

## Université de Limoges

### ED 615 - Sciences Biologiques et Santé (SBS)

CNRS UMR 7276 et INSERM U1262 - Laboratoire CRIBL : Contrôle de la Réponse Immune B et des Lymphoproliférations

Thèse pour obtenir le grade de

Docteur de l'Université de Limoges

Discipline : Biologie Sciences Santé - Spécialité : Immunogénétique

Présentée et soutenue par

**Ophélie Alyssa Martin**

Le 3 octobre 2018

**Hypermutation somatique dans les cellules B normales et pathologiques :  
éléments *cis*-régulateurs et facteurs nucléaires impliqués.**

Thèse dirigée par Dr. Éric Pinaud

#### **JURY :**

##### **Président du jury**

Dr. Salvatore Spicuglia, Directeur de Recherche INSERM, INSERM U1090 Marseille

##### **Rapporteurs**

Dr. Said Aoufouchi, Directeur de Recherche Inserm, CNRS UMR 8200 Villejuif

Dr. Salvatore Spicuglia, Directeur de Recherche INSERM, INSERM U1090 Marseille

##### **Examineurs**

Dr Céline Delaloy, Chargée de Recherche CNRS, INSERM U1236 Rennes

Dr Thierry Forné, Directeur de Recherche CNRS, CNRS UMR 5535 Montpellier

##### **Invités**

Dr. Eric Pinaud, Directeur de Recherche CNRS, CNRS UMR 7276 INSERM U1262





**« L'essentiel est invisible pour les yeux. »**

*Antoine de Saint-Exupéry*



## Remerciements.

---

*Je remercie sincèrement Éric pour ces 3 années d'encadrement (et plus puisque j'ai pris un peu d'avance avant la thèse). Merci pour ta disponibilité, ta bonne humeur et ces quelques phrase cultes « l'ADN c'est comme des spaghettis », « la CHIP c'est comme la forêt » et « pendant mon absence Stevy sera le boss » (avec la bonne intonation pour cette dernière phrase). Merci pour tous tes encouragements et tes conseils, j'ai été ravie de travailler avec toi, ces 3 années de thèse sont passées très vite !*

*Un grand merci à Sandrine, également pour ton encadrement, tes encouragements et toutes nos discussions scientifiques et non scientifiques ! Ta présence durant ma thèse m'a beaucoup apporté. Merci d'avoir partagé tes protocoles et leurs nombreuses « petites subtilités » et de m'avoir initiée à la technique de FISH. Après des heures et des heures (et même quelques repas) passées au trieur, au microscope et à mesurer des distances je ne peux plus me passer des noyaux !! Bon courage pour la suite du projet « coloc », pour les souris, pour les mutations (la version okokok de chez ok numéro 277 est la bonne) et pour faire régner l'ordre dans la pièce de migration !! J'ai enfin réussi à finir un mot fléché seule (ou presque) je suis donc logiquement prête pour la soutenance (en tout cas je l'espère) !*

*Je voudrais également remercier le Pr. Michel Cogné de m'avoir accueilli dans son laboratoire.*

*Je remercie tous les membres du jury de me faire l'honneur de juger mon travail. Merci au Dr Salvatore Spicuglia de présider ce jury et d'avoir accepté d'être rapporteur de mon travail. Merci au Dr Said Aoufouchi d'avoir également accepté d'être rapporteur de mon travail. Merci au Dr Céline Delaloy et au Dr Thierry Forné d'avoir accepté d'examiner mon travail.*

*Merci aussi à l'équipe pédagogique de la faculté des Sciences et plus particulièrement à Fabrice Dupuy qui a joué le rôle de tuteur pédagogique pendant mon service d'enseignement.*

