C.-C., Yang, W.-X., 2016. Small non-coding RNAs and their associated proteins in spermatogenesis. *Gene* 578, 141–157. <https://doi.org/10.1016/j.gene.2015.12.020>.
- Luo, Z., Liu, Y., Chen, L., Ellis, M., Li, M., Wang, J., Zhang, Y., Fu, P., Wang, K., Li, X., Wang, L., 2015. microRNA profiling in three main stages during porcine spermatogenesis. *J. Assist. Reprod. Genet.* 32, 451–460. <https://doi.org/10.1007/s10815-014-0406-x>.
- Macaulay, I.C., Ponting, C.P., Voet, T., 2017. Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 33, 155–168. <https://doi.org/10.1016/j.tig.2016.12.003>.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., Trombetta, J.J., Weitz, D.A., Sanes, J.R., Shalek, A.K., Regev, A., McCarroll, S.A., 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>.
- Malone, J.H., Oliver, B., 2011. Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* 9, 34. <https://doi.org/10.1186/1741-7007-9-34>.
- Margolin, G., Khil, P.P., Kim, J., Bellani, M.A., Camerini-Otero, R.D., 2014. Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genom.* 15, 39. <https://doi.org/10.1186/1471-2164-15-39>.
- McClelland, K.S., Bell, K., Larney, C., Harley, V.R., Sinclair, A.H., Oshlack, A., Koopman, P., Bowles, J., 2015. Purification and transcriptomic analysis of mouse fetal Leydig cells reveals candidate genes for specification of gonadal steroidogenic cells1. *Biol. Reprod.* 92, 1–12. <https://doi.org/10.1095/biolreprod.115.128918>.
- McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., Shendure, J., 2016. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science (80-)* 353. <https://doi.org/10.1126/science.aaf7907> aaf7907.
- Meikar, O., Da Ros, M., Korhonen, H., Kotaja, N., 2011. Chromatoid body and small RNAs in male germ cells. *Reproduction* 142, 195–209. <https://doi.org/10.1530/REP-11-0057>.
- Meikar, O., Vagin, V.V., Chalmel, F., Söstar, K., Lardenois, A., Hammell, M., Jin, Y., Da Ros, M., Wasik, K.A., Toppari, J., Hannon, G.J., Kotaja, N., 2014. An atlas of chromatoid body components. *RNA* 20, 483–495. <https://doi.org/10.1261/rna.043729.113>.
- Meng, X., Liu, P., Jia, F., Li, J., Gao, B.-Q., 2015. De novo transcriptome analysis of portunus trituberculatus ovary and testis by RNA-Seq: identification of genes involved in gonadal development. *PLoS One* 10, e0128659. <https://doi.org/10.1371/journal.pone.0128659>.
- Mensaert, K., Denil, S., Trooskens, G., Van Criekeing, W., Thas, O., De Meyer, T., 2014. Next-generation technologies and data analytical approaches for epigenomics. *Environ. Mol. Mutagen.* 55, 155–170. <https://doi.org/10.1002/em.21841>.
- Mockler, T.C., Ecker, J.R., Sundaresan, A., Chen, H., Jacobsen, S.E., Ecker, J.R., 2005. Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 85, 1–15. <https://doi.org/10.1016/j.ygeno.2004.10.005>.
- Morozova, O., Hirst, M., Marra, M.A., 2009. Applications of new sequencing technologies for transcriptome analysis. *Annu. Rev. Genomics Hum. Genet.* 10, 135–151. <https://doi.org/10.1146/annurev-genom-082908-145957>.
- Munger, S.C., Natarajan, A., Looger, L.L., Ohler, U., Capel, B., 2013. Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination. *PLoS Genet.* 9, e1003630. <https://doi.org/10.1371/journal.pgen.1003630>.
- Mutoji, K., Singh, A., Nguyen, T., Gildersleeve, H., Kaucher, A.V., Oatley, M.J., Oatley, J.M., Velte, E.K., Geyer, C.B., Cheng, K., McCarrey, J.R., Hermann, B.P., 2016. TSPAN8 expression distinguishes spermatogonial stem cells in the prepubertal mouse testis. *Biol. Reprod.* 95. <https://doi.org/10.1095/biolreprod.116.144220>, 117–117.
- Naro, C., Jolly, A., Di Persio, S., Bielli, P., Setterblad, N., Alberdi, A.J., Vicini, E., Geremia, R., De la Grange, P., Sette, C., 2017. An Orchestrated intron retention program in meiosis controls timely usage of transcripts during germ cell differentiation. *Dev. Cell* 41, 82–93. <https://doi.org/10.1016/j.devcel.2017.03.003> e4.
- Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., Baker, J.C., Grützner, F., Kaessmann, H., 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640. <https://doi.org/10.1038/>

- nature12943.
- Nelson, N.J., 2001. Microarrays have arrived: gene expression tool matures. *J. Natl. Cancer Inst.* 93, 492–494.
- Nguyen, M.T., Delaney, D.P., Kolon, T.F., 2009. Gene expression alterations in cryptorchid males using spermatozoal microarray analysis. *Fertil. Steril.* 92, 182–187. <https://doi.org/10.1016/j.fertnstert.2008.05.043>.
- Niu, Z., Goodyear, S.M., Rao, S., Wu, X., Tobias, J.W., Avarbock, M.R., Brinster, R.L., 2011. MicroRNA-21 regulates the self-renewal of mouse spermatogonial stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 108, 12740–12745. <https://doi.org/10.1073/pnas.1109987108>.
- O'Neil, S.T., Emrich, S.J., 2013. Assessing De Novo transcriptome assembly metrics for consistency and utility. *BMC Genom.* 14, 465. <https://doi.org/10.1186/1471-2164-14-465>.
- Oakberg, E.F., 1957. Duration of spermatogenesis in the mouse. *Nature* 180, 1137–1138.
- Oakberg, E.F., 1956. Duration of spermatogenesis in the mouse and timing of stages of the cycle of the seminiferous epithelium. *Am. J. Anat.* 99, 507–516. <https://doi.org/10.1002/aja.1000990307>.
- Oatley, M.J., Kaucher, A.V., Racicot, K.E., Oatley, J.M., 2011. Inhibitor of DNA binding 4 is expressed selectively by single spermatogonia in the male germline and regulates the self-renewal of spermatogonial stem cells in mice. *Biol. Reprod.* 85, 347–356. <https://doi.org/10.1095/biolreprod.111.091330>.
- Okada, H., Tajima, A., Shichiri, K., Tanaka, A., Tanaka, K., Inoue, I., 2008. Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates ART3 in genetic susceptibility. *PLoS Genet.* 4, e26. <https://doi.org/10.1371/journal.pgen.0040026>.
- Orwig, K.E., Ryu, B.-Y., Master, S.R., Phillips, B.T., Mack, M., Avarbock, M.R., Chodosh, L., Brinster, R.L., 2008. Genes involved in post-transcriptional regulation are overrepresented in stem/progenitor spermatogonia of cryptorchid mouse testes. *Stem Cell.* 26, 927–938. <https://doi.org/10.1634/stemcells.2007-0893>.
- Park, P.J., 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10, 669–680. <https://doi.org/10.1038/nrg2641>.
- Paronetto, M.P., Sette, C., 2010. Role of RNA-binding proteins in mammalian spermatogenesis. *Int. J. Androl.* 33, 2–12. <https://doi.org/10.1111/j.1365-2605.2009.00959.x>.
- Parvinen, M., 2005. The chromatoid body in spermatogenesis. *Int. J. Androl.* 28, 189–201. <https://doi.org/10.1111/j.1365-2605.2005.00542.x>.
- Perli, S.D., Cui, C.H., Lu, T.K., 2016. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* 353. <https://doi.org/10.1126/science.aag0511>.
- Peruquetti, R.L., 2015. Perspectives on mammalian chromatoid body research. *Anim. Reprod. Sci.* 159, 8–16. <https://doi.org/10.1016/j.anireprosci.2015.05.018>.
- Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., Reik, W., 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101–1105. <https://doi.org/10.1038/nature08829>.
- Pozhitkov, A.E., Tautz, D., Noble, P.A., 2007. Oligonucleotide microarrays: widely applied poorly understood. *Briefings Funct. Genomics Proteomics* 6, 141–148. <https://doi.org/10.1093/bfpp/elm014>.
- Ramsköld, D., Wang, E.T., Burge, C.B., Sandberg, R., 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* 5, e1000598. <https://doi.org/10.1371/journal.pcbi.1000598>.
- Rastetter, R.H., Smith, C.A., Wilhelm, D., 2015. The role of non-coding RNAs in male sex determination and differentiation. *Reproduction* 150, R93–R107. <https://doi.org/10.1530/REP-15-0106>.
- Rhoads, A., Au, K.F., 2015. PacBio sequencing and its applications. *Dev. Reprod. Biol.* 13, 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>.
- Rockett, J.C., Christopher Luft, J., Brian Garges, J., Krawetz, S.A., Hughes, M.R., Hee Kirn, K., Oudes, A.J., Dix, D.J., 2001. Development of a 950-gene DNA array for examining gene expression patterns in mouse testis. *Genome Biol.* 2. RESEARCH0014.
- Rybak-Wolf, A., Stottmeister, C., Glazar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., Herzog, M., Schreyer, L., Papavasileiou, P., Ivanov, A., Ohman, M., Refojo, D., Kadener, S., Rajewsky, N., 2015. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885. <https://doi.org/10.1016/j.molcel.2015.03.027>.
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., Reich, D., 2014. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507, 354–357. <https://doi.org/10.1038/nature12961>.
- Sankararaman, S., Mallick, S., Patterson, N., Reich, D., 2016. The combined landscape of denisovan and neanderthal ancestry in present-day humans. *Curr. Biol.* 26, 1241–1247. <https://doi.org/10.1016/j.cub.2016.03.037>.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., Regev, A., 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502. <https://doi.org/10.1038/nbt.3192>.
- Schena, M., Shalun, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.
- Schmid, R., Grellscheid, S.N., Ehrmann, I., Dalgliesh, C., Danilenko, M., Paronetto, M.P., Pedrotti, S., Grellscheid, D., Dixon, R.J., Sette, C., Eperon, I.C., Elliott, D.J., 2013. The splicing landscape is globally reprogrammed during male meiosis. *Nucleic Acids Res.* 41, 10170–10184. <https://doi.org/10.1093/nar/gkt811>.
- Schmidt, S.T., Zimmerman, S.M., Wang, J., Kim, S.K., Quake, S.R., 2017. Quantitative analysis of synthetic cell lineage tracing using nuclease barcoding. *ACS Synth. Biol.* 6, 936–942. <https://doi.org/10.1021/acssynbio.6b00309>.
- Schuster, A., Tang, C., Xie, Y., Ortogero, N., Yuan, S., Yan, W., 2016. SpermBase: a database for sperm-borne RNA contents. *Biol. Reprod.* 95. <https://doi.org/10.1095/biolreprod.116.142190>, 99–99.
- Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., Reik, W., 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* 48, 849–862. <https://doi.org/10.1016/j.molcel.2012.11.001>.
- Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., Matsui, Y., 2005. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* 278, 440–458. <https://doi.org/10.1016/j.ydbio.2004.11.025>.
- Skylaki, S., Hilsenbeck, O., Schroeder, T., 2016. Challenges in long-term imaging and quantification of single-cell dynamics. *Nat. Biotechnol.* 34, 1137–1144. <https://doi.org/10.1038/nbt.3713>.
- Small, C.L., Shima, J.E., Uzumcu, M., Skinner, M.K., Griswold, M.D., 2005. Profiling Gene Expression during the Differentiation and Development of the Murine Embryonic Gonad, vol. 72, pp. 492–501. <https://doi.org/10.1095/biolreprod.104.033696>.
- Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., deMassy, B., Mikkelsen, T.S., Kaessmann, H., 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* 3, 2179–2190. <https://doi.org/10.1016/j.celrep.2013.05.031>.
- Spies, A.-N., Feig, C., Schulze, W., Chalmel, F., Cappallo-Obermann, H., Primig, M., Kirchhoff, C., 2007. Cross-platform gene expression signature of human spermatogenic failure reveals inflammatory-like response. *Hum. Reprod.* 22, 2936–2946. <https://doi.org/10.1093/humrep/dem292>.
- Stévant, I., Neirijnck, Y., Borel, C., Escoffier, J., Smith, L.B., Antonarakis, S.E., Dermitzakis, E.T., Nef, S., 2018. Deciphering cell lineage specification during male sex determination with single-cell RNA sequencing. *Cell Rep.* 22, 1589–1599. <https://doi.org/10.1016/j.celrep.2018.01.043>.
- Sun, J., Lin, Y., Wu, J., 2013. Long non-coding RNA expression profiling of mouse testis during postnatal development. *PLoS One* 8, e75750. <https://doi.org/10.1371/journal.pone.0075750>.
- Tan, T., Zhang, Y., Ji, W., Zheng, P., 2014. miRNA signature in mouse spermatogonial stem cells revealed by high-throughput sequencing. *BioMed Res. Int.* 2014. <https://doi.org/10.1155/2014/154251>, 154251.
- Tanay, A., Regev, A., 2017. Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. <https://doi.org/10.1038/nature21350>.
- Tang, W.W.C., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., Surani, M.A., 2015. A unique gene regulatory network resets the human germline epigenome for development. *Cell* 161, 1453–1467. <https://doi.org/10.1016/j.cell.2015.04.053>.
- Taylor, D.H., Chu, E.T.-J., Spektor, R., Soloway, P.D., 2015. Long non-coding RNA regulation of reproduction and development. *Mol. Reprod. Dev.* 82, 932–956. <https://doi.org/10.1002/mrd.22581>.
- van der Heijden, G.W., Eijpe, M., Baarends, W.M., 2011. The X and Y chromosome in meiosis: how and why they keep silent. *Asian J. Androl.* 13, 779–780. <https://doi.org/10.1038/aja.2011.93>.
- Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial analysis of gene expression. *Science* 270, 484–487.
- Venables, J.P., 2002. Alternative splicing in the testes. *Curr. Opin. Genet. Dev.* 12, 615–619.
- von Kopylow, K., Kirchhoff, C., Jezek, D., Schulze, W., Feig, C., Primig, M., Steinkraus, V., Spies, A.-N., 2010. Screening for biomarkers of spermatogonia within the human testis: a whole genome approach. *Hum. Reprod.* 25, 1104–1112. <https://doi.org/10.1093/humrep/deq053>.
- von Meyenn, F., Berrens, R.V., Andrews, S., Santos, F., Collier, A.J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P.J., Reik, W., 2016. Comparative principles of DNA methylation reprogramming during human and mouse in vitro primordial germ cell specification. *Dev. Cell* 39, 104–115. <https://doi.org/10.1016/j.devcel.2016.09.015>.
- Vourekas, A., Zheng, Q., Alexiou, P., Maragkakis, M., Kirino, Y., Gregory, B.D., Mourelatos, Z., 2012. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat. Struct. Mol. Biol.* 19, 773–781. <https://doi.org/10.1038/nsmb.2347>.
- Wagner, A., Regev, A., Yosef, N., 2016. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* 34, 1145–1160. <https://doi.org/10.1038/nbt.3711>.
- Walker, W.H., Delfino, F.J., Habener, J.F., 1999. RNA processing and the control of spermatogenesis. *Front. Horm. Res.* 25, 34–58.
- Wang, C., Gong, B., Bushel, P.R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., Fang, H., Hong, H., Shen, J., Su, Z., Meehan, J., Li, X., Yang, L., Li, H., Labaj, P.P., Kreil, D.P., Megherbi, D., Gaj, S., Caumont, F., van Delft, J., Kleinjans, J., Scherer, A., Devanarayan, V., Wang, J., Yang, Y., Qian, H.-R., Lancashire, L.J., Bessarabova, M., Nikolsky, Y., Furlanello, C., Chierici, M., Albanese, D., Jurman, G., Riccadonna, S., Filosi, M., Vissintainer, R., Zhang, K.K., Li, J., Hsieh, J.-H., Svoboda, D.L., Fuscoe, J.C., Deng, Y., Shi, L., Paules, R.S., Auerbach, S.S., Tong, W., 2014. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* 32, 926–932. <https://doi.org/10.1038/nbt.3001>.
- Wang, K.C., Chang, H.Y., 2011. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* 43, 904–914. <https://doi.org/10.1016/j.molcel.2011.08.018>.

- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. <https://doi.org/10.1038/nrg2484>.
- Wichman, L., Somasundaram, S., Breindel, C., Valerio, D.M., McCarrey, J.R., Hodges, C.A., Khalil, A.M., 2017. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. *Biol. Reprod.* 97, 313–323. <https://doi.org/10.1093/biolre/iox084>.
- Wu, Z., Wu, H., 2016. *Experimental Design and Power Calculation for RNA-seq Experiments*. Humana Press, New York, NY, pp. 379–390. https://doi.org/10.1007/978-1-4939-3578-9_18.
- Yadav, R.P., Kotaja, N., 2014. Small RNAs in spermatogenesis. *Mol. Cell. Endocrinol.* 382, 498–508. <https://doi.org/10.1016/j.mce.2013.04.015>.
- Yang, Q., Hua, J., Wang, L., Xu, B., Zhang, H., Ye, N., Zhang, Z., Yu, D., Cooke, H.J., Zhang, Y., Shi, Q., 2013. MicroRNA and piRNA profiles in normal human testis detected by next generation sequencing. *PLoS One* 8, e66809. <https://doi.org/10.1371/journal.pone.0066809>.
- Yuan, G.-C., Cai, L., Elowitz, M., Enver, T., Fan, G., Guo, G., Irizarry, R., Kharchenko, P., Kim, J., Orkin, S., Quackenbush, J., Saadatpour, A., Schroeder, T., Shivdasani, R., Tirosh, I., 2017. Challenges and emerging directions in single-cell analysis. *Genome Biol.* 18, 84. <https://doi.org/10.1186/s13059-017-1218-y>.
- Zhang, P., Kang, J.-Y., Gou, L.-T., Wang, J., Xue, Y., Skogerboe, G., Dai, P., Huang, D.-W., Chen, R., Fu, X.-D., Liu, M.-F., He, S., 2015. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res.* 25, 193–207. <https://doi.org/10.1038/cr.2015.4>.
- Zhu, Z.-J., Yang, S., Li, Z., 2015. Transcriptome research on spermatogenic molecular drive in mammals. *Asian J. Androl.* 17, 961–971. <https://doi.org/10.4103/1008-682X.159721>.
- Zhu, Z., Li, C., Yang, S., Tian, R., Wang, J., Yuan, Q., Dong, H., He, Z., Wang, S., Li, Z., 2016. Dynamics of the transcriptome during human spermatogenesis: predicting the potential key genes regulating male gametes generation. *Sci. Rep.* 6, 19069. <https://doi.org/10.1038/srep19069>.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., Enard, W., 2017. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643. <https://doi.org/10.1016/j.molcel.2017.01.023> e4.
- Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A.M., Mazutis, L., 2016. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* 12, 44–73. <https://doi.org/10.1038/nprot.2016.154>.
- Zimmermann, C., Stévant, I., Borel, C., Conne, B., Pitetti, J.-L., Calvel, P., Kaessmann, H., Jégou, B., Chalmel, F., Nef, S., 2015. Research resource: the dynamic transcriptional profile of sertoli cells during the progression of spermatogenesis. *Mol. Endocrinol.* 29, 627–642. <https://doi.org/10.1210/me.2014-1356>.

6. Conclusion

La différenciation des gonades fœtales humaines est un processus fin et complexe. À partir d'une gonade bipotentielle, des mécanismes de régulation transcriptionnelle permettent à l'organe de s'engager sur l'une des deux trajectoires antagonistes possibles, vers le testicule ou l'ovaire. Un grain de sable dans cette machinerie, comme une mutation non fonctionnelle d'un gène, peut provoquer des cas d'inversion du sexe, le sexe génétique ne correspondant plus au sexe phénotypique. À la suite de ce moment clé, de nombreux processus de différenciation et de développement cellulaires entrent en jeu afin de former le testicule et l'ovaire. Ces organes acquièrent des spécificités dans leurs modes de production de gamètes et d'hormones, mais resteront indissociables quant à leur parfaite complémentarité pour la reproduction.

Les travaux de mon équipe d'accueil s'attachent depuis plusieurs années à décrypter la physiologie du tractus urogénital des mammifères, comprenant les souris et les rats, et à évaluer les effets de perturbations induites par des produits chimiques ou des virus. Depuis 2012, l'équipe a reçu l'accréditation permettant l'accès aux produits d'avortements de l'Hôpital sud de Rennes, ce qui a permis des avancées notables sur l'étude des gonades fœtales humaines et leurs perturbations.

Mon projet de thèse s'est réparti en trois principaux axes de recherche. Le premier a abordé la différenciation et le développement des testicules et ovaires fœtaux humains. Grâce à une technologie de transcriptomique haut-débit, le RNA-seq, j'ai pu décrire le programme transcriptionnel gouvernant les divers mécanismes de différenciation cellulaire ayant lieu entre la 6^{ème} et la 17^{ème} semaine de développement au sein des deux gonades. Le deuxième axe de ma thèse avait pour but de décrire les mécanismes d'action de différents composés suspectés d'être des perturbateurs endocriniens sur le testicule fœtal humain, toujours grâce au RNA-seq. Enfin, j'ai participé au développement d'une nouvelle version du ReproGenomic Viewer, un outil de visualisation inter-espèces de données issues de diverses technologies haut-débit.