*Je souhaite bien sûr aussi remercier sincèrement toute l'équipe du laboratoire CRIBL. Merci à Armand de m'avoir passé le flambeau et pour tous tes conseils ! Merci à Mylène ex-membre du groupe d'Éric et spécialiste en « iGC », n'oublie pas Mylène la qPCR multiplex est ton amie ! Merci à Claire, pour tes interventions au microscope (« allo Claire j'ai un problème », j'espère que le nikon va bien se tenir) ! Merci à Cendrine, on a formé une belle équipe pour le nettoyage de la pièce de migration, je suis sûre que notre idée d'offrir des images aux bons binômes de nettoyage tient la route (ps : j'espère que ton passeport est fin prêt). Merci à Zéliha pour ton savoir-faire avec les souris. Merci à Christelle O. pour ton aide précieuse lors des manips de western ! Merci à Stéphanie, pour tes précieux conseils en matière de CV et de lettres de motivation. C'est aussi grâce à toi (et à Sandrine) que j'ai progressé aux mots fléchés (j'arrive même à trouver des mots de plus de 2 lettres dans les forces 4) !! Ah oui je sais aussi maintenant à quoi ressemblent les clématites (plus d'erreur possible) ! Merci à François pour toutes nos conversations sur le comment du pourquoi pour détecter les mutations rares...et vive DeMinEr ! Merci à Brice pour tes mots d'encouragement et de m'avoir fait confiance pour la préparation à la sauce « Limoges » des échantillons à séquencer !*

*Merci à toutes les autres personnes de l'unité, chacun contribuant à la bonne ambiance du laboratoire : Soazig, Alexis, Julie, Tiffany, et toutes les teams Christophe, Laurent, Jean-Claude, Sophie, Jeanne, Yves, Anne. Sophie (et Sandrine) j'ai bien pris notes de vos conseils en matière de vous savez quoi (persils et compagnie) ! Merci*



*de m'avoir « préparé » pour mon entretien ! Merci Anne de m'avoir donné l'occasion de participer aux ateliers immunoludiques, j'ai adoré !*

*Merci également au personnel de l'animalerie qui prend soin de nos amis les souris.*

*Je souhaite une bonne continuation à tous ceux qui viennent de soutenir (chanceux !) et ceux qui ne vont pas tarder à soutenir... Omar, Anderson, Hend, Victoria, Sébastien, Jean-Marie, Anne, Audrey, Batoul, Iman, Zeinab, Nour, Hussein, Lina !!*

*Merci Audrey pour les debriefs autour d'un thé et ne t'en fais pas : tant qu'il y a de la bande y'a de l'espoir !*

*Merci également à tous les membres de l'équipe du Pr. Jean Feuillard pour votre sympathie.*

*J'espère n'avoir oublié personne et je souhaite à chacun une bonne continuation et la concrétisation de vos projets ! C'est avec un réel plaisir que je viendrai vous rendre visite lors de mes passages à Limoges !*

*Je pense également à tous ceux qui répondent présents lors des « réunions des doctorants » autour d'un verre... Sabrina, Hugo et Martial (responsables du pôle évènementiel), Axel (merci pour ce petit tour dans ta « splach »), Alexis (c'était le temps des fleurs...), Robin, Hélène... c'est toujours un plaisir de vous retrouver et de partager des bons moments avec vous ! La saison des soutenances a démarré !!*

*Sabrina j'ai juste envie de dire « donne moi un P, un O... » la belle époque (ou pas) !! Merci à toi et à Benoit pour votre amitié précieuse ! Il faut qu'on reprogramme un voyage au « anta-é » ou alors en mode peignoirs – champagne sur notre balcon avec vue directe sur Disney Illumination !!*

*Enfin, je remercie bien évidemment ma famille pour leur amour et leur soutien inconditionnel.*

*Maman merci pour tout ce que tu fais pour moi. Tes encouragements, ta présence et ta tendresse m'ont porté jusqu'ici. Ta force à toutes épreuves est une source de motivation.*

*Papa et Josette, vous allez pouvoir arrêter de stresser pour moi : la thèse c'est fait ! Merci pour votre présence, votre affection et de croire aussi fort en moi.*

*Et Béren ma sœur, sur qui je peux toujours compter et qui a plus confiance en moi que moi ! Merci pour notre belle complicité et pour ce lien si fort qui nous unit.*

*Je remercie aussi Jean-Luc, Gaëlle, Théo et Maëva d'être présents ainsi que pour tous ces bons moments passés ensembles.*

*Merci à toutes les taties et les tontons qui suivent mes aventures : tatie Claude, tatie Anne Marie, tonton Carlos, tatie Bernadette et tatie Paule... et à mes nombreux et surtout nombreuses cousines ! Merci également à mamie Raymonde de nous avoir offert une si belle et grande famille !*

*J'ai également une pensée émue pour tous ceux qui ne sont plus avec nous, qui comptent énormément pour moi et avec qui j'aurai tant aimé partager ce moment : papi Charlou, mamie Marcelle et papi Claude.*





*Un merci tout particulier à tatie Mireille et à tonton Daniel, vous avez été une deuxième maman et un deuxième papa et m'avez toujours soutenue. Merci d'avoir toujours répondu présents.*

*Je n'oublie pas non plus Christelle et Yannick, Chantale et Claude, Cédric et Alexandra, Domie et Alain. Vous faites aussi partie de ma famille, je pense très souvent à tous les bons moments que nous avons partagés ensemble.*

*Un grand merci également à ma belle-famille, Stéphane, Christine, Alexis, Louis et Laura. Merci pour votre accueil chaleureux, votre générosité et votre bienveillance. Je suis heureuse de faire partie de votre famille.*

*Enfin merci à Justin, ma moitié, pour tout l'amour et le réconfort que tu sais m'apporter. Merci de m'avoir soutenue durant cette période de rédaction (quelque peu stressante) et de toujours m'encourager. Je vous souhaite une vie heureuse et pleine d'aventures !*

*Ces quelques lignes ne seront jamais à la hauteur de ce que vous m'avez apporté, votre soutien a été précieux à chaque moment.*



## Droits d'auteurs.

---

Cette création est mise à disposition selon le Contrat :

« **Attribution-Pas d'Utilisation Commerciale-Pas de modification 3.0 France** »

disponible en ligne : <http://creativecommons.org/licenses/by-nc-nd/3.0/fr/>





## Table des matières.

---

<b>Remerciements.</b> .....	3
<b>Droits d'auteurs.</b> .....	6
<b>Table des matières.</b> .....	7
<b>Table des illustrations.</b> .....	9
<b>Liste des abréviations couramment utilisés.</b> .....	10
<b>Synthèse bibliographique.</b> .....	12
<b>Introduction.</b> .....	13
<b>Chapitre I. Organisation des gènes des immunoglobulines.</b> .....	15
I.1. Les loci des chaînes légères des immunoglobulines. ....	15
I.1.1. Le locus kappa ( <i>Igκ</i> ).....	15
I.1.2. Le locus lambda ( <i>Igλ</i> ).....	15
I.2. Le locus des chaînes lourdes des immunoglobulines ( <i>IgH</i> ).....	15
<b>Chapitre II. Développement des lymphocytes B.</b> .....	17
II.1. Phase précoce du développement B et recombinaisons V(D)J.....	17
II.1.1. Stades précoces de développement B : statut des gènes d'immunoglobulines.....	17
II.1.2. Stades précoces de développement B : facteurs essentiels à la progression dans le lignage B. .....	20
II.1.3. Régulation de l'accessibilité au locus <i>IgH</i> lors des recombinaisons VDJ. ....	21
II.1.3.1. La transcription germinale.....	21
II.1.3.2. Les modifications de la chromatine.....	22
II.2. Phase tardive du développement B et recombinaisons secondaires. ....	22
II.2.1. Le stade B transitionnel. ....	23
II.2.2. Signalisation induite par le BCR et devenir des cellules B. ....	24
II.2.3. Le stade B mature.....	24
II.2.3.1. Les cellules B1.....	25
II.2.3.2. Les cellules B2.....	25
II.2.3.3. La formation du centre germinatif.....	25
II.2.3.4. Facteurs essentiels à la régulation du centre germinatif.....	27
II.2.4. Les cellules B mémoires.....	28
II.2.5. Les plasmocytes : cellules B sécrétrices d'anticorps. ....	28
<b>Chapitre III. AID : enzyme clé du développement B tardif.</b> .....	30
III.1. AID : structure, fonction et régulation. ....	30
III.1.1. Structure de AID. ....	30
III.1.2. Fonction de AID. ....	30
III.1.3. Régulation de l'activité de AID. ....	31
III.2. Recrutement de AID aux loci des gènes d'immunoglobulines.....	33
III.2.1. Ciblage de AID lors de la SHM et mécanisme moléculaire de la SHM. ....	35
III.2.1.1. Ciblage de AID lors de la SHM.....	35
III.2.1.2. Mécanisme moléculaire de la SHM.....	36
III.2.2. Ciblage de AID lors de la CSR et mécanisme moléculaire de la CSR.....	39