La première partie de cette thèse s'est donc attachée à décrypter le programme transcriptionnel qui permettra, à partir d'une gonade indifférenciée, de former des organes fonctionnels essentiels pour la perpétuation de l'espèce. L'étude longitudinale du développement du testicule et de l'ovaire est inédite en bien des aspects, considérant les tissus humains étudiés, la large fenêtre de 12 semaines de développement couverte (entre 6 et 17 SD) et la sensibilité de description au niveau moléculaire.

Les résultats bruts du RNA-seq ont été raffinés afin de discriminer 35 194 transcrits exprimés dans nos échantillons, parmi lesquels des transcrits codant pour des protéines, des longs ARN non codants ainsi que des nouveaux gènes. Une approche intégrant des données de protéomique de gonades

foétales humaines avec nos données de transcriptomique (« Proteomics informed by Transcriptomics ») m'a permis d'identifier certains peptides appartenant à la séquence protéique prédite de nos transcrits, ce qui a mis en évidence le caractère codant de certains gènes non-codants. Cette procédure a également permis de confirmer que les nouveaux gènes identifiés ne codaient pas pour des protéines.

Après une filtration statistique, ces transcrits séquencés ont été définis comme « sexuellement dimorphique », c'est-à-dire qu'ils sont préférentiellement exprimés dans le testicule ou dans l'ovaire, ou « sexuellement non dimorphique », leur niveau d'expression variant au cours du développement de la gonade, de manière similaire dans les deux sexes. Une classification des gènes sur la base de leurs profils d'expression a permis de définir 20 groupes de gènes co-exprimés et une analyse fonctionnelle a permis d'associer chacun de ces groupes à certains processus biologiques.

La suite de l'analyse des données que j'ai effectué s'est orientée vers la caractérisation des phénomènes les plus précoces de la différenciation du testicule et de l'ovaire, tentant de définir l'ensemble des gènes et transcrits impliqués dans la différenciation sexuelle ou dans la cascade d'évènements transcriptionnels subséquente. Parmi l'ensemble des gènes différenciellement exprimés à des stades précoces, certains d'entre eux présentent un profil similaire à celui de SRY, principal acteur de la différenciation testiculaire connu. De même, certains gènes présentaient une sur-expression dans l'ovaire dès la 6^{ème} semaine de développement. Depuis quelques années, le dogme selon lequel la différenciation ovarienne est une voie « par défaut » de la différenciation gonadique est constamment remis en cause, grâce à la découverte de gènes « anti-testicule » et « pro-ovarien », tel que NR2F2 (COUPTF2) (Bashamboo et al., 2018; Zhao et al., 2017). De fait, le programme transcriptionnel femelle reste très mystérieux et globalement moins étudié que la trajectoire du transcriptome testiculaire. Certains nouveaux gènes ont également été décrits comme exprimés de manière très précoce dans le testicule et dans l'ovaire, et requièrent de plus amples investigations et validations.

Dans l'optique d'intégrer nos données avec d'autres études haut-débit, j'ai également développé une méthode afin d'étudier la conservation du développement gonadique. Cette étude de corrélation des profils d'expression des gènes au cours du développement des gonades humaines et souris m'a permis d'apprécier le socle de gènes communs aux deux espèces, dont la fonction au sein du développement gonadique est potentiellement conservée. Les résultats obtenus nous encouragent à persévérer dans cette voie, en intégrant de nouvelles données de RNA-seq (Zhao et al., 2018). La souris reste l'un des modèles les mieux maîtrisés pour étudier la biologie et la physiologie des mammifères. Le travail sur ces organismes possède de nombreux avantages : la parfaite maîtrise du génome, de l'âge des échantillons, la possibilité d'effectuer des manipulations génétiques (*Knock-out*, *Crisper-Cas9*...) ou toxicologiques,

et la rapidité de la reproduction. Cette grande maîtrise du modèle permet d'étudier de nombreux processus biologiques qu'il serait difficile d'observer dans les organes humains. Considérant cela, ma démarche s'inscrit dans une étude évolutive des processus de différenciation gonadique, afin de mettre en évidence les différences interspèces, mais également de mettre en exergue les similarités afin de donner plus de robustesse aux extrapolations entre les rongeurs et l'Homme.

Enfin, l'exploitation des résultats de cette étude transcriptomique de grande envergure n'est pas terminée. En effet, elle figure comme une ressource riche pour l'étude des processus de différenciation plus tardifs, il faudra plusieurs années afin de pleinement l'exploiter. Considérant les limites techniques du travail sur des organes entiers, le projet de recherche de l'équipe s'est naturellement orienté vers une description à l'échelle de la cellule unique du développement gonadique par des approches de scRNA-seq. La complémentarité de ces deux technologies va permettre des avancées toujours plus significatives dans la compréhension de la différenciation des gonades humaines.

La seconde partie de cette thèse s'est attachée à décrire les mécanismes transcriptionnels altérés suite à l'exposition des testicules fœtaux humains à deux composés suspectés d'être des perturbateurs endocriniens, grâce au RNA-seq. Le protocole de culture organotypique de testicule et d'ovaire humains développé par mon équipe permet d'évaluer la toxicité des substances d'intérêt sur les populations cellulaires des gonades et l'impact sur la stéroïdogenèse. Les résultats préliminaires de deux études toxicogénomique d'envergure ont permis de mettre en évidence plusieurs points. D'une part, l'ibuprofène et le chlordécone sont deux composés induisant une baisse significative du niveau de production de testostérone ainsi qu'une forte perturbation du programme d'expression du testicule fœtal. Cela a permis souligner que, si la cellule de Sertoli a un rôle central dans la différenciation testiculaire, sécrétant des facteurs paracrines et hormonaux permettant de verrouiller la différenciation de la gonade vers son destin masculin, elle apparaît également comme une cible potentielle des deux substances testées. De fait, une perturbation de l'activité de cette clé de voute du développement testiculaire pourrait entraîner une cascade de conséquences délétères pour l'organe en développement. Les signatures toxicogénomiques de l'ibuprofène et du chlordécone demandent de plus amples investigations afin de décrypter leurs mécanismes d'action. Les résultats globaux de ces deux projets de toxicogénomiques ont également mis en évidence l'absolue nécessité d'augmenter le nombre de réplicats biologiques pour pallier à la variabilité inter-individuelle importante des organes fœtaux humains. En effet, 9 des 11 substances testées n'ont pas engendré d'altération du transcriptome assez homogène entre les explants testiculaires pour pouvoir décrire une signature toxicogénomique robuste. Les technologies haut-débit

apparaissent toutes désignées pour enfin comprendre l'effet de notre environnement sur notre physiologie, mais les protocoles doivent être adaptés au travail sur du tissu humain.

Dans la dernière partie de ma thèse, j'ai participé au développement d'une nouvelle version du Reprogenomic Viewer (Darde et al., 2015). L'une des préoccupations constantes de mon travail a été de tirer le meilleur parti des différentes technologies haut-débit et des études effectuées dans les autres laboratoires afin d'affiner ma compréhension des mécanismes en jeu dans la différenciation gonadique. Dans cette démarche d'intégration des connaissances générées dans le domaine, le recensement des études a permis de mettre en exergue la quantité impressionnante de données haut-débit générées dans le domaine de la reproduction et la diversité des projets développés. Cette nouvelle version du Reprogenomic Viewer a donc pour but de proposer un contenu aussi exhaustif que possible afin de fournir une ressource précieuse couvrant un large spectre de la recherche en reproduction.

De manière plus personnelle, ce travail effectué au cours de ces trois dernières années m'a montré la grande complexité de la physiologie de la reproduction, sous différents aspects. D'une part, la différenciation des gonades fœtales, humaines ou d'autres espèces, est un processus issu de nombreuses adaptations évolutives, qui s'est diversifié à travers le règne animal. Ce processus fondamental pour la survie des espèces m'apparaît comme un mécanisme fascinant dont de nombreux mystères restent à découvrir. D'autre part, l'évolution des dogmes en biologie moléculaire et la découverte de nouvelles classes d'ARN laisse présager de nouveaux défis quant à notre compréhension des processus biologiques, impliquant de nouveaux systèmes de régulations. La bio-informatique est un domaine en constante mutation, conséquence des nombreux développements technologiques actuels. La complémentarité des nouvelles technologies et des approches plus classiques du travail en laboratoire permettra des avancées toujours plus grandes dans tous les domaines de la recherche.

Considérant le contexte plus global de ma thèse, les formations interdisciplinaires dispensées par l'École des Hautes Études en Santé Publique (EHESP) sur l'évaluation des exposition des populations, l'épidémiologie et l'évaluation des risques m'ont permis d'apprécier la complémentarité du travail de chaque discipline dans le cadre plus large de la santé publique. La rencontre avec les autres doctorants faisant partie du réseau doctoral de l'EHESP m'a également permis de comprendre l'ampleur des défis actuels, de l'étude de la physiologie des Hommes, de l'interaction avec leur environnement, jusqu'à la mise en place de politiques de santé publique cohérentes avec les aspects scientifiques, éthiques et législatifs. Chaque domaine m'est apparu comme une pierre à l'édifice du mieux vivre. Les bouleversements environnementaux dont nous sommes témoins, comprenant les changements

climatiques, mais également la pollution environnementale croissante, des sols, des nappes phréatiques et plus généralement de la biosphère, nécessitent un effort collectif de tous les domaines de recherche.

Références

- Achermann, J.C., Ito, M., Ito, M., Hindmarsh, P.C., and Jameson, J.L. (1999). A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. *Nat. Genet.* *22*, 125–126.
- Adami, H.O., Bergström, R., Möhner, M., Zatoński, W., Storm, H., Ekblom, A., Tretli, S., Teppo, L., Ziegler, H., and Rahu, M. (1994). Testicular cancer in nine northern European countries. *Int. J. Cancer* *59*, 33–38.
- Adams, I.R., and McLaren, A. (2002). Sexually dimorphic development of mouse primordial germ cells: switching from oogenesis to spermatogenesis. *Dev. Camb. Engl.* *129*, 1155–1164.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., and Moreno, R.F. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* *252*, 1651–1656.
- Akbas, G.E., Song, J., and Taylor, H.S. (2004). A HOXA10 estrogen response element (ERE) is differentially regulated by 17 beta-estradiol and diethylstilbestrol (DES). *J. Mol. Biol.* *340*, 1013–1023.
- Alano, M.A., Ngougma, E., Ostrea, E.M., and Konduri, G.G. (2001). Analysis of nonsteroidal antiinflammatory drugs in meconium and its relation to persistent pulmonary hypertension of the newborn. *Pediatrics* *107*, 519–523.
- Albert, O., and Jégou, B. (2014). A critical assessment of the endocrine susceptibility of the human testis to phthalates from fetal life to adulthood. *Hum. Reprod. Update* *20*, 231–249.
- Albrecht, K.H., and Eicher, E.M. (2001). Evidence That Sry Is Expressed in Pre-Sertoli Cells and Sertoli and Granulosa Cells Have a Common Precursor. *Dev. Biol.* *240*, 92–107.
- Allen, F.W. (1941). The Biochemistry of the Nucleic Acids, Purines, and Pyrimidines. *Annu. Rev. Biochem.* *10*, 221–244.
- Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* *17*, 487–500.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* *11*, R106.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* *31*, 166–169.
- Andersen, A.G., Jensen, T.K., Carlsen, E., Jørgensen, N., Andersson, A.M., Krarup, T., Keiding, N., and Skakkebaek, N.E. (2000). High frequency of sub-optimal semen quality in an unselected population of young men. *Hum. Reprod.* *15*, 366–372.
- Andersen, H.R., Schmidt, I.M., Grandjean, P., Jensen, T.K., Budtz-Jørgensen, E., Kjaerstad, M.B., Baelum, J., Nielsen, J.B., Skakkebaek, N.E., and Main, K.M. (2008). Impaired reproductive development in sons of women occupationally exposed to pesticides during pregnancy. *Environ. Health Perspect.* *116*, 566–572.
- Anderson, R.A., Fulton, N., Cowan, G., Coutts, S., and Saunders, P.T. (2007). Conserved and divergent patterns of expression of DAZL, VASA and OCT4 in the germ cells of the human fetal ovary and testis. *BMC Dev. Biol.* *7*, 136.
- Angelini, C., and Costa, V. (2014). Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. *Front. Cell Dev. Biol.* *2*, 51.
- Aparicio, O., Geisberg, J.V., and Struhl, K. (2004). Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo. *Curr. Protoc. Cell Biol. Chapter 17*, Unit 17.7.
- Ara, T., Nakamura, Y., Egawa, T., Sugiyama, T., Abe, K., Kishimoto, T., Matsui, Y., and Nagasawa, T. (2003). Impaired colonization of the gonads by primordial germ cells in mice lacking a chemokine, stromal cell-derived factor-1 (SDF-1). *Proc. Natl. Acad. Sci. U. S. A.* *100*, 5319–5323.

- Arango, N.A., Lovell-Badge, R., and Behringer, R.R. (1999). Targeted Mutagenesis of the Endogenous Mouse *Mis* Gene Promoter: In Vivo Definition of Genetic Pathways of Vertebrate Sexual Development. *Cell* 99, 409–419.
- Ascoli, M., Fanelli, F., and Segaloff, D.L. (2002). A 2002 Perspective. 141–174.
- Asmann, Y.W., Klee, E.W., Thompson, E.A., Perez, E.A., Middha, S., Oberg, A.L., Therneau, T.M., Smith, D.I., Poland, G.A., Wieben, E.D., et al. (2009). 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genomics* 10, 531.
- Auger, J., and Jouannet, P. (1997). Evidence for regional differences of semen quality among fertile French men. *Fédération Française des Centres d'Etude et de Conservation des Oeufs et du Sperme humains*. *Hum. Reprod. Oxf. Engl.* 12, 740–745.
- Auger, J., Kunstmann, J.M., Czyglik, F., and Jouannet, P. (1995). Decline in Semen Quality among Fertile Men in Paris during the Past 20 Years. *N. Engl. J. Med.* 332, 281–285.
- Auguste, A., Chassot, A.-A., Grégoire, E.P., Renault, L., Pannetier, M., Treier, M., Pailhoux, E., and Chaboissier, M.-C. (2011). Loss of R-Spondin1 and *Foxl2* Amplifies Female-to-Male Sex Reversal in XX Mice. *Sex. Dev.* 5, 304–317.
- Avery, O.T., MacLeod, C.M., and McCarty, M. (1944). STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES. *J. Exp. Med.* 79, 137–158.
- Ayers, K.L., Davidson, N.M., Demiyah, D., Roeszler, K.N., Grützner, F., Sinclair, A.H., Oshlack, A., and Smith, C.A. (2013). RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol.* 14, R26.
- Ayers, K.L., Lambeth, L.S., Davidson, N.M., Sinclair, A.H., Oshlack, A., and Smith, C.A. (2015). Identification of candidate gonadal sex differentiation genes in the chicken embryo using RNA-seq. *BMC Genomics* 16, 704.
- Baarends, W.M., van Helmond, M.J., Post, M., van der Schoot, P.J., Hoogerbrugge, J.W., de Winter, J.P., Uilenbroek, J.T., Karels, B., Wilming, L.G., and Meijers, J.H. (1994). A novel member of the transmembrane serine/threonine kinase receptor family is specifically expressed in the gonads and in mesenchymal cells adjacent to the müllerian duct. *Dev. Camb. Engl.* 120, 189–197.
- Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex Determination: Why So Many Ways of Doing It? *PLoS Biol.* 12, e1001899.
- Baerwald, A.R., Adams, G.P., and Pierson, R.A. (2012). Ovarian antral folliculogenesis during the human menstrual cycle: a review. *Hum. Reprod. Update* 18, 73–91.
- Baetu, T.M. (2012). Genes after the human genome project. *Stud. Hist. Philos. Biol. Biomed. Sci.* 43, 191–201.
- Bagheri-Fam, S., Ono, M., Li, L., Zhao, L., Ryan, J., Lai, R., Katsura, Y., Rossello, F.J., Koopman, P., Scherer, G., et al. (2015). *FGFR2* mutation in 46,XY sex reversal with craniosynostosis. *Hum. Mol. Genet.* 24, 6699–6710.
- Bahr, J.M. (2018). Ovary, Overview. In *Encyclopedia of Reproduction (Second Edition)*, M.K. Skinner, ed. (Oxford: Academic Press), pp. 3–7.
- Ballouz, S., Verleyen, W., and Gillis, J. (2015). Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* 31, 2123–2130.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995.
- Barrionuevo, F., Bagheri-Fam, S., Klattig, J., Kist, R., Taketo, M.M., Englert, C., and Scherer, G. (2006). Homozygous Inactivation of *Sox9* Causes Complete XY Sex Reversal in Mice. *Biol. Reprod.* 74, 195–201.

- Bashamboo, A., Eozenou, C., Rojo, S., and McElreavey, K. (2017). Anomalies in human sex determination provide unique insights into the complex genetic interactions of early gonad development. *Clin. Genet.* *91*, 143–156.
- Bashamboo, A., Eozenou, C., Jorgensen, A., Bignon-Topalovic, J., Siffroi, J.-P., Hyon, C., Tar, A., Nagy, P., Sólyom, J., Halász, Z., et al. (2018). Loss of Function of the Nuclear Receptor NR2F2, Encoding COUP-TF2, Causes Testis Development and Cardiac Defects in 46,XX Children. *Am. J. Hum. Genet.* *102*, 487–493.
- Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., et al. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* *45*, D158–D169.
- Battle, A., Mostafavi, S., Zhu, X., Potash, J.B., Weissman, M.M., McCormick, C., Haudenschild, C.D., Beckman, K.B., Shi, J., Mei, R., et al. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* *24*, 14–24.
- Beadle, G.W., and Tatum, E.L. (1941). Genetic Control of Biochemical Reactions in *Neurospora*. *Proc. Natl. Acad. Sci. U. S. A.* *27*, 499–506.
- Behringer, R.R., Cate, R.L., Froelick, G.J., Palmiter, R.D., and Brinster, R.L. (1990). Abnormal sexual development in transgenic mice chronically expressing müllerian inhibiting substance. *Nature*.
- Behringer, R.R., Finegold, M.J., and Cate, R.L. (1994). Mullerian-inhibiting substance function during mammalian sexual development. *Cell* *79*, 415–425.
- Ben Maamar, M., Lesné, L., Desdoits-Lethimonier, C., Coiffec, I., Lassurguère, J., Lavoué, V., Deceuninck, Y., Antignac, J.-P., Le Bizec, B., Perdu, E., et al. (2015). An investigation of the endocrine-disruptive effects of bisphenol a in human and rat fetal testes. *PLoS One* *10*, e0117226.
- Ben Maamar, M., Lesné, L., Hennig, K., Desdoits-Lethimonier, C., Kilcoyne, K.R., Coiffec, I., Rolland, A.D., Chevrier, C., Kristensen, D.M., Lavoué, V., et al. (2017). Ibuprofen results in alterations of human fetal testis development. *Sci. Rep.* *7*, 44184.
- Bendel-Stenzel, M., Anderson, R., Heasman, J., and Wylie, C. (1998). The origin and migration of primordial germ cells in the mouse. *Semin. Cell Dev. Biol.* *9*, 393–400.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* *456*, 53–59.
- Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). *Protein Structure and Function*. Biochem. 5th Ed.
- Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U. S. A.* *74*, 3171–3175.
- Bergman, J.E.H., Loane, M., Vrijheid, M., Pierini, A., Nijman, R.J.M., Addor, M.-C., Barisic, I., Béres, J., Braz, P., Budd, J., et al. (2015). Epidemiology of hypospadias in Europe: a registry-based study. *World J. Urol.* *33*, 2159–2167.
- Bergstrom, D.E., Young, M., Albrecht, K.H., and Eicher, E.M. (2000). Related function of mouse SOX3, SOX9, and SRY HMG domains assayed by male sex determination. *Genesis* *28*, 111–124.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* *28*, 235–242.
- Bernard, D.J., Schang, G., Li, Y., Ongaro, L., and Thompson, T.B. (2018). Activins and Inhibins in Female Reproduction. In *Encyclopedia of Reproduction (Second Edition)*, M.K. Skinner, ed. (Oxford: Academic Press), pp. 202–210.
- Beverdam, A., and Koopman, P. (2006). Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. *Hum. Mol. Genet.* *15*, 417–431.

- Biason-Lauber, A., Konrad, D., Navratil, F., and Schoenle, E.J. (2004). A *WNT4* Mutation Associated with Müllerian-Duct Regression and Virilization in a 46,XX Woman. *N. Engl. J. Med.* *351*, 792–798.
- Biason-Lauber, A., Konrad, D., Meyer, M., deBeaufort, C., and Schoenle, E.J. (2009). Ovaries and Female Phenotype in a Girl with 46,XY Karyotype and Mutations in the *CBX2* Gene.
- Bilezikjian, L.M., Blount, A.L., Leal, A.M.O., Donaldson, C.J., Fischer, W.H., and Vale, W.W. (2004). Autocrine/paracrine regulation of pituitary function by activin, inhibin and follistatin. *Mol. Cell. Endocrinol.* *225*, 29–36.
- Birk, O.S., Casiano, D.E., Wassif, C.A., Cogliati, T., Zhao, L., Zhao, Y., Grinberg, A., Huang, S., Kreidberg, J.A., Parker, K.L., et al. (2000). The LIM homeobox gene *Lhx9* is essential for mouse gonad formation. *Nature* *403*, 909–913.
- Bishop, C.E., Whitworth, D.J., Qin, Y., Agoulnik, A.I., Agoulnik, I.U., Harrison, W.R., Behringer, R.R., and Overbeek, P.A. (2000). A transgenic insertion upstream of *Sox9* is associated with dominant XX sex reversal in the mouse. *Nat. Genet.* *26*, 490–494.
- Boitani, C., Persio, S.D., Esposito, V., and Vicini, E. (2016). Spermatogonial cells: mouse, monkey and man comparison. *Semin. Cell Dev. Biol.* *59*, 79–88.
- Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinforma. Oxf. Engl.* *19*, 185–193.
- Boulanger, L., Pannetier, M., Gall, L., Allais-Bonnet, A., Elzaïat, M., Le Bourhis, D., Daniel, N., Richard, C., Cotinot, C., Ghyselinck, N.B., et al. (2014). *FOXL2* Is a Female Sex-Determining Gene in the Goat. *Curr. Biol.* *24*, 404–408.
- Bouma, G.J., Washburn, L.L., Albrecht, K., and Eicher, E.M. (2007). Correct dosage of *Fog2* and *Gata4* transcription factors is critical for fetal testis development in mice. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 14994–14999.
- Bourdon-Lacombe, J.A., Moffat, I.D., Deveau, M., Husain, M., Auerbach, S., Krewski, D., Thomas, R.S., Bushel, P.R., Williams, A., and Yauk, C.L. (2015). Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals. *Regul. Toxicol. Pharmacol.* *RTP 72*, 292–309.
- Bowles, J., and Koopman, P. (2010). Sex determination in mammalian germ cells: extrinsic versus intrinsic factors. *Reprod. Camb. Engl.* *139*, 943–958.
- Bowles, J., Knight, D., Smith, C., Wilhelm, D., Richman, J., Mamiya, S., Yashiro, K., Chawengsaksophak, K., Wilson, M.J., Rossant, J., et al. (2006). Retinoid signaling determines germ cell fate in mice. *Science* *312*, 596–600.
- Bowles, J., Feng, C.-W., Spiller, C., Davidson, T.-L., Jackson, A., and Koopman, P. (2010). *FGF9* suppresses meiosis and promotes male germ cell fate in mice. *Dev. Cell* *19*, 440–449.
- Brauner, R., Picard-Dieval, F., Lottmann, H., Rouget, S., Bignon-Topalovic, J., Bashamboo, A., and McElreavey, K. (2016). Familial forms of disorders of sex development may be common if infertility is considered a comorbidity. *BMC Pediatr.* *16*, 195.
- Brennan, J., Tilmann, C., and Capel, B. (2003). *Pdgfr-alpha* mediates testis cord organization and fetal Leydig cell development in the XY gonad. *Genes Dev.* *17*, 800–810.
- Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* *190*, 576–581.
- Breschi, A., Gingeras, T.R., and Guigó, R. (2017). Comparative transcriptomics in human and mouse. *Nat. Rev. Genet.* *18*, 425–440.
- Briño-Enríquez, M.A., Reig-Viader, R., Cabero, L., Toran, N., Martínez, F., Roig, I., and Garcia Caldés, M. (2012). Gene expression is altered after bisphenol A exposure in human fetal oocytes in vitro. *Mol. Hum. Reprod.* *18*, 171–183.

- Bromer, J.G., Wu, J., Zhou, Y., and Taylor, H.S. (2009). Hypermethylation of homeobox A10 by in utero diethylstilbestrol exposure: an epigenetic mechanism for altered developmental programming. *Endocrinology* 150, 3376–3382.
- Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R., et al. (2015). Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.* 43, D36–D42.
- Brucker-Davis, F., Wagner-Mahler, K., Delattre, I., Ducot, B., Ferrari, P., Bongain, A., Kurzenne, J.-Y., Mas, J.-C., and Fenichel, P. (2008). Cryptorchidism at birth in Nice area (France) is associated with higher prenatal exposure to PCBs and DDE, as assessed by colostrum concentrations. *Hum. Reprod.* 23, 1708–1718.
- Buck Louis, G.M., Cooney, M.A., and Peterson, C.M. (2011). The ovarian dysgenesis syndrome. *J. Dev. Orig. Health Dis.* 2, 25–35.
- Bulyk, M.L. (2006). DNA microarray technologies for measuring protein-DNA interactions. *Curr. Opin. Biotechnol.* 17, 422–430.
- Callier, P., Calvel, P., Matevossian, A., Makrythanasis, P., Bernard, P., Kurosaka, H., Vannier, A., Thauvin-Robinet, C., Borel, C., Mazaud-Guittot, S., et al. (2014). Loss of function mutation in the palmitoyl-transferase HHAT leads to syndromic 46,XY disorder of sex development by impeding Hedgehog protein palmitoylation and signaling. *PLoS Genet.* 10, e1004340.
- Cameron, F.J., Hageman, R.M., Cooke-Yarborough, C., Kwok, C., Goodwin, L.L., Sillence, D.O., and Sinclair, A.H. (1996). A novel germ line mutation in SOX9 causes familial campomelic dysplasia and sex reversal. *Hum. Mol. Genet.* 5, 1625–1630.
- Cammack, R.C., Atwood, T.A., Campbell, P.C., Parish, H.P., Smith, A.S., Vella, F.V., and Stirling, J.S. (2006). Thymus nucleic acid. In *Oxford Dictionary of Biochemistry and Molecular Biology*, (Oxford University Press), p.
- Campbell, D.M., Webb, J.A., and Hargreave, T.B. (1987). Cryptorchidism in Scotland. *Br. Med. J. Clin. Res. Ed* 295, 1235–1236.
- Canto, P., Söderlund, D., Reyes, E., and Méndez, J.P. (2004). Mutations in the desert hedgehog (DHH) gene in patients with 46,XY complete pure gonadal dysgenesis. *J. Clin. Endocrinol. Metab.* 89, 4480–4483.
- Canto, P., Vilchis, F., Söderlund, D., Reyes, E., and Méndez, J.P. (2005). A heterozygous mutation in the desert hedgehog gene in patients with mixed gonadal dysgenesis. *Mol. Hum. Reprod.* 11, 833–836.
- Cao, G., Zhao, L., Stangl, H., Hasegawa, T., Richardson, J. a, Parker, K.L., and Hobbs, H.H. (1999). Developmental and hormonal regulation of murine scavenger receptor, class B, type 1. *Mol. Endocrinol. Baltim. Md* 13, 1460–1473.
- Carlberg, C., and Molnár, F. (2018). What Is Epigenomics? In *Human Epigenomics*, (Springer, Singapore), pp. 3–18.
- Carlsen, E., Giwercman, A., Keiding, N., and Skakkebaek, N.E. (1992). Evidence for decreasing quality of semen during past 50 years. *BMJ* 305, 609–613.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.
- Carr, B.R., Parker, C.R., Ohashi, M., MacDonald, P.C., and Simpson, E.R. (1983). Regulation of human fetal testicular secretion of testosterone: low-density lipoprotein-cholesterol and cholesterol synthesized de novo as steroid precursor. *Am. J. Obstet. Gynecol.* 146, 241–247.
- Carré, G.-A., Siggers, P., Xipolita, M., Brindle, P., Lutz, B., Wells, S., and Greenfield, A. (2017). Loss of p300 and CBP disrupts histone acetylation at the mouse *Sry* promoter and causes XY gonadal sex reversal. *Hum. Mol. Genet.*
- Carson, R. (1962). *Silent spring* (Houghton Mifflin).

- Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell* 157, 77–94.
- Chaboissier, M.-C. (2004). Functional analysis of Sox8 and Sox9 during sex determination in the mouse. *Development* 131, 1891–1901.
- Chalmel, F., and Primig, M. (2008). The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology. *BMC Bioinformatics* 9, 86.
- Chalmel, F., Lardenois, a., Evrard, B., Rolland, a. D., Sallou, O., Dumargne, M.-C., Coiffec, I., Collin, O., Primig, M., and Jegou, B. (2014). High-Resolution Profiling of Novel Transcribed Regions During Rat Spermatogenesis. *Biol. Reprod.* 91, 5–5.
- Chang, H., Gao, F., Guillou, F., Taketo, M.M., Huff, V., and Behringer, R.R. (2008). Wt1 negatively regulates beta-catenin signaling during testis development. *Dev. Camb. Engl.* 135, 1875–1885.
- Chassot, A.-A., Ranc, F., Gregoire, E.P., Roepers-Gajadien, H.L., Taketo, M.M., Camerino, G., de Rooij, D.G., Schedl, A., and Chaboissier, M.-C. (2008). Activation of beta-catenin signaling by Rspo1 controls differentiation of the mammalian ovary. *Hum. Mol. Genet.* 17, 1264–1277.
- Chassot, A.-A., Bradford, S.T., Auguste, A., Gregoire, E.P., Pailhoux, E., de Rooij, D.G., Schedl, A., and Chaboissier, M.-C. (2012). WNT4 and RSPO1 together are required for cell proliferation in the early mouse gonad. *Dev. Camb. Engl.* 139, 4461–4472.
- Chen, E.-Q., Bai, L., Gong, D.-Y., and Tang, H. (2015). Employment of digital gene expression profiling to identify potential pathogenic and therapeutic targets of fulminant hepatic failure. *J. Transl. Med.* 13, 22.
- Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., Zhang, Q., Yan, G., and Cui, Q. (2013). LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* 41, D983-986.
- Chen, M., Zhang, L., Cui, X., Lin, X., Li, Y., Wang, Y., Wang, Y., Qin, Y., Chen, D., Han, C., et al. (2017a). Wt1 directs the lineage specification of sertoli and granulosa cells by repressing Sfl expression. *Development*.
- Chen, X., Li, X., Guo, J., Zhang, P., and Zeng, W. (2017b). The roles of microRNAs in regulation of mammalian spermatogenesis. *J. Anim. Sci. Biotechnol.* 8, 35.
- Chia, V.M., Quraishi, S.M., Devesa, S.S., Purdue, M.P., Cook, M.B., and McGlynn, K.A. (2010). International trends in the incidence of testicular cancer, 1973-2002. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* 19, 1151–1159.
- Childs, A.J., and Anderson, R.A. (2009). Activin A selectively represses expression of the membrane-bound isoform of Kit ligand in human fetal ovary. *Fertil. Steril.* 92, 1416–1419.
- Childs, A.J., Cowan, G., Kinnell, H.L., Anderson, R.A., and Saunders, P.T.K. (2011). Retinoic Acid Signalling and the Control of Meiotic Entry in the Human Fetal Gonad. *PLoS ONE* 6, e20249.
- Childs, A.J., Kinnell, H.L., He, J., and Anderson, R.A. (2012). LIN28 is selectively expressed by primordial and pre-meiotic germ cells in the human fetal ovary. *Stem Cells Dev.* 21, 2343–2349.
- Chilvers, C., Pike, M.C., Forman, D., Fogelman, K., and Wadsworth, M.E. (1984). Apparent doubling of frequency of undescended testis in England and Wales in 1962-81. *Lancet Lond. Engl.* 2, 330–332.
- Chocu, S., Evrard, B., Lavigne, R., Rolland, A.D., Aubry, F., Jégou, B., Chalmel, F., and Pineau, C. (2014). Forty-Four Novel Protein-Coding Loci Discovered Using a Proteomics Informed by Transcriptomics (PIT) Approach in Rat Male Germ Cells1. *Biol. Reprod.* 91, 123–123.
- Clark, A.M., Garland, K.K., and Russell, L.D. (2000). Desert hedgehog (Dhh) gene is required in the mouse testis for formation of adult-type Leydig cells and normal development of peritubular cells and seminiferous tubules. *Biol. Reprod.*
- Clements, J.A., Reyes, F.I., Winter, J.S., and Faiman, C. (1976). Studies on human sexual development. III. Fetal pituitary and serum, and amniotic fluid concentrations of LH, CG, and FSH. *J. Clin. Endocrinol. Metab.* 42, 9–19.

- Coffey, J.C., Aronin, P.A., French, F.S., and Nayfeh, S.N. (1972). Steroid metabolism by testicular homogenates of the Stanley-Gumbreck pseudohermaphrodite male rat. I. Increased formation of androsterone and androstenediol. *Steroids* *19*, 433–454.
- Cohen, P.E., and Holloway, J.K. (2010). Predicting Gene Networks in Human Oocyte Meiosis. *Biol. Reprod.* *82*, 469–472.
- Colborn, T., Dumanoski, D., and Myers, J.P. (1996). Our stolen future : are we threatening our fertility, intelligence, and survival? : a scientific detective story.
- Cole, B., Hensinger, K., Maciel, G.A.R., Chang, R.J., and Erickson, G.F. (2006). Human fetal ovary development involves the spatiotemporal expression of p450c17 protein. *J. Clin. Endocrinol. Metab.* *91*, 3654–3661.
- Combes, A.N., Spiller, C.M., Harley, V.R., Sinclair, A.H., Dunwoodie, S.L., Wilhelm, D., and Koopman, P. (2010). Gonadal defects in Cited2-mutant mice indicate a role for SF1 in both testis and ovary differentiation. *Int. J. Dev. Biol.* *54*, 683–689.
- Combes, A.N., Bowles, J., Feng, C.-W., Chiu, H.S., Khoo, P.-L., Jackson, A., Little, M.H., Tam, P.P.L., and Koopman, P. (2011). Expression and Functional Analysis of Dkk1 during Early Gonadal Development. *Sex. Dev.* *5*, 124–130.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* *17*, 13.
- Contreras, R., Rogiers, R., Van de Voorde, A., and Fiers, W. (1977). Overlapping of the VP2-VP3 gene and the VP1 gene in the SV40 genome. *Cell* *12*, 529–538.
- Cook, M.B., Akre, O., Forman, D., Madigan, M.P., Richiardi, L., and McGlynn, K.A. (2010). A systematic review and meta-analysis of perinatal variables in relation to the risk of testicular cancer--experiences of the son. *Int. J. Epidemiol.* *39*, 1605–1618.
- Cool, J., Carmona, F.D., Szucsik, J.C., and Capel, B. (2008). Peritubular myoid cells are not the migrating population required for testis cord formation in the XY gonad. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determ. Differ.* *2*, 128–133.
- Couse, J.F., Hewitt, S.C., Bunch, D.O., Sar, M., Walker, V.R., Davis, B.J., and Korach, K.S. (1999). Postnatal sex reversal of the ovaries in mice lacking estrogen receptors alpha and beta. *Science* *286*, 2328–2331.
- Couse, J.F., Yates, M.M., Deroo, B.J., and Korach, K.S. (2005). Estrogen receptor-beta is critical to granulosa cell differentiation and the ovulatory response to gonadotropins. *Endocrinology* *146*, 3247–3262.
- Coveney, D., Cool, J., Oliver, T., and Capel, B. (2008). Four-dimensional analysis of vascularization during primary development of an organ, the gonad. *Proc. Natl. Acad. Sci.* *105*, 7212–7217.
- Crick, F. (1958). On Protein Synthesis.
- Crick, F. (1970). Central dogma of molecular biology. *Nature* *227*, 561–563.
- Crisponi, L., Deiana, M., Loi, A., Chiappe, F., Uda, M., Amati, P., Bisceglia, L., Zelante, L., Nagaraja, R., Porcu, S., et al. (2001). The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat. Genet.* *27*, 159–166.
- Cui, S., Ross, A., Stallings, N., Parker, K.L., Capel, B., and Quaggin, S.E. (2004). Disrupted gonadogenesis and male-to-female sex reversal in Pod1 knockout mice. *Dev. Camb. Engl.* *131*, 4095–4105.
- Czerwinski, M., Natarajan, A., Barske, L., Looger, L.L., and Capel, B. (2016). A timecourse analysis of systemic and gonadal effects of temperature on sexual development of the red-eared slider turtle *Trachemys scripta elegans*. *Dev. Biol.* *420*, 166–177.
- Dai, M., Thompson, R.C., Maher, C., Contreras-Galindo, R., Kaplan, M.H., Markovitz, D.M., Omenn, G., and Meng, F. (2010). NGSQC: cross-platform quality analysis pipeline for deep sequencing data. *BMC Genomics* *11 Suppl 4*, S7.

- Damgaard, I.N., Skakkebaek, N.E., Toppari, J., Virtanen, H.E., Shen, H., Schramm, K.-W., Petersen, J.H., Jensen, T.K., Main, K.M., and Nordic Cryptorchidism Study Group (2006). Persistent pesticides in human breast milk and cryptorchidism. *Environ. Health Perspect.* *114*, 1133–1138.
- Darde, T.A., Sallou, O., Becker, E., Evrard, B., Monjeaud, C., Le Bras, Y., Jégou, B., Collin, O., Rolland, A.D., and Chalmel, F. (2015). The ReproGenomics Viewer: an integrative cross-species toolbox for the reproductive science community. *Nucleic Acids Res.* *43*, W109-16.
- Das, D.K., Sanghavi, D., Gawde, H., Idicula-Thomas, S., and Vasudevan, L. Novel homozygous mutations in Desert hedgehog gene in patients with 46,XY complete gonadal dysgenesis and prediction of its structural and functional implications by computational methods. *Eur. J. Med. Genet.* *54*, e529-34.
- David, M.P., Amit, A., Bergman, A., Yedwab, G., Paz, G.F., and Homonnai, Z.T. (1979). Sperm penetration in vitro: correlations between parameters of sperm quality and the penetration capacity. *Fertil. Steril.* *32*, 676–680.
- Davis, M.P., Carrieri, C., Saini, H.K., van Dongen, S., Leonardi, T., Bussotti, G., Monahan, J.M., Auchynnikava, T., Bitetti, A., Rappsilber, J., et al. (2017). Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. *EMBO Rep.* *18*, 1231–1247.
- De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* *4*, 49–60.
- De Felici, M., and Farini, D. (2012). The control of cell cycle in mouse primordial germ cells: old and new players. *Curr. Pharm. Des.* *18*, 233–244.
- De Felici, M., Klinger, F.G., Farini, D., Scaldaferrri, M.L., Iona, S., and Lobascio, M. (2005). Establishment of oocyte population in the fetal ovary: primordial germ cell proliferation and oocyte programmed cell death. *Reprod. Biomed. Online* *10*, 182–191.
- De Pol, A., Vaccina, F., Forabosco, A., Cavazzuti, E., and Marzona, L. (1997). Apoptosis of germ cells during human prenatal oogenesis. *Hum. Reprod.* *12*, 2235–2241.
- DeFalco, T., Takahashi, S., and Capel, B. (2011). Two distinct origins for Leydig cell progenitors in the fetal testis. *Dev. Biol.* *352*, 14–26.
- Deng, Y., Ren, X., Yang, L., Lin, Y., and Wu, X. (2003). A JNK-dependent pathway is required for TNF α -induced apoptosis. *Cell* *115*, 61–70.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* *22*, 1775–1789.
- Dewailly, D., Andersen, C.Y., Balen, A., Broekmans, F., Dilaver, N., Fanchin, R., Griesinger, G., Kelsey, T.W., La Marca, A., Lambalk, C., et al. (2014). The physiology and clinical utility of anti-Müllerian hormone in women. *Hum. Reprod. Update* *20*, 370–385.
- Dinger, M.E., Pang, K.C., Mercer, T.R., and Mattick, J.S. (2008). Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities. *PLOS Comput. Biol.* *4*, e1000176.
- Dissen, G.A., Garcia-Rudaz, C., and Ojeda, S.R. (2009). Role of neurotrophic factors in early ovarian development. *Semin. Reprod. Med.* *27*, 24–31.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner.
- Domenice, S., Machado, A.Z., Ferreira, F.M., Ferraz-de-Souza, B., Lerario, A.M., Lin, L., Nishi, M.Y., Gomes, N.L., da Silva, T.E., Silva, R.B., et al. (2016). Wide spectrum of NR5A1-related phenotypes in 46,XY and 46,XX individuals. *Birth Defects Res. Part C Embryo Today Rev.* *108*, 309–320.
- Dresser, D.W., Hacker, A., Lovell-Badge, R., and Guerrier, D. (1995). The genes for a spliceosome protein (SAP62) and the anti-Müllerian hormone (AMH) are contiguous. *Hum. Mol. Genet.* *4*, 1613–1618.