III.2.2.1 Ciblage de AID lors de la CSR.....	39
III.2.2.2. Mécanisme moléculaire de la CSR.....	41
III.3. Cibles illégitimes de AID. ....	42
<b>Chapitre IV. Les régions régulatrices au locus <i>IgH</i>.....</b>	<b>47</b>
IV.1. Les promoteurs au locus <i>IgH</i> .....	47
IV.1.1. Les promoteurs pV <sub>H</sub> .....	47
IV.1.2. Le promoteur pDQ52. ....	48
IV.1.3. Les promoteurs I. ....	48
IV.2. Les activateurs transcriptionnels ou « enhancers » au locus <i>IgH</i> . ....	48
IV.2.1. La région intronique E $\mu$ .....	48
IV.2.1.1. L'activateur intronique core E $\mu$ (cE $\mu$ ).....	49
IV.2.1.2. Les régions d'attachements à la matrice nucléaire (MAR <sub>E<math>\mu</math></sub> ).....	51
IV.2.2. La région régulatrice en 3' du locus <i>IgH</i> (3'RR). ....	55
IV.2.2.1. Structure de la région 3'RR.....	55
IV.2.2.2. Fonction de la région 3'RR.....	56
IV.2.2.2.1. Modèles de souris invalidant hs1-2 ou hs3a par insertion du gène neo <sup>r</sup> ou délétion précise.....	57
IV.2.2.2.2. Modèles de souris déficientes pour hs3b.....	57
IV.2.2.2.3. Modèles de souris déficientes pour hs4.....	57
IV.2.2.2.4. Modèle de souris déficientes pour plusieurs éléments composants la région 3'RR.....	58
IV.3. Les régions isolatrices au locus <i>IgH</i> . ....	59
IV.3.1. La région en 5' du locus <i>IgH</i> . ....	59
IV.3.2. La région intergénique V <sub>H</sub> -D.....	60
IV.3.3. La région en 3' du locus <i>IgH</i> . ....	60
IV.4. Régulation spatiale du locus <i>IgH</i> . ....	60
IV.4.1. Position des loci des gènes d'Ig.....	61
IV.4.2. Topologie tridimensionnelle du locus <i>IgH</i> .....	63
<b>Résultats.....</b>	<b>64</b>
<b>Article 1</b> : Detecting rare AID-induced mutations in B lineage oncogenes from high throughput sequencing data using DeMinEr method.....	<b>68</b>
<b>Article 2</b> : Antisens transcripts driven by E $\mu$ matrix attachment regions optimize somatic hypermutation upstream from the enhancer.....	<b>70</b>
<b>Article 3</b> : Evidence for a <i>trans</i> -dominant impact of <i>IgH</i> E $\mu$ matrix attachment regions on somatic hypermutation and subnuclear positioning of both Ig genes and AID off targets.....	<b>71</b>
<b>Discussion et perspectives.....</b>	<b>73</b>
<b>Annexes. ....</b>	<b>82</b>
<b>Article 1</b> : Physiological and druggable skipping of immunoglobulin variable exons in plasma cells.....	<b>83</b>
<b>Article 2</b> : Radiolabelled polymeric IgA: from biodistribution to a new molecular imaging tool in colorectal cancer lung metastases. ....	<b>84</b>
<b>Article 3</b> : Nuclear proximity of Mtr4 to RNA exosome restricts DNA mutational asymmetry.....	<b>85</b>
<b>Références bibliographiques.....</b>	<b>86</b>