- van den Driesche, S., Walker, M., McKinnell, C., Scott, H.M., Eddie, S.L., Mitchell, R.T., Seckl, J.R., Drake, A.J., Smith, L.B., Anderson, R.A., et al. (2012). Proposed role for COUP-TFII in regulating fetal leydig cell steroidogenesis, perturbation of which leads to masculinization disorders in rodents. *PLoS ONE*.
- van den Driesche, S., Macdonald, J., Anderson, R.A., Johnston, Z.C., Chetty, T., Smith, L.B., McKinnell, C., Dean, A., Homer, N.Z., Jorgensen, A., et al. (2015). Prolonged exposure to acetaminophen reduces testosterone production by the human fetal testis in a xenograft model. *Sci. Transl. Med.* 7, 288ra80.
- Dubé, C., Bergeron, F., Vaillant, M.J., Robert, N.M., Brousseau, C., and Tremblay, J.J. (2009). The nuclear receptors SF1 and LRH1 are expressed in endometrial cancer cells and regulate steroidogenic gene transcription by cooperating with AP-1 factors. *Cancer Lett.* 275, 127–138.
- Dufau, M.L., Catt, K.J., and Tsuruhara, T. (1972). A sensitive gonadotropin responsive system: radioimmunoassay of testosterone production by the rat testis in vitro. *Endocrinology* 90, 1032–1040.
- Dweep, H., Sticht, C., Pandey, P., and Gretz, N. (2011). miRWalk--database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J. Biomed. Inform.* 44, 839–847.
- Edson, M.A., Nagaraja, A.K., and Matzuk, M.M. (2009). The Mammalian Ovary from Genesis to Revelation. *Endocr. Rev.* 30, 624–712.
- Eggers, S., and Sinclair, A. (2012). Mammalian sex determination—insights from humans and mice. *Chromosome Res.* 20, 215–238.
- Eggers, S., Sadedin, S., van den Bergen, J.A., Robevska, G., Ohnesorg, T., Hewitt, J., Lambeth, L., Bouty, A., Knarston, I.M., Tan, T.Y., et al. (2016). Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort. *Genome Biol.* 17, 243.
- Eid, W., Opitz, L., and Biason-Lauber, A. (2015). Genome-Wide Identification of CBX2 Targets: Insights in the Human Sex Development Network. *Mol. Endocrinol.* 29, 247–257.
- Escalante-Alcalde, D., and Merchant-Larios, H. (1992). Somatic and germ cell interactions during histogenetic aggregation of mouse fetal testes. *Exp. Cell Res.* 198, 150–158.
- Eskenazi, B., Chevrier, J., Rosas, L.G., Anderson, H.A., Bornman, M.S., Bouwman, H., Chen, A., Cohn, B.A., de Jager, C., Henshel, D.S., et al. (2009). The Pine River statement: human health consequences of DDT use. *Environ. Health Perspect.* 117, 1359–1367.
- Evans, J.P., and Robinson, D.N. (2011). The spatial and mechanical challenges of female meiosis. *Mol. Reprod. Dev.* 78, 769–777.
- Evans, V.C., Barker, G., Heesom, K.J., Fan, J., Bessant, C., and Matthews, D.A. (2012). De novo derivation of proteomes from transcriptomes for transcript and protein identification. *Nat. Methods* 9, 1207–1211.
- Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X., et al. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 46, D308–D314.
- Fang, Z., Martin, J., and Wang, Z. (2012). Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. *Cell Biosci.* 2, 26.
- Feng, C.-W., Bowles, J., and Koopman, P. (2014). Control of mammalian germ cell entry into meiosis. *Mol. Cell. Endocrinol.* 382, 488–497.
- Findlay, J.K., Drummond, A.E., Dyson, M.L., Baillie, A.J., Robertson, D.M., and Ethier, J.-F. (2002). Recruitment and development of the follicle; the roles of the transforming growth factor-beta superfamily. *Mol. Cell. Endocrinol.* 191, 35–43.
- Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–37.
- Fisch, H., Goluboff, E.T., Olson, J.H., Feldshuh, J., Broder, S.J., and Barad, D.H. (1996). Semen analyses in 1,283 men from the United States over a 25-year period: no decline in quality. *Fertil. Steril.* 65, 1009–1014.

- Fisch, H., Hyun, G., and Hensle, T.W. (2010). Rising hypospadias rates: disproving a myth. *J. Pediatr. Urol.* 6, 37–39.
- Forabosco, A., and Sforza, C. (2007). Establishment of ovarian reserve: a quantitative morphometric study of the developing human ovary. *Fertil. Steril.* 88, 675–683.
- Foster, J.W., and Graves, J.A. (1994). An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. *Proc. Natl. Acad. Sci.* 91, 1927–1931.
- Foster, J.W., Dominguez-Steglich, M.A., Guioli, S., Kwok, C., Weller, P.A., Stevanović, M., Weissenbach, J., Mansour, S., Young, I.D., Goodfellow, P.N., et al. (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* 372, 525–530.
- Fowler, P.A., Flannigan, S., Mathers, A., Gillanders, K., Lea, R.G., Wood, M.J., Maheshwari, A., Bhattacharya, S., Collie-Duguid, E.S.R., Baker, P.J., et al. (2009). Gene expression analysis of human fetal ovarian primordial follicle formation. *J. Clin. Endocrinol. Metab.* 94, 1427–1435.
- Fowler, P.A., Anderson, R.A., Saunders, P.T., Kinnell, H., Mason, J.I., Evans, D.B., Bhattacharya, S., Flannigan, S., Franks, S., Monteiro, A., et al. (2011). Development of steroid signaling pathways during primordial follicle formation in the human fetal ovary. *J. Clin. Endocrinol. Metab.* 96, 1754–1762.
- Francavilla, S., Cordeschi, G., Properzi, G., Concordia, N., Cappa, F., and Pozzi, V. (1990). Ultrastructure of fetal human gonad before sexual differentiation and during early testicular and ovarian development. *J. Submicrosc. Cytol. Pathol.* 22, 389–400.
- Frankish, A., Uszczynska, B., Ritchie, G.R.S., Gonzalez, J.M., Pervouchine, D., Petryszak, R., Mudge, J.M., Fonseca, N., Brazma, A., Guigo, R., et al. (2015). Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. *BMC Genomics* 16 Suppl 8, S2.
- Franklin, R.E., and Gosling, R.G. (1953a). Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* 172, 156–157.
- Franklin, R.E., and Gosling, R.G. (1953b). Molecular configuration in sodium thymonucleate. *Nature* 171, 740–741.
- Freeman, B. (2003). The active migration of germ cells in the embryos of mice and men is a myth. *Reprod. Camb. Engl.* 125, 635–643.
- Frith, M.C., Forrest, A.R., Nourbakhsh, E., Pang, K.C., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T.L., and Grimmond, S.M. (2006). The Abundance of Short Proteins in the Mammalian Proteome. *PLOS Genet.* 2, e52.
- Fu, Q., and Wang, P.J. (2014). Mammalian piRNAs: Biogenesis, function, and mysteries. *Spermatogenesis* 4, e27889.
- Fukami, M., Wada, Y., Miyabayashi, K., Nishino, I., Hasegawa, T., Nordenskjöld, A., Camerino, G., Kretz, C., Buj-Bello, A., Laporte, J., et al. (2006). CXorf6 is a causative gene for hypospadias. *Nat. Genet.* 38, 1369–1371.
- Fukuda, T., Hedinger, C., and Groscurth, P. (1975). Ultrastructure of developing germ cells in the fetal human testis. *Cell Tissue Res.* 161, 55–70.
- Gaffney, D.J. (2013). Global properties and functional complexity of human gene regulatory variation. *PLoS Genet.* 9, e1003501.
- Gan, H., Wen, L., Liao, S., Lin, X., Ma, T., Liu, J., Song, C.-X., Wang, M., He, C., Han, C., et al. (2013). Dynamics of 5-hydroxymethylcytosine during mouse spermatogenesis. *Nat. Commun.* 4.
- Ganter, B., Snyder, R.D., Halbert, D.N., and Lee, M.D. (2006). Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix database. *Pharmacogenomics* 7, 1025–1044.
- Garalde, D.R., Snell, E.A., Jachimowicz, D., Sipos, B., Lloyd, J.H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., et al. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods.*

- García-López, J., Alonso, L., Cárdenas, D.B., Artaza-Alvarez, H., Hourcade, J. de D., Martínez, S., Briño-Enríquez, M.A., and Del Mazo, J. (2015). Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization. *RNA N. Y. N* 21, 946–962.
- García-Ortiz, J., Pelosi, E., Omari, S., Nedorezov, T., Piao, Y., Karmazin, J., Uda, M., Cao, A., Cole, S.W., Forabosco, A., et al. (2009). Foxl2 functions in sex determination and histogenesis throughout mouse ovary development. *BMC Dev. Biol.* 9, 36.
- García-Rudaz, C., Luna, F., Tapia, V., Kerr, B., Colgin, L., Galimi, F., Dissen, G.A., Rawlings, N.D., and Ojeda, S.R. (2007). Fxna, a novel gene differentially expressed in the rat ovary at the time of folliculogenesis, is required for normal ovarian histogenesis. *Dev. Camb. Engl.* 134, 945–957.
- Gaskell, T.L., Esnal, A., Robinson, L.L.L., Anderson, R. a, and Saunders, P.T.K. (2004). Immunohistochemical profiling of germ cells within the human fetal testis: identification of three subpopulations. *Biol. Reprod.* 71, 2012–2021.
- Gaspari, L., Paris, F., Jandel, C., Kalfa, N., Orsini, M., Daurès, J.P., and Sultan, C. (2011). Prenatal environmental risk factors for genital malformations in a population of 1442 French male newborns: a nested case–control study. *Hum. Reprod.* 26, 3155–3162.
- Gaudriault, P., Mazaud-Guittot, S., Lavoué, V., Coiffec, I., Lesné, L., Dejucq-Rainsford, N., Scholze, M., Kortenkamp, A., and Jégou, B. (2017). Endocrine Disruption in Human Fetal Testis Explants by Individual and Combined Exposures to Selected Pharmaceuticals, Pesticides, and Environmental Pollutants. *Environ. Health Perspect.* 125, 087004.
- Gene Ontology Consortium, T.G.O. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43, D1049-56.
- George, F.W., and Wilson, J.D. (1978). Conversion of androgen to estrogen by the human fetal ovary. *J. Clin. Endocrinol. Metab.* 47, 550–555.
- Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korb, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S., and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 17, 669–681.
- Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 10, e1004383.
- Gierl, M.S., Gruhn, W.H., von Seggern, A., Maltry, N., and Niehrs, C. (2012a). GADD45G functions in male sex determination by promoting p38 signaling and Sry expression. *Dev. Cell* 23, 1032–1042.
- Gierl, M.S., Gruhn, W.H., von Seggern, A., Maltry, N., and Niehrs, C. (2012b). GADD45G functions in male sex determination by promoting p38 signaling and Sry expression. *Dev. Cell* 23, 1032–1042.
- Gilad, Y., Rifkin, S.A., and Pritchard, J.K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* TIG 24, 408–415.
- Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202.
- Giulivo, M., Lopez de Alda, M., Capri, E., and Barceló, D. (2016). Human exposure to endocrine disrupting compounds: Their role in reproductive systems, metabolic syndrome and breast cancer. A review. *Environ. Res.* 151, 251–264.
- Gkoutela, S., Li, Z., Vincent, J.J., Zhang, K.X., Chen, A., Pellegrini, M., and Clark, A.T. (2013). The ontogeny of cKIT+ human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and in vitro differentiation. *Nat. Cell Biol.* 15, 113–122.
- Gkoutela, S., Zhang, K.X., Shafiq, T.A., Liao, W.W., Hargan-Calvopi??a, J., Chen, P.Y., and Clark, A.T. (2015). DNA demethylation dynamics in the human prenatal germline. *Cell* 161, 1425–1436.

- Gonen, N., Quinn, A., O'Neill, H.C., Koopman, P., Lovell-Badge, R., and Jaenisch, R. (2017). Normal Levels of Sox9 Expression in the Developing Mouse Testis Depend on the TES/TESCO Enhancer, but This Does Not Act Alone. *PLOS Genet.* *13*, e1006520.
- Gonen, N., Futtner, C.R., Wood, S., Garcia-Moreno, S.A., Salamone, I.M., Samson, S.C., Sekido, R., Poulat, F., Maatouk, D.M., and Lovell-Badge, R. (2018). Sex reversal following deletion of a single distal enhancer of Sox9. *Science* eaas9408.
- Gould, D.A., Moscoso, G.J., Young, M.P., and Barton, D.P. (2000). Human first trimester fetal ovaries express oncofetal antigens and steroid receptors. *J. Soc. Gynecol. Investig.* *7*, 131–138.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* *29*, 644–652.
- Graves, J.A.M. (2002). The rise and fall of SRY. *Trends Genet.* *18*, 259–264.
- Graves, J.A.M. (2006). Sex Chromosome Specialization and Degeneration in Mammals. *Cell* *124*, 901–914.
- Gregoire, E.P., Stevant, I., Chassot, A.-A., Martin, L., Lachambre, S., Mondin, M., de Rooij, D.G., Nef, S., and Chaboissier, M.-C. (2018). NRG1 signalling regulates the establishment of Sertoli cell stock in the mouse testis. *Mol. Cell. Endocrinol.*
- Grewal, S.I.S., and Elgin, S.C.R. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature* *447*, 399–406.
- Grinspon, R.P., Nevado, J., Mori Alvarez, M. de los Á., del Rey, G., Castera, R., Venara, M., Chiesa, A., Podestá, M., Lapunzina, P., and Rey, R.A. (2016). 46,XX ovotesticular DSD associated with a *SOX3* gene duplication in a *SRY*-negative boy. *Clin. Endocrinol. (Oxf.)* *85*, 673–675.
- Griswold, S.L., and Behringer, R.R. (2009). Fetal Leydig Cell Origin and Development. *Sex. Dev.* *3*, 1–15.
- Groothuis, P.G., Dassen, H.H.N.M., Romano, A., and Punyadeera, C. (2007). Estrogen and the endometrium: lessons learned from gene expression profiling in rodents and human. *Hum. Reprod. Update* *13*, 405–417.
- Grzybowska, E.A. (2012). Human intronless genes: Functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. *Biochem. Biophys. Res. Commun.* *424*, 1–6.
- Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* *346*, 245–250.
- Guillette, L.J., Gross, T.S., Masson, G.R., Matter, J.M., Percival, H.F., and Woodward, A.R. (1994). Developmental abnormalities of the gonad and abnormal sex hormone concentrations in juvenile alligators from contaminated and control lakes in Florida. *Environ. Health Perspect.* *102*, 680–688.
- Gunes, S.O., Metin Mahmutoglu, A., and Agarwal, A. (2016). Genetic and epigenetic effects in sex determination. *Birth Defects Res. Part C Embryo Today Rev.* *108*, 321–336.
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015a). The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* *161*, 1437–1452.
- Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015b). The transcriptome and DNA methylome landscapes of human primordial germ cells. *Cell* *161*, 1437–1452.
- Guo, H., Hu, B., Yan, L., Yong, J., Wu, Y., Gao, Y., Guo, F., Hou, Y., Fan, X., Dong, J., et al. (2017). DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. *Cell Res.* *27*, 165–183.
- Gustin, S.E., Hogg, K., Stringer, J.M., Rastetter, R.H., Pelosi, E., Miles, D.C., Sinclair, A.H., Wilhelm, D., and Western, P.S. (2016). WNT/ β -catenin and p27/FOXL2 differentially regulate supporting cell proliferation in the developing ovary. *Dev. Biol.*
- Hacker, A., Capel, B., Goodfellow, P., and Lovell-Badge, R. (1995). Expression of Sry, the mouse sex determining gene. *Dev. Camb. Engl.* *121*, 1603–1614.

- Haines, B., Hughes, J., Corbett, M., Shaw, M., Innes, J., Patel, L., Gecz, J., Clayton-Smith, J., and Thomas, P. (2015). Interchromosomal Insertional Translocation at Xq26.3 Alters *SOX3* Expression in an Individual With XX Male Sex Reversal. *J. Clin. Endocrinol. Metab.* *100*, E815–E820.
- Hammes, A., Guo, J.K., Lutsch, G., Leheste, J.R., Landrock, D., Ziegler, U., Gubler, M.C., and Schedl, A. (2001). Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. *Cell* *106*, 319–329.
- Han, H., Shim, H., Shin, D., Shim, J.E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., et al. (2015a). TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* *5*, 11432.
- Han, Y., Guo, S., Muegge, K., Zhang, W., and Zhou, B. (2015b). Advanced Applications of RNA Sequencing and Challenges. *Bioinforma. Biol. Insights* *9*, 29.
- Hanada, K., Zhang, X., Borevitz, J.O., Li, W.-H., and Shiu, S.-H. (2007). A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. *Genome Res.* *17*, 632–640.
- Handel, M.A., and Eppig, J.J. (1979). Sertoli cell differentiation in the testes of mice genetically deficient in germ cells. *Biol. Reprod.* *20*, 1031–1038.
- Hanley, N., Hagan, D., Clement-Jones, M., Ball, S., Strachan, T., Salas-Cortés, L., McElreavey, K., Lindsay, S., Robson, S., Bullen, P., et al. (2000). SRY, SOX9, and DAX1 expression patterns during human sex determination and gonadal development. *Mech. Dev.* *91*, 403–407.
- Hansen, K.R., Knowlton, N.S., Thyer, A.C., Charleston, J.S., Soules, M.R., and Klein, N.A. (2008). A new model of reproductive aging: the decline in ovarian non-growing follicle number from birth to menopause. *Hum. Reprod. Oxf. Engl.* *23*, 699–708.
- Hara, K., Kanai-Azuma, M., Uemura, M., Shitara, H., Taya, C., Yonekawa, H., Kawakami, H., Tsunekawa, N., Kurohmaru, M., and Kanai, Y. (2009). Evidence for crucial role of hindgut expansion in directing proper migration of primordial germ cells in mouse early embryogenesis. *Dev. Biol.* *330*, 427–439.
- Harding, S.D., Armit, C., Armstrong, J., Brennan, J., Cheng, Y., Haggarty, B., Houghton, D., Lloyd-MacGilp, S., Pi, X., Roochun, Y., et al. (2011). The GUDMAP database – an online resource for genitourinary research. *Dev. Camb. Engl.* *138*, 2845–2853.
- Hardwick, S.A., Deveson, I.W., and Mercer, T.R. (2017). Reference standards for next-generation sequencing. *Nat. Rev. Genet.* *18*, 473–484.
- Harley, V.R., and Goodfellow, P.N. (1994). The biochemical role of SRY in sex determination. *Mol. Reprod. Dev.* *39*, 184–193.
- Harris, A., Siggers, P., Corrochano, S., Warr, N., Sagar, D., Grimes, D.T., Suzuki, M., Burdine, R.D., Cong, F., Koo, B.-K., et al. (2018). ZNRF3 functions in mammalian sex determination by inhibiting canonical WNT signaling. *Proc. Natl. Acad. Sci. U. S. A.* *115*, 5474–5479.
- Harris, S.E., Chand, A.L., Winship, I.M., Gersak, K., Aittomäki, K., and Shelling, A.N. (2002). Identification of novel mutations in FOXL2 associated with premature ovarian failure. *Mol. Hum. Reprod.* *8*, 729–733.
- Hashimoto, Y., Kondo, T., and Kageyama, Y. (2008). Lilliputians get into the limelight: novel class of small peptide genes in morphogenesis. *Dev. Growth Differ.* *50 Suppl 1*, S269-276.
- Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., and Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. *BioTechniques* *56*, 61–64, 66, 68, passim.
- Hedlund, E., and Deng, Q. (2018). Single-cell RNA sequencing: Technical advancements and biological applications. *Mol. Aspects Med.* *59*, 36–46.
- Heller, M.J. (2002). DNA Microarray Technology: Devices, Systems, and Applications. *Annu. Rev. Biomed. Eng.* *4*, 129–153.

- Hersmus, R., van der Zwan, Y.G., Stoop, H., Bernard, P., Sreenivasan, R., Oosterhuis, J.W., Brüggewirth, H.T., de Boer, S., White, S., Wolffenbuttel, K.P., et al. (2012). A 46,XY Female DSD Patient with Bilateral Gonadoblastoma, a Novel SRY Missense Mutation Combined with a WT1 KTS Splice-Site Mutation. *PLoS ONE* 7, e40858.
- Hezroni, H., Koppstein, D., Schwartz, M.G., Avrutin, A., Bartel, D.P., and Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* 11, 1110–1122.
- Hiramatsu, R., Harikae, K., Tsunekawa, N., Kurohmaru, M., Matsuo, I., and Kanai, Y. (2010). FGF signaling directs a center-to-pole expansion of tubulogenesis in mouse testis differentiation. *Dev. Camb. Engl.* 137, 303–312.
- Hossain, A., and Saunders, G.F. (2001). The human sex-determining gene SRY is a direct target of WT1. *J. Biol. Chem.* 276, 16817–16823.
- Houmard, B., Small, C., Yang, L., Nalwai-Cecchini, T., Cheng, E., Hassold, T., and Griswold, M. (2009). Global Gene Expression in the Human Fetal Testis and Ovary. *Biol Reprod* 443, [biolreprod.108.075747](https://doi.org/10.1093/biolreprod/108.075747).
- Høyer, P.E., Byskov, A.G., and Møllgård, K. (2005). Stem cell factor and c-Kit in human primordial germ cells and fetal ovaries. *Mol. Cell. Endocrinol.* 234, 1–10.
- Hrdlickova, R., Toloue, M., and Tian, B. (2017). RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. RNA* 8, e1364.
- Hsueh, A.J., and Rauch, R. (2012). Ovarian Kaleidoscope Database: Ten Years and Beyond. *Biol. Reprod.* 86.
- Hu, L., Xu, Z., Hu, B., and Lu, Z.J. (2017). COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Res.* 45, e2.
- Hu, W., Alvarez-Dominguez, J.R., and Lodish, H.F. (2012). Regulation of mammalian cell differentiation by long non-coding RNAs. *EMBO Rep.* 13, 971–983.
- Hu, Y.-C., Okumura, L.M., and Page, D.C. (2013). Gata4 is required for formation of the genital ridge in mice. *PLoS Genet.* 9, e1003629.
- Hua, J., Xu, B., Yang, Y., Ban, R., Iqbal, F., Cooke, H.J., Zhang, Y., and Shi, Q. (2015). Follicle Online: an integrated database of follicle assembly, development and ovulation. *Database J. Biol. Databases Curation* 2015, bav036.
- Huang, B., Wang, S., Ning, Y., Lamb, A.N., and Bartley, J. (1999). Autosomal XX sex reversal caused by duplication of SOX9. *Am. J. Med. Genet.* 87, 349–353.
- Huang, H.-C., Niu, Y., and Qin, L.-X. (2015). Differential Expression Analysis for RNA-Seq: An Overview of Statistical Methods and Computational Software. *Cancer Inform.* 14, 57–67.
- Huang, S., Zhang, J., Li, R., Zhang, W., He, Z., Lam, T.-W., Peng, Z., and Yiu, S.-M. (2011). SOAPsplice: Genome-Wide ab initio Detection of Splice Junctions from RNA-Seq Data. *Front. Genet.* 2, 46.
- Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* 19, 342–347.
- Huhtaniemi, I.T., Korenbrot, C.C., and Jaffe, R.B. (1977). HCG binding and stimulation of testosterone biosynthesis in the human fetal testis. *J. Clin. Endocrinol. Metab.* 44, 963–967.
- Huhtaniemi, I.T., Katikineni, M., and Catt, K.J. (1981). Regulation of luteinizing-hormone-receptors and steroidogenesis in the neonatal rat testis. *Endocrinology* 109, 588–595.
- Huhtaniemi, I.T., Nozu, K., Warren, D.W., Dufau, M.L., and Catt, K.J. (1982). Acquisition of regulatory mechanisms for gonadotropin receptors and steroidogenesis in the maturing rat testis. *Endocrinology* 111, 1711–1720.

- Hume, R., Kelly, R.W., Taylor, P.L., and Boyd, G.S. (1984). The catalytic cycle of cytochrome P-450_{scc} and intermediates in the conversion of cholesterol to pregnenolone. *Eur. J. Biochem. FEBS* *140*, 583–591.
- Hummitzsch, K., Irving-Rodgers, H.F., Hatzirodos, N., Bonner, W., Sabatier, L., Reinhardt, D.P., Sado, Y., Ninomiya, Y., Wilhelm, D., and Rodgers, R.J. (2013). A New Model of Development of the Mammalian Ovary and Follicles. *PLoS ONE* *8*, e55578.
- Huyghe, E., Plante, P., and Thonneau, P.F. (2007). Testicular cancer variations in time and space in Europe. *Eur. Urol.* *51*, 621–628.
- Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T., and Yamada, H. (2015). Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* *43*, D921-927.
- Ikeda, Y. (1994). Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases. *Mol. Endocrinol.* *8*, 654–662.
- Imbeaud, S., Carre-Eusebe, D., Rey, R., Belville, C., Josso, N., and Picard, J.-Y. (1994). Molecular genetics of the persistent Mullerian duct syndrome: a study of 19 families. *Hum. Mol. Genet.* *3*, 125–131.
- Imbeaud, S., Faure, E., Lamarre, I., Mattéi, M.-G., di Clemente, N., Tizard, R., Carré-Eusèbe, D., Belville, C., Tragethon, L., Tonkin, C., et al. (1995). Insensitivity to anti-Müllerian hormone due to a mutation in the human anti-Müllerian hormone receptor. *Nat. Genet.* *11*, 382–388.
- Inagaki, A., Roset, R., and Petrini, J.H.J. (2016). Functions of the MRE11 complex in the development and maintenance of oocytes. *Chromosoma* *125*, 151–162.
- Inoue, M., Shima, Y., Miyabayashi, K., Tokunaga, K., Sato, T., Baba, T., Ohkawa, Y., Akiyama, H., Suyama, M., and Morohashi, K. (2016). Isolation and Characterization of Fetal Leydig Progenitor Cells of Male Mice. *Endocrinology* *157*, 1222–1233.
- Irie, N., Tang, W.W.C., and Azim Surani, M. (2014). Germ cell specification and pluripotency in mammals: a perspective from early embryogenesis. *Reprod. Med. Biol.* *13*, 203–215.
- Irie, N., Weinberger, L., Tang, W.W.C.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H.H., and Surani, M.A.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. *Cell* *160*, 253–268.
- Irie, N., Sybirna, A., and Surani, M.A. (2018). Chapter Two - What Can Stem Cell Models Tell Us About Human Germ Cell Biology? In *Current Topics in Developmental Biology*, A.H. Brivanlou, ed. (Academic Press), pp. 25–65.
- Irvine, S., Cawood, E., Richardson, D., MacDonald, E., and Aitken, J. (1996). Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 577 men in Scotland over 11 years. *BMJ* *312*, 467–471.
- Ito, M., Yokouchi, K., Yoshida, K., Kano, K., Naito, K., Miyazaki, J.-I., and Tojo, H. (2006). Investigation of the fate of Sry-expressing cells using an in vivo Cre/loxP system. *Dev. Growth Differ.* *48*, 41–47.
- Jacobs, P.A., and Strong, J.A. (1959). A Case of Human Intersexuality Having a Possible XXY Sex-Determining Mechanism. *Nature* *183*, 302–303.
- Jäger, R.J., Harley, V.R., Pfeiffer, R.A., Goodfellow, P.N., and Scherer, G. (1992). A familial mutation in the testis-determining gene SRY shared by both sexes. *Hum. Genet.* *90*, 350–355.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* *343*, 776–779.
- Jameson, S.A., Lin, Y.-T., and Capel, B. (2012a). Testis development requires the repression of Wnt4 by Fgf signaling. *Dev. Biol.* *370*, 24–32.
- Jameson, S.A., Natarajan, A., Cool, J., DeFalco, T., Maatouk, D.M., Mork, L., Munger, S.C., and Capel, B. (2012b). Temporal transcriptional profiling of somatic and germ cells reveals biased lineage priming of sexual fate in the fetal mouse gonad. *PLoS Genet.* *8*, e1002575.

- Jeanes, A., Wilhelm, D., Wilson, M.J., Bowles, J., McClive, P.J., Sinclair, A.H., and Koopman, P. (2005). Evaluation of candidate markers for the peritubular myoid cell lineage in the developing mouse testis. *Reproduction* *130*, 509–516.
- Jeays-Ward, K. (2003). Endothelial and steroidogenic cell migration are regulated by WNT4 in the developing mammalian gonad. *Development* *130*, 3663–3670.
- Jégou, B. (2015). Paracetamol-induced endocrine disruption in human fetal testes. *Nat. Rev. Endocrinol.* *11*, 453–454.
- Jegou, B., and Velez De La Calle, J.F. (1993). [Spermatogenesis protection: myth or reality?]. *Contracept. Fertil. Sex. (Paris)* *21*, 744–748.
- Jiao, Y., Widschwendter, M., and Teschendorff, A.E. (2014). A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. *Bioinforma. Oxf. Engl.* *30*, 2360–2366.
- Johannsen, W. (1909). *Elemente der exakten erblichkeitslehre. Deutsche wesentlich erweiterte ausgabe in fünfundzwanzig vorlesungen*, (Jena,: G. Fischer,).
- Johnson, R., and Guigó, R. (2014). The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* *20*, 959–976.
- Johnson, K.J., Heger, N.E., and Boekelheide, K. (2012). Of mice and men (and rats): Phthalate-induced fetal testis endocrine disruption is species-dependent. *Toxicol. Sci.*
- Jones, R.L., and Pepling, M.E. (2013). KIT signaling regulates primordial follicle formation in the neonatal mouse ovary. *Dev. Biol.* *382*, 186–197.
- Jordan, B.K., Mohammed, M., Ching, S.T., Délot, E., Chen, X.-N., Dewing, P., Swain, A., Rao, P.N., Elejalde, B.R., and Vilain, E. (2001). Up-Regulation of WNT-4 Signaling and Dosage-Sensitive Sex Reversal in Humans. *Am. J. Hum. Genet.* *68*, 1102–1109.
- Jørgensen, A., and Rajpert-De Meyts, E. (2014). Regulation of meiotic entry and gonadal sex differentiation in the human: normal and disrupted signaling. *Biomol. Concepts* *5*, 331–341.
- Jørgensen, N., Andersen, A.G., Eustache, F., Irvine, D.S., Suominen, J., Petersen, J.H., Andersen, A.N., Auger, J., Cawood, E.H., Horte, A., et al. (2001). Regional differences in semen quality in Europe. *Hum. Reprod. Oxf. Engl.* *16*, 1012–1019.
- Juengel, J.L., Heath, D.A., Quirke, L.D., and McNatty, K.P. (2006). Oestrogen receptor alpha and beta, androgen receptor and progesterone receptor mRNA and protein localisation within the developing ovary and in small growing follicles of sheep. *Reprod. Camb. Engl.* *131*, 81–92.
- Kageyama, Y., Kondo, T., and Hashimoto, Y. (2011). Coding vs non-coding: Translatability of short ORFs found in putative non-coding transcripts. *Biochimie* *93*, 1981–1986.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* *44*, D457–D462.
- Kaprara, A., and Huhtaniemi, I.T. (2017). The hypothalamus-pituitary-gonad axis: tales of mice and men. *Metabolism.*
- Karl, J., and Capel, B. (1998). Sertoli Cells of the Mouse Testis Originate from the Coelomic Epithelium. *Dev. Biol.* *203*, 323–333.
- Kashimada, K., and Koopman, P. (2010). Sry: the master switch in mammalian sex determination. *Development* *137*, 3921–3930.
- Kashimada, K., Pelosi, E., Chen, H., Schlessinger, D., Wilhelm, D., and Koopman, P. (2011). FOXL2 and BMP2 act cooperatively to regulate follistatin gene expression during ovarian development. *Endocrinology* *152*, 272–280.

- Kastenmayer, J.P., Ni, L., Chu, A., Kitchen, L.E., Au, W.-C., Yang, H., Carter, C.D., Wheeler, D., Davis, R.W., Boeke, J.D., et al. (2006). Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* *16*, 365–373.
- Katoh-Fukui, Y., Miyabayashi, K., Komatsu, T., Owaki, A., Baba, T., Shima, Y., Kidokoro, T., Kanai, Y., Schedl, A., Wilhelm, D., et al. (2012). *Cbx2*, a Polycomb Group Gene, Is Required for *Sry* Gene Expression in Mice. *Endocrinology* *153*, 913–924.
- Kerr, B., Garcia-Rudaz, C., Dorfman, M., Paredes, A., and Ojeda, S.R. (2009). NTRK1 and NTRK2 receptors facilitate follicle assembly and early follicular development in the mouse ovary. *Reprod. Camb. Engl.* *138*, 131–140.
- Kerr, C.L., Hill, C.M., Blumenthal, P.D., and Gearhart, J.D. (2008). Expression of pluripotent stem cell markers in the human fetal ovary. *Hum. Reprod.* *23*, 589–599.
- Khatri, P., Sirota, M., and Butte, A.J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput. Biol.* *8*, e1002375.
- Kilcoyne, K.R., Smith, L.B., Atanassova, N., Macpherson, S., McKinnell, C., van den Driesche, S., Jobling, M.S., Chambers, T.J.G., De Gendt, K., Verhoeven, G., et al. (2014). Fetal programming of adult Leydig cell function by androgenic effects on stem/progenitor cells. *Proc. Natl. Acad. Sci.* *111*, E1924–E1932.
- Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* *14*, R36.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* *12*, 357–360.
- Kim, J.H., Dhanasekaran, S.M., Prensner, J.R., Cao, X., Robinson, D., Kalyana-Sundaram, S., Huang, C., Shankar, S., Jing, X., Iyer, M., et al. (2011). Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer. *Genome Res.* *21*, 1028–1041.
- Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. *Nature* *509*, 575–581.
- Kim, Y., Kobayashi, A., Sekido, R., DiNapoli, L., Brennan, J., Chaboissier, M.-C., Poulat, F., Behringer, R.R., Lovell-Badge, R., and Capel, B. (2006). *Fgf9* and *Wnt4* Act as Antagonistic Signals to Regulate Mammalian Sex Determination. *PLoS Biol.* *4*, e187.
- Kitamura, K., Yanazawa, M., Sugiyama, N., Miura, H., Iizuka-Kogo, A., Kusaka, M., Omichi, K., Suzuki, R., Kato-Fukui, Y., Kamiirisa, K., et al. (2002). Mutation of ARX causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. *Nat. Genet.* *32*, 359–369.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* *161*, 1187–1201.
- Klip, H., Verloop, J., van Gool, J.D., Koster, M.E., Burger, C.W., van Leeuwen, F.E., and OMEGA Project Group (2002). Hypospadias in sons of women exposed to diethylstilbestrol in utero: a cohort study. *The Lancet* *359*, 1102–1107.
- Knebelmann, B., Boussin, L., Guerrier, D., Legeai, L., Kahn, A., Josso, N., and Picard, J.Y. (1991). Anti-Müllerian hormone Bruxelles: a nonsense mutation associated with the persistent Müllerian duct syndrome. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 3767–3771.
- Knight, P.G., and Glistler, C. (2003). Local roles of TGF-beta superfamily members in the control of ovarian follicle development. *Anim. Reprod. Sci.* *78*, 165–183.
- Knight, P.G., and Glistler, C. (2006). TGF-beta superfamily members and ovarian follicle development. *Reprod. Camb. Engl.* *132*, 191–206.

- Kobayashi, T., Zhang, H., Tang, W.W.C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S., Contreras, D.A., Webb, R., et al. (2017). Principles of early human development and germ cell program from conserved model systems. *Nature* 546, 416–420.
- Köhler, S., Doelken, S.C., Mungall, C.J., Bauer, S., Firth, H.V., Bailleul-Forestier, I., Black, G.C.M., Brown, D.L., Brudno, M., Campbell, J., et al. (2014). The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* 42, D966-74.
- Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y.A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T., et al. (2015). ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116.
- Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Mol. Cell* 58, 610–620.
- Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, W345-9.
- Koopman, P., Münsterberg, A., Capel, B., Vivian, N., and Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* 348, 450–452.
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. *Nature* 351, 117–121.
- Koubova, J., Menke, D.B., Zhou, Q., Capel, B., Griswold, M.D., and Page, D.C. (2006). Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc. Natl. Acad. Sci. U. S. A.* 103, 2474–2479.
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73.
- Krebs, J.E., Goldstein, E.S., and Kilpatrick, S.T. (2017). Lewin’s GENES XII (Jones & Bartlett Learning).
- Kreidberg, J.A., Sariola, H., Loring, J.M., Maeda, M., Pelletier, J., Housman, D., and Jaenisch, R. (1993). *WT-1* is required for early kidney development. *Cell* 74, 679–691.
- de Kretser, D.M., Stanton, P., and O’Donnell, L. (2018). Structure/Cells Overview. In *Encyclopedia of Reproduction (Second Edition)*, M.K. Skinner, ed. (Oxford: Academic Press), pp. 10–16.
- Kristensen, D.M., Hass, U., Lesné, L., Lottrup, G., Jacobsen, P.R., Desdoits-Lethimonier, C., Boberg, J., Petersen, J.H., Toppari, J., Jensen, T.K., et al. (2011). Intrauterine exposure to mild analgesics is a risk factor for development of male reproductive disorders in human and rat. *Hum. Reprod. Oxf. Engl.* 26, 235–244.
- Kristensen, D.M., Lesné, L., Le Fol, V., Desdoits-Lethimonier, C., Dejucq-Rainsford, N., Leffers, H., and Jégou, B. (2012). Paracetamol (acetaminophen), aspirin (acetylsalicylic acid) and indomethacin are anti-androgenic in the rat foetal testis. *Int. J. Androl.* 35, 377–384.
- Kristensen, D.M., Mazaud-Guittot, S., Gaudriault, P., Lesné, L., Serrano, T., Main, K.M., and Jégou, B. (2016). Analgesic use - prevalence, biomonitoring and endocrine and reproductive effects. *Nat. Rev. Endocrinol.* 12, 381–393.
- Krysiak-Baltyn, K., Toppari, J., Skakkebaek, N.E., Jensen, T.S., Virtanen, H.E., Schramm, K.-W., Shen, H., Vartiainen, T., Kiviranta, H., Taboureau, O., et al. (2010). Country-specific chemical signatures of persistent environmental compounds in breast milk. *Int. J. Androl.* 33, 270–278.
- Krysiak-Baltyn, K., Toppari, J., Skakkebaek, N.E., Jensen, T.S., Virtanen, H.E., Schramm, K.-W., Shen, H., Vartiainen, T., Kiviranta, H., Taboureau, O., et al. (2012). Association between chemical pattern in breast milk and congenital cryptorchidism: modelling of complex human exposures. *Int. J. Androl.* 35, 294–302.
- Kumar, T.R., Wang, Y., Lu, N., and Matzuk, M.M. (1997). Follicle stimulating hormone is required for ovarian follicle maturation but not male fertility. *Nat. Genet.* 15, 201–204.

- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Kheradpour, P., Zhang, Z., Heravi-Moussavi, A., Liu, Y., Amin, V., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* *518*, 317–330.
- Kuroki, S., Matoba, S., Akiyoshi, M., Matsumura, Y., Miyachi, H., Mise, N., Abe, K., Ogura, A., Wilhelm, D., Koopman, P., et al. (2013). Epigenetic regulation of mouse sex determination by the histone demethylase *Jmjd1a*. *Science* *341*, 1106–1109.
- Kusaka, M., Katoh-Fukui, Y., Ogawa, H., Miyabayashi, K., Baba, T., Shima, Y., Sugiyama, N., Sugimoto, Y., Okuno, Y., Kodama, R., et al. (2010). Abnormal Epithelial Cell Polarity and Ectopic Epidermal Growth Factor Receptor (EGFR) Expression Induced in *Emx2* KO Embryonic Gonads. *Endocrinology* *151*, 5893–5904.
- Kvam, V.M., Liu, P., and Si, Y. (2012). A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.* *99*, 248–256.
- LaFramboise, T. (2009). Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res.* *37*, 4181–4193.
- Lalli, E., and Sassone-Corsi, P. (2003). DAX-1, an unusual orphan receptor at the crossroads of steroidogenic function and sexual differentiation. *Mol. Endocrinol. Baltim. Md* *17*, 1445–1453.
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* *313*, 1929–1935.
- Lambrot, R., Muczynski, V., Lécureuil, C., Angenard, G., Coffigny, H., Pairault, C., Moison, D., Frydman, R., Habert, R., and Rouiller-Fabre, V. (2009). Phthalates impair germ cell development in the human fetal testis in vitro without change in testosterone production. *Environ. Health Perspect.* *117*, 32–37.
- Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* *44*, D862–868.
- Lane, a. H., and Donahoe, P.K. (1998). New insights into Mullerian inhibiting substance and its mechanism of action. *J. Endocrinol.* *158*, 1–6.
- Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A.C., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* *501*, 506–511.
- Lardenois, A., Gattiker, A., Collin, O., Chalmel, F., and Primig, M. (2010). GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database J. Biol. Databases Curation* *2010*.
- Larney, C., Bailey, T.L., and Koopman, P. (2014). Switching on sex: transcriptional regulation of the testis-determining gene *Sry*. *Development* *141*, 2195–2205.
- Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K., and Studholme, D.J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. Quantif.* *3*, 1–8.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* *15*, R29.
- Le Bouffant, R., Guerquin, M.J., Duquenne, C., Frydman, N., Coffigny, H., Rouiller-Fabre, V., Frydman, R., Habert, R., and Livera, G. (2010). Meiosis initiation in the human ovary requires intrinsic retinoic acid synthesis. *Hum. Reprod.* *25*, 2579–2590.
- Le Guyader, H. (2001). Qu'est-ce qu'un gène ? Une petite histoire du concept. *Courr. Environ. INRA* *53*–63.
- Lee, C.Y., and Ryan, R.J. (1973). Interaction of ovarian receptors with human luteinizing hormone and human chorionic gonadotropin. *Biochemistry* *12*, 4609–4615.

- Lee, P.A., Nordenström, A., Houk, C.P., Ahmed, S.F., Auchus, R., Baratz, A., Baratz Dalke, K., Liao, L.-M., Lin-Su, K., Looijenga, L.H.J., et al. (2016). Global Disorders of Sex Development Update since 2006: Perceptions, Approach and Care. *Horm. Res. Paediatr.* 85, 158–180.
- Lee, T.-L., Cheung, H.-H., Claus, J., Sastry, C., Singh, S., Vu, L., Rennert, O., and Chan, W.-Y. (2009). GermSAGE: a comprehensive SAGE database for transcript discovery on male germ cell development. *Nucleic Acids Res.* 37, D891-7.
- Lee, T.-L., Li, Y., Cheung, H.-H., Claus, J., Singh, S., Sastry, C., Rennert, O.M., Lau, Y.-F.C., and Chan, W.-Y. (2010). GonadSAGE: a comprehensive SAGE database for transcript discovery on male embryonic gonad development. *Bioinforma. Oxf. Engl.* 26, 585–586.
- Lee, T.-L., Rennert, O.M., and Chan, W.-Y. (2012). Revealing the transcriptome landscape of mouse spermatogonial cells by tiling microarray. *Methods Mol. Biol. Clifton NJ* 825, 75–92.
- Lei, L., and Spradling, A.C. (2016). Mouse oocytes differentiate through organelle enrichment from sister cyst germ cells. *Science* 352, 95–99.
- Leinonen, R., Sugawara, H., and Shumway, M. (2011). The Sequence Read Archive. *Nucleic Acids Res.* 39, D19–D21.
- Leitch, H.G., Tang, W.W.C., and Surani, M.A. (2013). Primordial germ-cell development and epigenetic reprogramming in mammals. *Curr. Top. Dev. Biol.* 104, 149–187.
- Leverrier-Penna, S., Mitchell, R.T., Becker, E., Lecante, L., Ben Maamar, M., Homer, N., Lavoué, V., Kristensen, D.M., Dejuçq-Rainsford, N., Jégou, B., et al. (2018). Ibuprofen is deleterious for the development of first trimester human fetal ovary ex vivo. *Hum. Reprod. Oxf. Engl.*
- Lherminier, P. (1998). L'hérédité avant la génétique. *médecine/sciences* 14, I.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
- Li, H., and Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinform.* 11, 473–483.
- Li, Y., and Tollefsbol, T.O. (2011). DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods Mol. Biol. Clifton NJ* 791, 11–21.
- Li, J., Li, Y.-X., and Li, Y.-Y. (2016). Differential Regulatory Analysis Based on Coexpression Network in Cancer Research. *BioMed Res. Int.* 2016.
- Li, J.-L., Mazar, J., Zhong, C., Faulkner, G.J., Govindarajan, S.S., Zhang, Z., Dinger, M.E., Meredith, G., Adams, C., Zhang, S., et al. (2013a). Genome-wide methylated CpG island profiles of melanoma cells reveal a melanoma coregulation network. *Sci. Rep.* 3, 2962.
- Li, L., Dong, J., Yan, L., Yong, J., Liu, X., Hu, Y., Fan, X., Wu, X., Guo, H., Wang, X., et al. (2017a). Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* 20, 891–892.
- Li, L., Zirkin, B.R., and Papadopoulos, V. (2018). Leydig Cell Androgen Synthesis. In *Encyclopedia of Reproduction (Second Edition)*, M.K. Skinner, ed. (Oxford: Academic Press), pp. 215–221.
- Li, M.J., Wang, P., Liu, X., Lim, E.L., Wang, Z., Yeager, M., Wong, M.P., Sham, P.C., Chanock, S.J., and Wang, J. (2012). GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* 40, D1047–D1054.
- Li, P., Piao, Y., Shon, H.S., and Ryu, K.H. (2015). Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC Bioinformatics* 16.
- Li, W., Yang, W., and Wang, X.-J. (2013b). Pseudogenes: Pseudo or Real Functional Elements? *J. Genet. Genomics* 40, 171–177.