## Table des illustrations.

---

### Synthèse bibliographique :

Figure 1 : Structure générale d'une immunoglobuline, exemple d'une IgG.

Figure 2 : Représentation schématique des loci des gènes d'immunoglobulines.

Figure 3 : Evènements de remaniements s'opérant aux loci des chaînes lourdes (*IgH*) et légères (*IgL*) d'Ig.

Figure 4 : Représentation schématique des différents stades associés au développement B précoce.

Figure 5 : Règle dite « 12/23 » et recombinaisons VDJ au locus *IgH*.

Figure 6 : Transcrits germinaux sens et antisens initiés au locus *IgH*.

Figure 7 : Représentation schématique des différents stades associés au développement B tardif.

Figure 8 : Localisation et devenir des cellules B transitionnelles de la rate.

Figure 9 : Représentation schématique de la structure d'un centre germinatif.

Figure 10 : Structure et activité de l'enzyme AID.

Figure 11 : Caractéristiques générales du processus de SHM.

Figure 12 : Ciblage de AID vers les régions d'intérêt lors du processus de SHM.

Figure 13 : Devenir du mésappariement initié par AID lors du processus de SHM.

Figure 14 : Schématisation du processus de CSR.

Figure 15 : Ciblage de AID vers les régions d'intérêt lors du processus de CSR.

Figure 16 : Voies de réparation des cassures de l'ADN induites lors de la CSR.

Figure 17 : Les cibles illégitimes de AID sont réparties en 3 groupes.

Figure 18 : Intervention de AID sur les cibles « off targets ».

Figure 19 : Modèles murins disponibles au laboratoire pour étudier les rôles de la région  $E\mu$  et 3'RR.

Figure 20 : Représentation schématique de la région  $E\mu$  et des sites de fixation des facteurs *trans*-régulateurs.

Figure 21 : Schématisation des 2 types de régions MARs.

Figure 22 : Représentation schématique du positionnement des loci des gènes d'Ig au cours du développement B.

Figure 23 : Exemples de topologie adoptée par le locus *IgH*.

Table 1 : Principales différences entre les processus de SHM et de CSR.

Table 2 : Modèles murins de délétion de la région  $E\mu$  et conséquences sur les évènements de recombinaisons VDJ, d'hypermutation somatique (SHM) et de recombinaison de classe (CSR).

Table 3 : Modèles murins de délétion de la région 3'RR et conséquences sur les évènements de recombinaisons VDJ, d'hypermutation somatique (SHM) et de recombinaison de classe (CSR).

### Discussion et perspectives :

Figure 24 : Hypothèse d'un mode d'action en *trans* pour les ARN non codants.

Figure 25 : Données de RNA-seq et MAR-BP.

Figure 26 : Stratégie de croisement pour l'obtention des modèles de souris KO conditionnels pour ARID3A et SATB1.