- Li, X., Brock, G.N., Rouchka, E.C., Cooper, N.G.F., Wu, D., O'Toole, T.E., Gill, R.S., Eteleeb, A.M., O'Brien, L., and Rai, S.N. (2017b). A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. *PLoS One* *12*, e0176185.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* *30*, 923–930.
- Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinforma. Oxf. Engl.* *27*, i275–282.
- Lin, Y., Golovkina, K., Chen, Z.-X., Lee, H.N., Negron, Y.L.S., Sultana, H., Oliver, B., and Harbison, S.T. (2016). Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. *BMC Genomics* *17*, 28.
- Lin, Y.-T., Barske, L., DeFalco, T., and Capel, B. (2017). Numb regulates somatic cell lineage commitment during early gonadogenesis in mice. *Dev. Camb. Engl.* *144*, 1607–1618.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. (1999). High density synthetic oligonucleotide arrays. *Nat. Genet.* *21*, 20–24.
- Liu, B., Li, J., and Cairns, M.J. (2014). Identifying miRNAs, targets and functions. *Brief. Bioinform.* *15*, 1–19.
- Liu, C., Peng, J., Matzuk, M.M., and Yao, H.H.-C. (2015). Lineage specification of ovarian theca cells requires multicellular interactions via oocyte and granulosa cells. *Nat. Commun.* *6*, 6934.
- Liu, C., Rodriguez, K., and Yao, H.H.-C. (2016). Mapping lineage progression of somatic progenitor cells in the mouse fetal testis. *Dev. Camb. Engl.* *143*, 3700–3710.
- Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). *Molecular Cell Biology* (W. H. Freeman).
- Lord, T., and Oatley, J.M. (2017). A revised Asingle model to explain stem cell dynamics in the mouse male germline. *Reprod. Camb. Engl.* *154*, R55–R64.
- Lourenco, D., Brauner, R., Rybczynska, M., Nihoul-Fekete, C., McElreavey, K., and Bashamboo, A. (2011). Loss-of-function mutation in GATA4 causes anomalies of human testicular development. *Proc. Natl. Acad. Sci.* *108*, 1597–1602.
- Loveland, K.L. (2018). Activins in Male Reproduction. In *Encyclopedia of Reproduction* (Second Edition), M.K. Skinner, ed. (Oxford: Academic Press), pp. 208–214.
- Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* *14*, 265–279.
- Luk, A.C.-S., Gao, H., Xiao, S., Liao, J., Wang, D., Tu, J., Rennert, O.M., Chan, W.-Y., and Lee, T.-L. (2015). GermLncRNA: a unique catalogue of long non-coding RNAs and associated regulations in male germ cell development. *Database J. Biol. Databases Curation* *2015*, bav044.
- Lund, L., Engebjerg, M.C., Pedersen, L., Ehrenstein, V., Nørgaard, M., and Sørensen, H.T. (2009). Prevalence of hypospadias in Danish boys: a longitudinal study, 1977–2005. *Eur. Urol.* *55*, 1022–1026.
- Luo, L.-F., Hou, C.-C., and Yang, W.-X. (2016). Small non-coding RNAs and their associated proteins in spermatogenesis. *Gene* *578*, 141–157.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* *1*, 18.
- Luo, X., Ikeda, Y., and Parker, K.L. (1994). A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* *77*, 481–490.
- Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. *J. Mach. Learn. Res.* *9*, 2579–2605.
- Maatouk, D.M., DiNapoli, L., Alvers, A., Parker, K.L., Taketo, M.M., and Capel, B. (2008). Stabilization of β -catenin in XY gonads causes male-to-female sex-reversal. *Hum. Mol. Genet.* *17*, 2949–2955.

- Maatouk, D.M., Mork, L., Hinson, A., Kobayashi, A., McMahon, A.P., and Capel, B. (2012). Germ Cells Are Not Required to Establish the Female Pathway in Mouse Fetal Gonads. *PLoS ONE* 7, e47238.
- MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45, D896–D901.
- Macleod, J., and Gold, R.Z. (1951). The male factor in fertility and infertility. II. Spermatozoon counts in 1000 men of known fertility and in 1000 cases of infertile marriage. *J. Urol.* 66, 436–449.
- Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214.
- Madrigal, P., and Krajewski, P. (2015). Uncovering correlated variability in epigenomic datasets using the Karhunen-Loeve transform. *BioData Min.* 8, 20.
- Maheshwari, A., and Fowler, P.A. (2008). Primordial follicular assembly in humans – revisited. *Zygote* 16, 285–296.
- Mainwaring, W., Haining, S., and Harper, B. (1988). The functions of testosterone and its metabolites. *New Compr.*
- Malki, S., van der Heijden, G.W., O’Donnell, K.A., Martin, S.L., and Bortvin, A. (2014). A Role for Retrotransposon LINE-1 in Fetal Oocyte Attrition in Mice. *Dev. Cell* 29, 521–533.
- Mamsen, L.S., Brøchner, C.B., Byskov, A.G., and Møllgaard, K. (2012). The migration and loss of human primordial germ stem cells from the hind gut epithelium towards the gonadal ridge. *Int. J. Dev. Biol.* 56, 771–778.
- Mamsen, L.S., Ernst, E.H., Borup, R., Larsen, A., Olesen, R.H., Ernst, E., Anderson, R.A., Kristensen, S.G., and Andersen, C.Y. (2017). Temporal expression pattern of genes during the period of sex differentiation in human embryonic gonads. *Sci. Rep.* 7, 15961.
- Manuylov, N.L., Smagulova, F.O., Leach, L., and Tevosian, S.G. (2008). Ovarian development in mice requires the GATA4-FOG2 transcription complex. *Development* 135, 3731–3743.
- Marchese, F.P., Raimondi, I., and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* 18, 206.
- Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* 24, 496–510.
- Martineau, J., Nordqvist, K., Tilmann, C., Lovell-Badge, R., and Capel, B. (1997). Male-specific cell migration into the developing gonad. *Curr. Biol.* CB 7, 958–968.
- Matlai, P., and Beral, V. (1985). Trends in congenital malformations of external genitalia. *Lancet Lond. Engl.* 1, 108.
- Matson, C.K., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* 476, 101–104.
- Matsuda, F., Inoue, N., Manabe, N., and Ohkura, S. (2012). Follicular growth and atresia in mammalian ovaries: regulation by survival and death of granulosa cells. *J. Reprod. Dev.* 58, 44–50.
- Matsui, Y., Zsebo, K.M., and Hogan, B.L. (1990). Embryonic expression of a haematopoietic growth factor encoded by the Sl locus and the ligand for c-kit. *Nature* 347, 667–669.
- Mazaud, S., Oréal, E., Guigon, C.J., Carré-Eusèbe, D., and Magre, S. (2002). Lhx9 expression during gonadal morphogenesis as related to the state of cell differentiation. *Gene Expr. Patterns GEP* 2, 373–377.

- Mazaud, S., Guyot, R., Guigon, C.J., Coudouel, N., Le Magueresse-Battistoni, B., and Magre, S. (2005). Basal membrane remodeling during follicle histogenesis in the rat ovary: contribution of proteinases of the MMP and PA families. *Dev. Biol.* 277, 403–416.
- Mazaud-Guittot, S., Nicolaz, C.N., Desdoits-Lethimonier, C., Coiffec, I., Maamar, M.B., Balaguer, P., Kristensen, D.M., Chevrier, C., Lavoué, V., Poulain, P., et al. (2013). Paracetamol, Aspirin, and Indomethacin Induce Endocrine Disturbances in the Human Fetal Testis Capable of Interfering With Testicular Descent. *J. Clin. Endocrinol. Metab.* 98, E1757–E1767.
- McClelland, K.S., Bell, K., Larney, C., Harley, V.R., Sinclair, A.H., Oshlack, A., Koopman, P., and Bowles, J. (2015). Purification and Transcriptomic Analysis of Mouse Fetal Leydig Cells Reveals Candidate Genes for Specification of Gonadal Steroidogenic Cells¹. *Biol. Reprod.* 92, 1–12.
- McCoshen, J.A. (1982). In vivo sex differentiation of congenic germinal cell aplastic gonads. *Am. J. Obstet. Gynecol.* 142, 83–88.
- McCoshen, J.A. (1983). Quantitation of sex chromosomal influence(s) on the somatic growth of fetal gonads in vivo. *Am. J. Obstet. Gynecol.* 145, 469–473.
- McEwan, I.J., and Smith, L.B. (2018). Androgen Receptor. In *Encyclopedia of Reproduction (Second Edition)*, M.K. Skinner, ed. (Oxford: Academic Press), pp. 222–226.
- McLaren, A. (1984). Meiosis and differentiation of mouse germ cells. *Symp. Soc. Exp. Biol.* 38, 7–23.
- McLaren, A. (1991). Development of the mammalian gonad: the fate of the supporting cell lineage. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 13, 151–156.
- Meeks, J.J. (2003). Dax1 regulates testis cord organization during gonadal differentiation. *Development* 130, 1029–1036.
- Meeks, J.J., Weiss, J., and Jameson, J.L. (2003a). Dax1 is required for testis determination. *Nat. Genet.* 34, 32–33.
- Meeks, J.J., Russell, T.A., Jeffs, B., Huhtaniemi, I., Weiss, J., and Jameson, J.L. (2003b). Leydig cell-specific expression of DAX1 improves fertility of the Dax1-deficient mouse. *Biol. Reprod.* 69, 154–160.
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877.
- Mendis, S.H.S., Meachem, S.J., Sarraj, M.A., and Loveland, K.L. (2011). Activin A balances Sertoli and germ cell proliferation in the fetal mouse testis. *Biol. Reprod.* 84, 379–391.
- Merchant, H. (1975). Rat gonadal and ovarioan organogenesis with and without germ cells. An ultrastructural study. *Dev. Biol.* 44, 1–21.
- Merchant-Larios, H., Moreno-Mendoza, N., and Buehr, M. (1993). The role of the mesonephros in cell differentiation and morphogenesis of the mouse fetal testis. *Int. J. Dev. Biol.* 37, 407–415.
- von Meyenn, F., and Reik, W. (2015). Forget the Parents: Epigenetic Reprogramming in Human Germ Cells. *Cell* 161, 1248–1251.
- von Meyenn, F., Berrens, R.V.V., Andrews, S., Santos, F., Collier, A.J.J., Krueger, F., Osorno, R., Dean, W., Rugg-Gunn, P.J.J., Reik, W., et al. (2016). Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. *Dev. Cell* 39, 104–115.
- Milham, M., Craddock, C., Fleischmann, M., Son, J., Clucas, J., Xu, H., Koo, B., Krishnakumar, A., Biswal, B., Castellanos, F., et al. (2017). Assessment of the impact of shared data on the scientific literature. *BioRxiv* 183814.
- Minkina, A., Matson, C.K., Lindeman, R.E., Ghyselinck, N.B., Bardwell, V.J., and Zarkower, D. (2014). DMRT1 Protects Male Gonadal Cells from Retinoid-Dependent Sexual Transdifferentiation. *Dev. Cell* 29, 511–520.

- Miossec, P., Archambeaud-Mouvieroux, F., and Teissier, M.P. (1997). [Inhibition of steroidogenesis by ketoconazole. Therapeutic uses]. *Ann. Endocrinol.* 58, 494–502.
- Mistry, J., Bateman, A., and Finn, R.D. (2007). Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics* 8, 298.
- Miyabayashi, K., Katoh-Fukui, Y., Ogawa, H., Baba, T., Shima, Y., Sugiyama, N., Kitamura, K., and Morohashi, K. (2013). Aristaless related homeobox gene, *Arx*, is implicated in mouse fetal Leydig cell differentiation possibly through expressing in the progenitor cells. *PLoS One* 8, e68050.
- Miyamoto, Y., Taniguchi, H., Hamel, F., Silversides, D.W., and Viger, R.S. (2008). A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex determination and differentiation. *BMC Mol. Biol.* 9, 44.
- Moalem, S., Babul-Hirji, R., Stavropolous, D.J., Wherrett, D., Bägli, D.J., Thomas, P., and Chitayat, D. (2012). XX male sex reversal with genital abnormalities associated with a de novo *SOX3* gene duplication. *Am. J. Med. Genet. A.* 158A, 1759–1764.
- Mockler, T.C., Ecker, J.R., Sundaresan, A., Chen, H., Jacobsen, S.E., and Ecker, J.R. (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics* 85, 1–15.
- Moffatt, M.F., Kabesch, M., Liang, L., Dixon, A.L., Strachan, D., Heath, S., Depner, M., von Berg, A., Bufe, A., Rietschel, E., et al. (2007). Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* 448, 470–473.
- MØLLER, H. (1998). Trends in sex-ratio, testicular cancer and male reproductive hazards: Are they connected? *APMIS* 106, 232–239.
- Møllgård, K., Jespersen, A., Lutterodt, M.C., Yding Andersen, C., Høyer, P.E., and Byskov, A.G. (2010). Human primordial germ cells migrate along nerve fibers and Schwann cells from the dorsal hind gut mesentery to the gonadal ridge. *Mol. Hum. Reprod.* 16, 621–631.
- Molyneaux, K.A., Zinszner, H., Kunwar, P.S., Schaible, K., Stebler, J., Sunshine, M.J., O'Brien, W., Raz, E., Littman, D., Wylie, C., et al. (2003). The chemokine *SDF1/CXCL12* and its receptor *CXCR4* regulate mouse germ cell migration and survival. *Dev. Camb. Engl.* 130, 4279–4286.
- Moniot, B., Declosmenil, F., Barrionuevo, F., Scherer, G., Aritake, K., Malki, S., Marzi, L., Cohen-Solal, A., Georg, I., Klattig, J., et al. (2009). The *PGD2* pathway, independently of *FGF9*, amplifies *SOX9* activity in Sertoli cells during male sexual differentiation. *Development* 136, 1813–1821.
- Monniaux, D., Drouilhet, L., Rico, C., Estienne, A., Jarrier, P., Touzé, J.-L., Sapa, J., Phocas, F., Dupont, J., Dalbiès-Tran, R., et al. (2012). Regulation of anti-Müllerian hormone production in domestic animals. *Reprod. Fertil. Dev.* 25, 1–16.
- Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R., and Dermitzakis, E.T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464, 773–777.
- Moorthy, A.V., Chesney, R.W., and Lubinsky, M. (1987). Chronic renal failure and XY gonadal dysgenesis: “Frasier” syndrome—a commentary on reported cases. *Am. J. Med. Genet. Suppl.* 3, 297–302.
- Moraes, F., and Góes, A. (2016). A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochem. Mol. Biol. Educ. Bimon. Publ. Int. Union Biochem. Mol. Biol.* 44, 215–223.
- Moreno, S.G., Attali, M., Allemand, I., Messiaen, S., Fouchet, P., Coffigny, H., Romeo, P.-H., and Habert, R. (2010). TGFbeta signaling in male germ cells regulates gonocyte quiescence and fertility in mice. *Dev. Biol.* 342, 74–84.
- Mork, L., Maatouk, D.M., McMahon, J.A., Guo, J.J., Zhang, P., McMahon, A.P., and Capel, B. (2012). Temporal Differences in Granulosa Cell Specification in the Ovary Reflect Distinct Follicle Fates in Mice. *Biol. Reprod.* 86.

- Ostrer, H., Huang, H.Y., Masch, R.J., and Shapiro, E. (2007). A cellular study of human testis development. *Sex. Dev.* *1*, 286–292.
- Ottolenghi, C., Pelosi, E., Tran, J., Colombino, M., Douglass, E., Nedorezov, T., Cao, A., Forabosco, A., and Schlessinger, D. (2007). Loss of Wnt4 and Foxl2 leads to female-to-male sex reversal extending to germ cells. *Hum. Mol. Genet.* *16*, 2795–2804.
- Oyama, M., Itagaki, C., Hata, H., Suzuki, Y., Izumi, T., Natsume, T., Isobe, T., and Sugano, S. (2004). Analysis of Small Human Proteins Reveals the Translation of Upstream Open Reading Frames of mRNAs. *Genome Res.* *14*, 2048–2052.
- Oyelade, J., Isewon, I., Oladipupo, F., Aromolaran, O., Uwoghiren, E., Ameh, F., Achas, M., and Adebisi, E. (2016). Clustering Algorithms: Their Application to Gene Expression Data. *Bioinforma. Biol. Insights* *10*, 237–253.
- Palmer, M.S., Sinclair, A.H., Berta, P., Ellis, N.A., Goodfellow, P.N., Abbas, N.E., and Fellous, M. (1989). Genetic evidence that ZFY is not the testis-determining factor. *Nature* *342*, 937–939.
- Pannetier, M., Chassot, A.-A., Chaboissier, M.-C., and Pailhoux, E. (2016). Involvement of FOXL2 and RSPO1 in Ovarian Determination, Development, and Maintenance in Mammals. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determ. Differ.* *10*, 167–184.
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Sci. Rep.* *6*, 25533.
- Park, S.Y., Tong, M., and Jameson, J.L. (2007). Distinct roles for steroidogenic factor 1 and desert hedgehog pathways in fetal and adult Leydig cell development. *Endocrinology* *148*, 3704–3710.
- Parma, P., Radi, O., Vidal, V., Chaboissier, M.C., Dellambra, E., Valentini, S., Guerra, L., Schedl, A., and Camerino, G. (2006). R-spondin1 is essential in sex determination, skin differentiation and malignancy. *Nat. Genet.* *38*, 1304–1309.
- Patro, R., Mount, S.M., and Kingsford, C. (2014). Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat. Biotechnol.* *32*, 462–464.
- Paulozzi, L.J. (1999). International trends in rates of hypospadias and cryptorchidism. *Environ. Health Perspect.* *107*, 297–302.
- Paulozzi, L.J., Erickson, J.D., Jackson, R.J., Angerpointer, T., Källén, B., Bertolini, E., Castilla, A., Harris, E., Beaty, T., Allen, T., et al. (1997). Hypospadias trends in two US surveillance systems. *Pediatrics* *100*, 831–834.
- Pavlopoulou, A., and Michalopoulos, I. (2011). State-of-the-art bioinformatics protein structure prediction tools (Review). *Int. J. Mol. Med.* *28*, 295–310.
- Payne, A.H., and Hales, D.B. (2004). Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. *Endocr. Rev.* *25*, 947–970.
- Pearlman, A., Loke, J., Le Caignec, C., White, S., Chin, L., Friedman, A., Warr, N., Willan, J., Brauer, D., Farmer, C., et al. (2010). Mutations in MAP3K1 Cause 46,XY Disorders of Sex Development and Implicate a Common Signal Transduction Pathway in Human Testis Determination.
- Pek, J.W., and Okamura, K. (2015). Regulatory RNAs discovered in unexpected places. *Wiley Interdiscip. Rev. RNA* *6*, 671–686.
- Pelletier, G., Labrie, C., and Labrie, F. (2000). Localization of oestrogen receptor alpha, oestrogen receptor beta and androgen receptors in the rat reproductive organs. *J. Endocrinol.* *165*, 359–370.
- Pepling, M.E. (2012). Follicular assembly: mechanisms of action. *Reprod. Camb. Engl.* *143*, 139–149.
- Pepling, M.E., Sundman, E.A., Patterson, N.L., Gephardt, G.W., Medico, L., and Wilson, K.I. (2010). Differences in oocyte development and estradiol sensitivity among mouse strains. *Reproduction* *139*, 349–357.
- Perdigão, N., Heinrich, J., Stolte, C., Sabir, K.S., Buckley, M.J., Tabor, B., Signal, B., Gloss, B.S., Hammang, C.J., Rost, B., et al. (2015). Unexpected features of the dark proteome. *Proc. Natl. Acad. Sci.* *112*, 15898–15903.

- Perteau, M., Perteau, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* *33*, 290–295.
- Petersen, C., and Söder, O. (2006). The Sertoli cell - A hormonal target and “super” nurse for germ cells that determines testicular size. *Horm. Res.*
- Philibert, P., Biason-Lauber, A., Gueorguieva, I., Stuckens, C., Pienkowski, C., Lebon-Labich, B., Paris, F., and Sultan, C. (2011). Molecular analysis of WNT4 gene in four adolescent girls with müllerian duct abnormality and hyperandrogenism (atypical Mayer-Rokitansky-Küster-Hauser syndrome). *Fertil. Steril.* *95*, 2683–2686.
- Phillippy, A.M. (2017). New advances in sequence assembly. *Genome Res.* *27*, xi–xiii.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* *10*, 1096–1098.
- Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* *464*, 768–772.
- Pierucci-Alves, F., Clark, A.M., and Russell, L.D. (2001). A developmental study of the Desert hedgehog-null mouse testis. *Biol. Reprod.* *65*, 1392–1402.
- Pirim, H., Eksioğlu, B., Perkins, A., and Yüceer, C. (2012). Clustering of High Throughput Gene Expression Data. *Comput. Oper. Res.* *39*, 3046–3061.
- Pitetti, J.-L., Calvel, P., Zimmermann, C., Conne, B., Papaioannou, M.D., Aubry, F., Cederroth, C.R., Uner, F., Fumel, B., Crausaz, M., et al. (2013a). An essential role for insulin and IGF1 receptors in regulating sertoli cell proliferation, testis size, and FSH action in mice. *Mol. Endocrinol. Baltim. Md* *27*, 814–827.
- Pitetti, J.-L., Calvel, P., Romero, Y., Conne, B., Truong, V., Papaioannou, M.D., Schaad, O., Docquier, M., Herrera, P.L., Wilhelm, D., et al. (2013b). Insulin and IGF1 Receptors Are Essential for XX and XY Gonadal Differentiation and Adrenal Development in Mice. *PLoS Genet.* *9*, e1003160.
- Polanco, J.C., and Koopman, P. (2007). Sry and the hesitant beginnings of male development. *Dev. Biol.* *302*, 13–24.
- Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* *32*, 1053–1058.
- Pollier, J., Rombauts, S., and Goossens, A. (2013). Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol. Biol. Clifton NJ* *1011*, 305–315.
- Pop, R., Zaragoza, M.V., Gaudette, M., Dohrmann, U., and Scherer, G. (2005). A homozygous nonsense mutation in SOX9 in the dominant disorder campomelic dysplasia: a case of mitotic gene conversion. *Hum. Genet.* *117*, 43–53.
- Porrás-Gómez, T.J., and Moreno-Mendoza, N. (2017). Neo-oogenesis in mammals. *Zygote Camb. Engl.* *25*, 404–422.
- Pozhitkov, A.E., Tautz, D., and Noble, P.A. (2007). Oligonucleotide microarrays: widely applied poorly understood. *Brief. Funct. Genomic. Proteomic.* *6*, 141–148.
- Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., et al. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* *42*, D756–D763.
- Pundir, S., Magrane, M., Martin, M.J., O’Donovan, C., and UniProt Consortium (2015). Searching and Navigating UniProt Databases. *Curr. Protoc. Bioinforma.* *50*, 1.27.1-10.
- Purdue, M.P., Devesa, S.S., Sigurdson, A.J., and McGlynn, K.A. (2005). International patterns and trends in testis cancer incidence. *Int. J. Cancer* *115*, 822–827.

- Putney, S.D., Herlihy, W.C., and Schimmel, P. (1983). A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* *302*, 718–721.
- Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* *17*, 47–62.
- Rabinovici, J., and Jaffe, R.B. (1990). Development and regulation of growth and differentiated function in human and subhuman primate fetal gonads. *Endocr Rev* *11*, 532–557.
- Racine, C., Rey, R., Forest, M.G., Louis, F., Ferré, A., Huhtaniemi, I., Josso, N., and di Clemente, N. (1998). Receptors for anti-müllerian hormone on Leydig cells are responsible for its effects on steroidogenesis and cell differentiation. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 594–599.
- Radovick, S., Levine, J.E., and Wolfe, A. (2012). Estrogenic Regulation of the GnRH Neuron. *Front. Endocrinol.* *3*.
- Rahman, N.A., and Rao, C.V. (2009). Recent progress in luteinizing hormone/human chorionic gonadotrophin hormone research. *Mol. Hum. Reprod.*
- Rahmoun, M., Lavery, R., Laurent-Chaballier, S., Bellora, N., Philip, G.K., Rossitto, M., Symon, A., Pailhoux, E., Cammas, F., Chung, J., et al. (2017). In mammalian foetal testes, SOX9 regulates expression of its target genes by binding to genomic regions with conserved signatures. *Nucleic Acids Res.* *45*, 7191–7211.
- Rainsford, K.D. (2009). Ibuprofen: pharmacology, efficacy and safety. *Inflammopharmacology* *17*, 275–342.
- Rajkovic, A., Pangas, S.A., Ballow, D., Suzumori, N., and Matzuk, M.M. (2004). NOBOX deficiency disrupts early folliculogenesis and oocyte-specific gene expression. *Science* *305*, 1157–1159.
- Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* *30*, 777–782.
- Reijo, R.A., Dorfman, D.M., Slee, R., Renshaw, A.A., Loughlin, K.R., Cooke, H., and Page, D.C. (2000). DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. *Biol. Reprod.* *63*, 1490–1496.
- Reyes, A., and Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* *46*, 582–592.
- Reynaud, K., Cortvrindt, R., Verlinde, F., De Schepper, J., Bourgain, C., and Smitz, J. (2004). Number of ovarian follicles in human fetuses with the 45,X karyotype. *Fertil. Steril.* *81*, 1112–1119.
- Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* *13*, 278–289.
- Richard, G.-F., Kerrest, A., and Dujon, B. (2008). Comparative Genomics and Molecular Dynamics of DNA Repeats in Eukaryotes. *Microbiol. Mol. Biol. Rev.* *72*, 686–727.
- Richards, J.S., and Pangas, S.A. (2010). The ovary: basic biology and clinical implications. *J. Clin. Invest.* *120*, 963–972.
- Richardson, B.E., and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat. Rev. Mol. Cell Biol.* *11*, 37–49.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* *43*, e47.
- Robertson, D.M. (2018). Inhibin and Male. In *Encyclopedia of Reproduction (Second Edition)*, M.K. Skinner, ed. (Oxford: Academic Press), pp. 204–207.
- Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* *11*, R25.

- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* 26, 139–140.
- Robles, J.A., Qureshi, S.E., Stephen, S.J., Wilson, S.R., Burden, C.J., and Taylor, J.M. (2012). Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics* 13, 484.
- Rolland, A.D., Lehmann, K.P., Johnson, K.J., Gaido, K.W., and Koopman, P. (2011). Uncovering gene regulatory networks during mouse fetal germ cell development. *Biol. Reprod.* 84, 790–800.
- Ross, A.J., Tilman, C., Yao, H., MacLaughlin, D., and Capel, B. (2003). AMH induces mesonephric cell migration in XX gonads. *Mol. Cell. Endocrinol.* 211, 1–7.
- Rouget, C., Papin, C., Boureux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature* 467, 1128–1132.
- Rouiller-Fabre, V., Muczynski, V., Lambrot, R., Lécureuil, C., Coffigny, H., Pairault, C., Moison, D., Angenard, G., Bakalska, M., Courtot, A.M., et al. (2009). Ontogenesis of testicular function in humans. *Folia Histochem. Cytobiol. Pol. Acad. Sci. Pol. Histochem. Cytochem. Soc.* 47, S19–S24.
- Ruiz-Arana, I.-L., Hübner, A., Cetingdag, C., Krude, H., Grüters, A., Fukami, M., Biebermann, H., and Köhler, B. (2015). A novel hemizygous mutation of MAMLD1 in a patient with 46,XY complete gonadal dysgenesis. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determ. Differ.* 9, 80–85.
- Sales, G., Coppe, A., Bisognin, A., Biasiolo, M., Bortoluzzi, S., and Romualdi, C. (2010). MAGIA, a web-based tool for miRNA and Genes Integrated Analysis. *Nucleic Acids Res.* 38, W352–359.
- Samanta, M.P., Tongprasit, W., and Stolc, V. (2007). In-depth query of large genomes using tiling arrays. *Methods Mol. Biol. Clifton NJ* 377, 163–174.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467.
- de Santa Barbara, P., Méjean, C., Moniot, B., Malclès, M.H., Berta, P., and Boizet-Bonhoure, B. (2001). Steroidogenic factor-1 contributes to the cyclic-adenosine monophosphate down-regulation of human SRY gene expression. *Biol. Reprod.* 64, 775–783.
- Schatz, M.C. (2017). Nanopore sequencing meets epigenetics. *Nat. Methods* 14, 347–348.
- Scheetz, D., Folger, J.K., Smith, G.W., and Ireland, J.J. (2012). Granulosa cells are refractory to FSH action in individuals with a low antral follicle count. *Reprod. Fertil. Dev.* 24, 327–336.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.
- Schug, T.T., Johnson, A.F., Birnbaum, L.S., Colborn, T., Guillette, L.J., Crews, D.P., Collins, T., Soto, A.M., vom Saal, F.S., McLachlan, J.A., et al. (2016). Minireview: Endocrine Disruptors: Past Lessons and Future Directions. *Mol. Endocrinol.* 30, 833–847.
- Schuster, A., Tang, C., Xie, Y., Ortogero, N., Yuan, S., and Yan, W. (2016). SpermBase: A Database for Sperm-Borne RNA Contents. *Biol. Reprod.* 95, 99–99.
- Scott, H.M., Mason, J.I., and Sharpe, R.M. (2009). Steroidogenesis in the fetal testis and its susceptibility to disruption by exogenous compounds. *Endocr. Rev.* 30, 883–925.
- Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* 453, 930–934.
- Sekido, R., and Lovell-Badge, R. (2013). Genetic Control of Testis Development. *Sex. Dev.* 7, 21–32.
- Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M., and Ligterink, W. (2016). Learning from Co-expression Networks: Possibilities and Challenges. *Front. Plant Sci.* 7, 444.

- Serrano, T., Chevrier, C., Multigner, L., Cordier, S., and Jegou, B. (2013). International geographic correlation study of the prevalence of disorders of male reproductive health. *Hum. Reprod.* 28, 1974–1986.
- Seyednasrollah, F., Laiho, A., and Elo, L.L. (2015). Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinform.* 16, 59–70.
- Shanmugalingam, T., Soultati, A., Chowdhury, S., Rudman, S., and Van Hemelrijck, M. (2013). Global incidence and outcome of testicular cancer. *Clin. Epidemiol.* 5, 417–427.
- Sharpe, R.M. (2012). Sperm counts and fertility in men: a rocky road ahead. *EMBO Rep.* 13, 398–403.
- She, Z.-Y., and Yang, W.-X. (2016). Sry and SoxE genes: How they participate in mammalian sex determination and gonadal development? *Semin. Cell Dev. Biol.* 63, 13–22.
- Sheikh, I.A., Yasir, M., Abu-Elmagd, M., Dar, T.A., Abuzenadah, A.M., Damanhour, G.A., Al-Qahtani, M., and Beg, M.A. (2016). Human sex hormone-binding globulin as a potential target of alternate plasticizers: an in silico study. *BMC Struct. Biol.* 16, 15.
- Shi, X., Sun, M., Liu, H., Yao, Y., and Song, Y. (2013). Long non-coding RNAs: a new frontier in the study of human diseases. *Cancer Lett.* 339, 159–166.
- Shi, X., Nie, F., Wang, Z., and Sun, M. (2016). Pseudogene-expressed RNAs: a new frontier in cancers. *Tumor Biol.* 37, 1471–1478.
- Shilin, Z., Guo, Y., and Shyr, Y. (2014). KEGGprofile: An annotation and visualization package for multi-types and multi-groups expression data in KEGG pathway.
- Shima, Y., Miyabayashi, K., Haraguchi, S., Arakawa, T., Otake, H., Baba, T., Matsuzaki, S., Shishido, Y., Akiyama, H., Tachibana, T., et al. (2013). Contribution of Leydig and Sertoli Cells to Testosterone Production in Mouse Fetal Testes. *Mol. Endocrinol.* 27, 63–73.
- Shimamura, R., Fraizer, G.C., Trapman, J., Lau YfC, and Saunders, G.F. (1997). The Wilms' tumor gene WT1 can regulate genes involved in sex determination and differentiation: SRY, Müllerian-inhibiting substance, and the androgen receptor. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 3, 2571–2580.
- Shimasaki, S., Moore, R.K., Otsuka, F., and Erickson, G.F. (2004). The Bone Morphogenetic Protein System In Mammalian Reproduction. *Endocr. Rev.* 25, 72–101.
- Shinawi, M., and Cheung, S.W. (2008). The array CGH and its clinical applications. *Drug Discov. Today* 13, 760–770.
- Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346, 240–244.
- Singh, R.P., and Carr, D.H. (1966). The anatomy and histology of XO human embryos and fetuses. *Anat. Rec.* 155, 369–383.
- Sirisena, N.D., McElreavey, K., Bashamboo, A., de Silva, K.S.H., Jayasekara, R.W., and Dissanayake, V.H.W. (2014). A child with a novel de novo mutation in the aristaless domain of the aristaless-related homeobox (ARX) gene presenting with ambiguous genitalia and psychomotor delay. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determ. Differ.* 8, 156–159.
- Skakkebaek, N.E., Rajpert-De Meyts, E., Buck Louis, G.M., Toppari, J., Andersson, A., Eisenberg, M.L., Jensen, T.K., Jørgensen, N., Swan, S.H., Sapra, K.J., et al. (2016). Male Reproductive Disorders and Fertility Trends: Influences of Environment and Genetic Susceptibility. *Physiol. Rev.* 96, 55–97.
- Skakkebak, N.E., Meyts, E.R.-D., and Main, K.M. (2001). Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum. Reprod.* 16, 972–978.
- Skytner, M.J., Sim, J.A., and Herbison, A.E. (1999). Detection of estrogen receptor alpha and beta messenger ribonucleic acids in adult gonadotropin-releasing hormone neurons. *Endocrinology* 140, 5195–5201.

- Small, C.L., Shima, J.E., Uzumcu, M., Skinner, M.K., and Griswold, M.D. (2005). Profiling Gene Expression During the Differentiation and Development of the Murine Embryonic Gonad. *72*, 492–501.
- Snyder, P.J., Ellenberg, S.S., and Farrar, J.T. (2016). Testosterone Treatment in Older Men. *N. Engl. J. Med.* *375*, 90.
- Soh, Y.Q.S., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A., and Page, D.C. (2015). A Gene Regulatory Program for Meiotic Prophase in the Fetal Ovary. *PLoS Genet.* *11*, e1005531.
- Sotero-Caio, C.G., Platt, R.N., Suh, A., and Ray, D.A. (2017). Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol. Evol.* *9*, 161–177.
- Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthès, P., Kokkinaki, M., Nef, S., Gnirke, A., et al. (2013). Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. *Cell Rep.* *3*, 2179–2190.
- Soyal, S.M., Amlah, A., and Dean, J. (2000). FIGalpha, a germ cell-specific transcription factor required for ovarian follicle formation. *Development* *127*, 4645–4654.
- Speir, M.L., Zweig, A.S., Rosenbloom, K.R., Raney, B.J., Paten, B., Nejad, P., Lee, B.T., Learned, K., Karolchik, D., Hinrichs, A.S., et al. (2016). The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.* *44*, D717–D725.
- Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., Flavell, R.A., Bien-Willner, G., Lupski, J., Cheung, S., Stankiewicz, P., Sippell, W., et al. (2005). Interchromosomal associations between alternatively expressed loci. *Nature* *435*, 637–645.
- Spiller, C.M., Feng, C.-W., Jackson, A., Gillis, A.J.M., Rolland, A.D., Looijenga, L.H.J., Koopman, P., and Bowles, J. (2012). Endogenous Nodal signaling regulates germ cell potency during mammalian testis development. *Dev. Camb. Engl.* *139*, 4123–4132.
- Sproll, P., Eid, W., Gomes, C.R., Mendonca, B.B., Gomes, N.L., Costa, E.M.-F., and Biason-Lauber, A. (2018). Assembling the jigsaw puzzle: CBX2 isoform 2 and its targets in disorders/differences of sex development. *Mol. Genet. Genomic Med.*
- Stasiak, A. (2001). Rosalind Franklin. *EMBO Rep.* *2*, 181.
- Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* *16*, 133–145.
- Stévant, I., Neirijnck, Y., Borel, C., Escoffier, J., Smith, L.B., Antonarakis, S.E., Dermitzakis, E.T., and Nef, S. (2018). Deciphering Cell Lineage Specification during Male Sex Determination with Single-Cell RNA Sequencing. *Cell Rep.* *22*, 1589–1599.
- Strauss, B.S. (2016). Biochemical Genetics and Molecular Biology: The Contributions of George Beadle and Edward Tatum. *Genetics* *203*, 13–20.
- Sumpter, J.P., and Jobling, S. (1995). Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. *Environ. Health Perspect.* 173–178.
- Sun, J., Lin, Y., and Wu, J. (2013a). Long non-coding RNA expression profiling of mouse testis during postnatal development. *PloS One* *8*, e75750.
- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., Liu, Y., Chen, R., and Zhao, Y. (2013b). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* *41*, e166.
- Sutton, E., Hughes, J., White, S., Sekido, R., Tan, J., Arboleda, V., Rogers, N., Knowler, K., Rowley, L., Eyre, H., et al. (2011). Identification of SOX3 as an XX male sex reversal gene in mice and humans. *J. Clin. Invest.* *121*, 328–341.
- Suzuki, A., and Saga, Y. (2008). Nanos2 suppresses meiosis and promotes male germ cell differentiation. *Genes Dev.* *22*, 430–435.

- Svechnikov, K., Landreh, L., Weisser, J., Izzo, G., Colón, E., Svechnikov, I., and Söder, O. (2010). Origin, Development and Regulation of Human Leydig Cells. *Horm. Res. Paediatr.* *73*, 93–101.
- Svingen, T., and Koopman, P. (2013). Building the mammalian testis: Origins, differentiation, and assembly of the component cell populations. *Genes Dev.* *27*, 2409–2426.
- Swain, A., Narvaez, V., Burgoyne, P., Camerino, G., and Lovell-Badge, R. (1998). Dax1 antagonizes Sry action in mammalian sex determination. *Nature* *391*, 761–767.
- Swan, S.H., Brazil, C., Drobnis, E.Z., Liu, F., Kruse, R.L., Hatch, M., Redmon, J.B., Wang, C., Overstreet, J.W., and Study For Future Families Research Group (2003). Geographic differences in semen quality of fertile U.S. males. *Environ. Health Perspect.* *111*, 414–420.
- Tachibana, M. (2015). Epigenetic regulation of mammalian sex determination. *J. Med. Investig. JMI* *62*, 19–23.
- Tanaka, S.S., and Nishinakamura, R. (2014). Regulation of male sex determination: genital ridge formation and Sry activation in mice. *Cell. Mol. Life Sci. CMLS* *71*, 4781–4802.
- Tang, H., Brennan, J., Karl, J., Hamada, Y., Raetzman, L., and Capel, B. (2008). Notch signaling maintains Leydig progenitor cells in the mouse testis. *Dev. Camb. Engl.* *135*, 3745–3753.
- Tang, W.W.C., Dietmann, S., Irie, N., Leitch, H.G., Floros, V.I., Bradshaw, C.R., Hackett, J.A., Chinnery, P.F., and Surani, M.A. (2015). A Unique Gene Regulatory Network Resets the Human Germline Epigenome for Development. *Cell* *161*, 1453–1467.
- Tang, W.W.C., Kobayashi, T., Irie, N., Dietmann, S., and Surani, M.A. (2016a). Specification and epigenetic programming of the human germ line. *Nat. Rev. Genet.* *17*, 585–600.
- Tang, W.W.C., Kobayashi, T., Irie, N., Dietmann, S., and Surani, M.A. (2016b). Specification and epigenetic programming of the human germ line. *Nat. Rev. Genet.* *17*, 585–600.
- d'Anglemont de Tassigny, X., and Colledge, W.H. (2010). The role of kisspeptin signaling in reproduction. *Physiol. Bethesda Md* *25*, 207–217.
- Teixeira, J., Fynn-Thompson, E., Payne, a. H., and Donahoe, P.K. (1999). Mullerian-inhibiting substance regulates androgen synthesis at the transcriptional level. *Endocrinology* *140*, 4732–4738.
- Tevosian, S.G., Albrecht, K.H., Crispino, J.D., Fujiwara, Y., Eicher, E.M., and Orkin, S.H. (2002). Gonadal differentiation, sex determination and normal Sry expression in mice require direct interaction between transcription partners GATA4 and FOG2. *Dev. Camb. Engl.* *129*, 4627–4634.
- The Gene Ontology Consortium (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* *45*, D331–D338.
- Themmen, A.P.N., and Huhtaniemi, I.T. (2000). Mutations of Gonadotropins and Gonadotropin Receptors: Elucidating the Physiology and Pathophysiology of Pituitary-Gonadal Function. *Endocr. Rev.* *21*, 551–583.
- Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* *14*, 178–192.
- Tingen, C., Kim, A., and Woodruff, T.K. (2009). The primordial pool of follicles and nest breakdown in mammalian ovaries. *Mol. Hum. Reprod.* *15*, 795–803.
- Tipson, R.S. (1957). Phoebus Aaron Theodor Levene, 1869-1940. *Adv. Carbohydr. Chem.* *12*, 1–12.
- Tomaselli, S., Megiorni, F., De Bernardo, C., Felici, A., Marrocco, G., Maggiulli, G., Grammatico, B., Remotti, D., Saccucci, P., Valentini, F., et al. (2008). Syndromic true hermaphroditism due to an R-spondin1 (*RSPO1*) homozygous mutation. *Hum. Mutat.* *29*, 220–226.
- Tomaselli, S., Megiorni, F., Lin, L., Mazzilli, M.C., Gerrelli, D., Majore, S., Grammatico, P., and Achermann, J.C. (2011). Human *RSPO1/R-spondin1* Is Expressed during Early Ovary Development and Augments β -Catenin Signaling. *PLoS ONE* *6*, e16366.

- Toppari, J., Larsen, J.C., Christiansen, P., Giwercman, A., Grandjean, P., Guillette, L.J., Jégou, B., Jensen, T.K., Jouannet, P., Keiding, N., et al. (1996). Male reproductive health and environmental xenoestrogens. *Environ. Health Perspect.* 741–803.
- Toppari, J., Kaleva, M., Virtanen, H.E., Main, K.M., and Skakkebaek, N.E. (2007). Luteinizing hormone in testicular descent. *Mol. Cell. Endocrinol.* 269, 34–37.
- Tostain, J., Rossi, D., and Martin, P.M. (2004). [Physiology of androgens in adult men]. *Progres En Urol. J. Assoc. Francaise Urol. Soc. Francaise Urol.* 14, 639–660.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.
- Trautmann, E., Guerquin, M.-J., Duquenne, C., Lahaye, J.-B., Habert, R., and Livera, G. (2008). Retinoic acid prevents germ cell mitotic arrest in mouse fetal testes. *Cell Cycle Georget. Tex* 7, 656–664.
- Tremblay, J.J., and Viger, R.S. (1999). Transcription Factor GATA-4 Enhances Müllerian Inhibiting Substance Gene Transcription through a Direct Interaction with the Nuclear Receptor SF-1. *Mol. Endocrinol.* 13, 1388–1401.
- Tress, M.L., Abascal, F., and Valencia, A. (2017). Alternative Splicing May Not Be the Key to Proteome Complexity. *Trends Biochem. Sci.* 42, 98–110.
- Trombly, D.J., Woodruff, T.K., and Mayo, K.E. (2009). Suppression of Notch signaling in the neonatal mouse ovary decreases primordial follicle formation. *Endocrinology* 150, 1014–1024.
- Tutar, Y. (2012). Pseudogenes. *Comp. Funct. Genomics* 2012.
- Uhlenhaut, N.H., Jakob, S., Anlag, K., Eisenberger, T., Sekido, R., Kress, J., Treier, A.-C., Klugmann, C., Klasen, C., Holter, N.I., et al. (2009). Somatic Sex Reprogramming of Adult Ovaries to Testes by FOXL2 Ablation. *Cell* 139, 1130–1142.
- Ulitsky, I., and Bartel, D.P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46.
- Umehara, F., Tate, G., Itoh, K., Yamaguchi, N., Douchi, T., Mitsuya, T., and Osame, M. (2000). A Novel Mutation of *desert hedgehog* in a Patient with 46,XY Partial Gonadal Dysgenesis Accompanied by Minifascicular Neuropathy. *Am. J. Hum. Genet.* 67, 1302–1305.
- Uszczynska-Ratajczak, B., Lagarde, J., Frankish, A., Guigó, R., and Johnson, R. (2018). Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.*
- del Valle, I., Buonocore, F., Duncan, A.J., Lin, L., Barenco, M., Parnaik, R., Shah, S., Hubank, M., Gerrelli, D., and Achermann, J.C. (2017). A genomic atlas of human adrenal and gonad development. *Wellcome Open Res.* 2, 25.
- Vandenberg, L.N., Colborn, T., Hayes, T.B., Heindel, J.J., Jacobs, D.R., Lee, D.-H., Shioda, T., Soto, A.M., vom Saal, F.S., Welshons, W. V., et al. (2012). Hormones and endocrine-disrupting chemicals: low-dose effects and nonmonotonic dose responses. *Endocr. Rev.* 33, 378–455.
- Vaskivuo, T.E., Mäentausta, M., Törn, S., Oduwole, O., Lönnberg, A., Herva, R., Isomaa, V., and Tapanainen, J.S. (2005). Estrogen receptors and estrogen-metabolizing enzymes in human ovaries during fetal development. *J. Clin. Endocrinol. Metab.* 90, 3752–3756.
- Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* 270, 484–487.

- Vetro, A., Dehghani, M.R., Kraoua, L., Giorda, R., Beri, S., Cardarelli, L., Merico, M., Manolagos, E., Parada-Bustamante, A., Castro, A., et al. (2015). Testis development in the absence of SRY: chromosomal rearrangements at SOX9 and SOX3. *Eur. J. Hum. Genet.* *23*, 1025–1032.
- Vidal, V.P.I., Chaboissier, M.-C., de Rooij, D.G., and Schedl, A. (2001). Sox9 induces testis development in XX transgenic mice. *Nat. Genet.* *28*, 216–217.
- Vigier, B., Watrin, F., Magre, S., Tran, D., and Josso, N. (1987). Purified bovine AMH induces a characteristic freemartin effect in fetal rat prospective ovaries exposed to it in vitro. *Dev. Camb. Engl.* *100*, 43–55.
- Virtanen, H.E., and Toppari, J. (2008). Epidemiology and pathogenesis of cryptorchidism. *Hum. Reprod. Update* *14*, 49–58.
- Virtanen, H.E., and Toppari, J. (2014). Embryology and physiology of testicular development and descent. *Pediatr. Endocrinol. Rev. PER* *11 Suppl 2*, 206–213.
- Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., et al. (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* *79*, 1111–1120.
- Wain, H.M., Bruford, E.A., Lovering, R.C., Lush, M.J., Wright, M.W., and Povey, S. (2002). Guidelines for human gene nomenclature. *Genomics* *79*, 464–470.
- Wainwright, E.N., and Wilhelm, D. (2010). The game plan: cellular and molecular mechanisms of mammalian testis development. *Curr. Top. Dev. Biol.* *90*, 231–262.
- Wallis, M.C., Waters, P.D., and Graves, J.A.M. (2008). Sex determination in mammals — Before and after the evolution of SRY. *Cell. Mol. Life Sci.* *65*, 3182–3195.
- Walschaerts, M., Muller, A., Auger, J., Bujan, L., Guérin, J.-F., Le Lannou, D., Clavert, A., Spira, A., Jouannet, P., and Thonneau, P. (2007). Environmental, occupational and familial risks for testicular cancer: a hospital-based case-control study. *Int. J. Androl.* *30*, 222–229.
- Wang, K.C., and Chang, H.Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* *43*, 904–914.
- Wang, Y., and Rekaya, R. (2009). A comprehensive analysis of gene expression evolution between humans and mice. *Evol. Bioinforma. Online* *5*, 81–90.
- Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., et al. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* *38*, e178.
- Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* *41*, e74.
- Wang, N., Zhang, P., Guo, X., Zhou, Z., and Sha, J. (2011). Hnrnpk, a protein differentially expressed in immature rat ovarian development, is required for normal primordial follicle assembly and development. *Endocrinology* *152*, 1024–1035.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* *10*, 57–63.
- Wanunu, M. (2012). Nanopores: A journey towards DNA sequencing. *Phys. Life Rev.* *9*, 125–158.
- Waring, R.H., and Harris, R.M. (2005). Endocrine disruptors: a human risk? *Mol. Cell. Endocrinol.* *244*, 2–9.
- Warr, N., Carre, G.-A., Siggers, P., Faleato, J.V., Brixey, R., Pope, M., Bogani, D., Childers, M., Wells, S., Scudamore, C.L., et al. (2012a). Gadd45 γ and Map3k4 Interactions Regulate Mouse Testis Determination via p38 MAPK-Mediated Control of Sry Expression. *Dev. Cell* *23*, 1020–1031.
- Warr, N., Carre, G.-A., Siggers, P., Faleato, J.V., Brixey, R., Pope, M., Bogani, D., Childers, M., Wells, S., Scudamore, C.L., et al. (2012b). Gadd45 γ and Map3k4 Interactions Regulate Mouse Testis Determination via p38 MAPK-Mediated Control of Sry Expression. *Dev. Cell* *23*, 1020–1031.

- Warren, D.W., Dufau, M.L., and Catt, K.J. (1982). Hormonal regulation of gonadotropin receptors and steroidogenesis in cultured fetal rat tests. *Science* 218, 375–377.
- Washietl, S., Kellis, M., and Garber, M. (2014). Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* 24, 616–628.
- Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* 20, 1732–1743.
- Watson, J.D., and Crick, F.H.C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737–738.
- Welsh, M., Saunders, P.T.K., Fiskens, M., Scott, H.M., Hutchison, G.R., Smith, L.B., and Sharpe, R.M. (2008). Identification in rats of a programming window for reproductive tract masculinization, disruption of which leads to hypospadias and cryptorchidism. *J. Clin. Invest.* 118, 1479–1490.
- Wen, Q., Cheng, C.Y.Y., and Liu, Y.-X.X. (2016). Development, function and fate of fetal Leydig cells.
- Werner, R., Merz, H., Birnbaum, W., Marshall, L., Schröder, T., Reiz, B., Kavran, J.M., Bäumer, T., Capetian, P., and Hiort, O. (2015). 46,XY Gonadal Dysgenesis due to a Homozygous Mutation in Desert Hedgehog (DHH) Identified by Exome Sequencing. *J. Clin. Endocrinol. Metab.* 100, E1022-9.
- Wilhelm, D. (2002). The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1. *Genes Dev.* 16, 1839–1851.
- Wilhelm, D., Martinson, F., Bradford, S., Wilson, M.J., Combes, A.N., Beverdam, A., Bowles, J., Mizusaki, H., and Koopman, P. (2005). Sertoli cell differentiation is induced both cell-autonomously and through prostaglandin signaling during mammalian sex determination. *Dev. Biol.* 287, 111–124.
- Wilhelm, D., Hiramatsu, R., Mizusaki, H., Widjaja, L., Combes, A.N., Kanai, Y., and Koopman, P. (2007). SOX9 Regulates Prostaglandin D Synthase Gene Transcription *in Vivo* to Ensure Testis Development. *J. Biol. Chem.* 282, 10553–10560.
- Wilson, C.M., and McPhaul, M.J. (1996). A and B forms of the androgen receptor are expressed in a variety of human tissues. *Mol. Cell. Endocrinol.* 120, 51–57.
- Witchi, E. (1948). Migration of the germ cells of human embryos from the yolk sac to the primitive gonadal folds. *Contrib Embryol Carnegie Inst* 32, 67–80.
- Wu, X., and Watson, M. (2009). CORNA: testing gene lists for regulation by microRNAs. *Bioinforma. Oxf. Engl.* 25, 832–833.
- Wu, Z., and Wu, H. (2016). *Experimental Design and Power Calculation for RNA-seq Experiments*. (Humana Press, New York, NY), pp. 379–390.
- Wu, F.C.W., Tajar, A., Beynon, J.M., Pye, S.R., Silman, A.J., Finn, J.D., O’Neill, T.W., Bartfai, G., Casanueva, F.F., Forti, G., et al. (2010). Identification of late-onset hypogonadism in middle-aged and elderly men. *N. Engl. J. Med.* 363, 123–135.
- Wu, H., Yang, L., and Chen, L.-L. (2017). The Diversity of Long Noncoding RNAs and Their Generation. *Trends Genet.* 33, 540–552.
- Wurster, C.F. (2015). DDT wars : rescuing our national bird, preventing cancer, and creating The Environmental Defense Fund.
- Wylie, C. (1999). Germ cells. *Cell* 96, 165–174.
- Xie, L., Weichel, B., Ohm, J.E., and Zhang, K. (2011). An integrative analysis of DNA methylation and RNA-Seq data for human heart, kidney and liver. *BMC Syst. Biol.* 5 Suppl 3, S4.
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., et al. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinforma. Oxf. Engl.* 30, 1660–1666.

- Xu, J., and Gridley, T. (2013). Notch2 is required in somatic cells for breakdown of ovarian germ-cell nests and formation of primordial follicles. *BMC Biol.* *11*, 13.
- Xu, J., and Zhang, J. (2016). Are Human Translated Pseudogenes Functional? *Mol. Biol. Evol.* *33*, 755–760.
- Yadav, R.P., and Kotaja, N. (2014). Small RNAs in spermatogenesis. *Mol. Cell. Endocrinol.* *382*, 498–508.
- Yang, L., Froberg, J.E., and Lee, J.T. (2014). Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem. Sci.* *39*, 35–43.
- Yao, H.H.-C., Whoriskey, W., and Capel, B. (2002). Desert Hedgehog/Patched 1 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes Dev.* *16*, 1433–1440.
- Yao, H.H.C., Matzuk, M.M., Jorgez, C.J., Menke, D.B., Page, D.C., Swain, A., and Capel, B. (2004a). Follistatin operates downstream of Wnt4 in mammalian ovary organogenesis. *Dev. Dyn.* *230*, 210–215.
- Yao, H.H.C., Matzuk, M.M., Jorgez, C.J., Menke, D.B., Page, D.C., Swain, A., and Capel, B. (2004b). Follistatin operates downstream of Wnt4 in mammalian ovary organogenesis. *Dev. Dyn.* *230*, 210–215.
- Yao, H.H.-C., Aardema, J., and Holthusen, K. (2006). Sexually dimorphic regulation of inhibin beta B in establishing gonadal vasculature in mice. *Biol. Reprod.* *74*, 978–983.
- Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. *Nucleic Acids Res.* *44*, D710–D716.
- Yazaki, J., Gregory, B.D., and Ecker, J.R. (2007). Mapping the genome landscape using tiling array technology. *Curr. Opin. Plant Biol.* *10*, 534–542.
- Yoshino, T., and Saito, D. (2018). Epithelial-to-mesenchymal transition–based morphogenesis of dorsal mesentery and gonad. *Semin. Cell Dev. Biol.*
- Yu, R.N., Ito, M., Saunders, T.L., Camper, S.A., and Jameson, J.L. (1998). Role of Ahch in gonadal development and gametogenesis. *Nat. Genet.* *20*, 353–357.
- Yusuf, D., Butland, S.L., Swanson, M.I., Bolotin, E., Ticoll, A., Cheung, W.A., Zhang, X.Y.C., Dickman, C.T.D., Fulton, D.L., Lim, J.S., et al. (2012). The transcription factor encyclopedia. *Genome Biol.* *13*, R24.
- Zanaria, E., Muscatelli, F., Bardoni, B., Strom, T.M., Guioli, S., Guo, W., Lalli, E., Moser, C., Walker, A.P., McCabe, E.R.B., et al. (1994). An unusual member of the nuclear hormone receptor superfamily responsible for X-linked adrenal hypoplasia congenita. *Nature* *372*, 635–641.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* *46*, D754–D761.
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* *4*, Article17.
- Zhang, H., Risal, S., Gorre, N., Busayavalasa, K., Li, X., Shen, Y., Bosbach, B., Brännström, M., and Liu, K. (2014). Somatic Cells Initiate Primordial Follicle Activation and Govern the Development of Dormant Oocytes in Mice. *Curr. Biol.* *24*, 2501–2508.
- Zhang, H.-Q., Zhang, X.-F., Zhang, L.-J., Chao, H.-H., Pan, B., Feng, Y.-M., Li, L., Sun, X.-F., and Shen, W. (2012). Fetal exposure to bisphenol A affects the primordial follicle formation by inhibiting the meiotic progression of oocytes. *Mol. Biol. Rep.* *39*, 5651–5657.
- Zhang, Y., Zhong, L., Xu, B., Yang, Y., Ban, R., Zhu, J., Cooke, H.J., Hao, Q., and Shi, Q. (2013). SpermatogenesisOnline 1.0: a resource for spermatogenesis based on manual literature curation and genome-wide data mining. *Nucleic Acids Res.* *41*, D1055–D1062.
- Zhao, S., and Zhang, B. (2015). A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genomics* *16*, 97.

Zhao, F., Franco, H.L., Rodriguez, K.F., Brown, P.R., Tsai, M.-J., Tsai, S.Y., and Yao, H.H.-C. (2017). Elimination of the male reproductive tract in the female embryo is promoted by COUP-TFII in mice. *Science* 357, 717–720.

Zhao, L., Wang, C., Lehman, M.L., He, M., An, J., Svingen, T., Spiller, C.M., Ng, E.T., Nelson, C.C., and Koopman, P. (2018). Transcriptomic analysis of mRNA expression and alternative splicing during mouse sex determination. *Mol. Cell. Endocrinol.*

Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A.M., and Mazutis, L. (2016). Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* 12, 44–73.

Zirkin, B.R., and Chen, H. (2000). Regulation of Leydig cell steroidogenic function during aging. *Biol. Reprod.* 63, 977–981.

(1965). Nobel Prizes, 1965. *Br. Med. J.* 2, 958.

Whole genome - Homo sapiens - Ensembl genome browser 93.

Résumé

Les organes centraux du tractus urogénital sont le testicule et l'ovaire, qui assurent la production de gamètes et d'hormones, et donc la fertilité de l'individu. Ces deux organes, parfaitement distincts et complémentaires, ont pour origine une gonade bipotentielle qui s'engagera vers une trajectoire de différenciation masculine ou féminine au cours de la vie fœtale. Les deux gonades vont par la suite subir plusieurs phases de différenciation et de développement de leurs populations cellulaires, afin d'acquérir leurs fonctions propres qui leur permettront d'assumer leur rôle à l'âge adulte. Depuis plus d'une quinzaine d'années, le concept de syndrome de dysgénésie testiculaire fait état d'un lien entre l'exposition du fœtus à des composés environnementaux et des anomalies du tractus urogénital. Bien que sujette à de vifs débats au sein de la communauté scientifique, cette hypothèse a attiré l'attention de la recherche sur les conséquences de l'exposition des mères aux xénobiotiques sur l'enfant à naître.

La différenciation et le développement des gonades fœtales sont gouvernés par des programmes d'expression spécifiques de chaque sexe, dont de nombreuses zones d'ombres subsistent, notamment concernant la fraction non-codante exprimée par le génome humain. La première partie de cette thèse de doctorat présente, pour la première fois, le paysage transcriptionnel contrôlant ces processus complexes entre la 6^{ième} et 17^{ième} semaine de développement chez l'Homme. Grâce à l'avènement des technologies de transcriptomique, il est désormais possible d'identifier et d'observer l'expression des gènes de manière sensible et sans *a priori*. Le RNA-seq m'a donc permis de décrire de manière exhaustive la dynamique d'expression des gènes, pendant les stades précoces de la différenciation sexuelle, jusqu'aux phénomènes plus tardifs conduisant aux lignées des différentes populations cellulaires spécifiques du testicule et de l'ovaire.

Dans une deuxième partie, mon travail de recherche s'est attaché à étudier l'impact de deux perturbateurs endocriniens suspectés, l'ibuprofène et le chlordécone, sur le programme d'expression du testicule fœtal humain. L'utilisation du RNA-seq m'a permis de définir et de comparer la signature toxicogénomique de chaque molécule afin de contribuer à la compréhension de leur mécanisme d'action et d'identifier les populations cellulaires affectées.

Enfin, face à l'essor des technologies ultra-haut-débit dans les sciences de la vie, y compris dans les domaines de la reproduction, j'ai activement participé au déploiement d'une nouvelle version du Reprogenomic Viewer dans la dernière partie de ma thèse (<http://rgv.genouest.org>). Cet outil internet a pour vocation de centraliser et de rendre accessibles les données de séquençage accumulées au sein de la communauté de la reproduction via des outils de visualisation intuitifs

Titre : Analyse par RNA-seq de la différenciation des gonades fœtales humaines et de son altération par des perturbateurs endocriniens.

Mots clés : Reproduction, Testis, Ovary, Transcriptomics, Endocrine disruptors

Résumé :

La différenciation et le développement des gonades fœtales sont gouvernés par des programmes d'expression spécifiques de chaque sexe, dont de nombreuses zones d'ombres subsistent, notamment concernant la fraction non-codante exprimée par le génome humain. La première partie de cette thèse de doctorat présente, pour la première fois, le paysage transcriptionnel contrôlant ces processus complexes entre la 6^{ème} et 17^{ème} semaine de développement chez l'Homme. Grâce à l'avènement des technologies de transcriptomique, il est désormais possible d'identifier et d'observer l'expression des gènes de manière sensible et sans a priori. Le RNA-seq m'a donc permis de décrire de manière exhaustive la dynamique d'expression des gènes, pendant les stades précoces de la différenciation sexuelle, jusqu'aux phénomènes plus tardifs conduisant aux lignées des différentes populations cellulaires spécifiques du testicule et de l'ovaire.

Dans une deuxième partie, mon travail de recherche s'est attaché à étudier l'impact de deux perturbateurs endocriniens suspectés, l'ibuprofène et le chlordécone, sur le programme d'expression du testicule fœtal humain. L'utilisation du RNA-seq m'a permis de définir et de comparer la signature toxicogénomique de chaque molécule afin de contribuer à la compréhension de leur mécanisme d'action et d'identifier les populations cellulaires affectées. Enfin, face à l'essor des technologies ultra-haut-débit dans les sciences de la vie, y compris dans les domaines de la reproduction, j'ai activement participé au déploiement d'une nouvelle version du Reprogenomic Viewer, un outil internet ayant pour vocation de centraliser et de rendre accessibles les données de séquençage générées par la communauté de la reproduction via des outils de visualisation intuitifs.

Title : Dynamics of the transcriptional landscape during human fetal gonad differentiation and development and its alteration by endocrine disruptors.

Keywords : Reproduction, Testis, Ovary, Transcriptomics, Endocrine disruptors

Abstract : Fetal life is a crucial period for sexual reproduction when bipotential gonads differentiate into either a testis or an ovary. Gaining insights into the complex molecular events underlying this process is central to a better understanding of disorders of sexual development. The present work intends to improve the knowledge on molecular pathways at play during gonad development in humans using RNA-sequencing. This project particularly seeks to identify early transcriptional events that may play critical role in the regulatory network driving human sexual differentiation.

To address this issue, we defined the transcriptional landscape of fetal human gonads by sequencing total RNA extracted from testes and ovaries between 6 and 17 gestational weeks.

The dynamic of transcriptional landscape occurring during human fetal gonads development has been described and new genes and interesting candidates, including new genes, have been highlighted as potential key genes governing this biological process.

The second interest of this work was the study of the impact of two endocrine disruptors, ibuprofene and chlordecone, on human fetal testis using RNA-seq. The transcriptional alteration induced by these compound in the gonad allowed a deeper understanding of their mechanisms of action of endocrine disruption.