## Liste des abréviations couramment utilisées.

<p><b>A</b></p> <p>A : adénine</p> <p>Ac : anticorps</p> <p>AEJ : alternative end joining</p> <p>AID : activation induced cytidine deaminase</p> <p>ADN : acide désoxyribonucléique</p> <p>Ag : antigène</p> <p>APOBEC : apolipoprotein B mRNA editing catalytic</p> <p>ARID : A-T rich interaction domain</p> <p>ARN : acide ribonucléique</p> <p>ARN pol : ARN polymérase</p>	<p>ChIP : chromatine immunoprecipitation</p> <p>CLP : common lymphoid progenitors</p> <p>CMH : complexe majeur d'histocompatibilité</p> <p>C-NHEJ (ou NHEJ) : classical non homologous end joining</p> <p>CSH : cellules souches hématopoïétique</p> <p>CSR : class switch recombination</p> <p>CTCF : CCCTC-binding factor</p>	<p>Ig<math>\kappa</math> : locus kappa de chaînes légères d'immunoglobulines</p> <p>Ig<math>\lambda</math> : locus lambda de chaînes légères d'immunoglobulines</p> <p><i>IgH</i> : locus des chaînes lourdes d'immunoglobulines</p> <p><i>IgL</i> : locus des chaînes légères d'immunoglobulines</p> <p>IL : interleukine</p> <p>IRIS : inverted repeat intervening sequence</p>
<p><b>B</b></p> <p>BCL : B cell lymphoma</p> <p>BCR : B Cell Receptor</p> <p>BER : base Excision Repair</p> <p>BRIGHT : B cell regulator of IgH transcription</p>	<p><b>D</b></p> <p>dC : deoxycytidine</p> <p>D : segment de diversité du locus <i>IgH</i></p> <p>DNase I : désoxyribonucléase I</p> <p>dU : déoxyuracile</p>	<p><b>J</b></p> <p>J<sub>H</sub> : segment de jonction au locus <i>IgH</i></p> <p>J<sub>L</sub> : segment de jonction au locus <i>IgL</i></p>
<p><b>C</b></p> <p>C : cytosine</p> <p>CBE : CTCT binding element</p> <p>CDF : cellules dendritique folliculaire</p> <p>CDR : complementary determining region</p> <p>cE<math>\mu</math> : core enhance E<math>\mu</math></p> <p>CG : centre germinatif</p> <p>C<sub>H</sub> : segment constant du locus <i>IgH</i></p>	<p><b>F</b></p> <p>FISH 3D : Fluorescent <i>in situ</i> hybridization and 3 dimensions</p> <p>FO : folliculaire</p>	<p><b>K</b></p> <p>kb : kilobase</p> <p>KD : knock down</p> <p>KI : knock in</p> <p>KO : knock out</p>
<p><b>G</b></p> <p>GFP : green fluorescent protein</p> <p>G : guanine</p>	<p><b>H</b></p> <p>hs : site hypersensible à la DNase I</p>	<p><b>L</b></p> <p>LB : lymphocyte B</p> <p>LT : lymphocyte</p> <p>LPS : lipopolysaccharide bactérien</p>
<p><b>I</b></p> <p>I : intervening</p> <p>Ig : immunoglobuline</p>	<p><b>M</b></p> <p>MARs : matrix attachment region, et MAR<sub>E<math>\mu</math></sub> : matrix attachment region encadrant l'activateur cE<math>\mu</math> au locus <i>IgH</i></p> <p>MAR-BP : MAR – binding proteins</p> <p>MMR : MisMatch Repair</p>	



MPP : multipotent progenitor

MZ : zone marginale

### N

neo<sup>r</sup> : gène de résistance à la néomycine

NES : nuclear export signal

NGS : next generation sequencing

NLS : nuclear location signal

pb : paire de bases

### P

pV<sub>H</sub> : promoteur des régions variables des chaînes lourdes d'immunoglobulines

### R

Rag : recombination activating gene

3'RR : 3' Regulatory Region

RSS : recombination signal sequence

RPA : replicative protein A

### S

S : région switch

SATB : special AT binding

SHM : Somatic Hypermutation

SI : système immunitaire

### T

T1/T2/T3 : cellules B transitionnelles

TFH : cellules T follicular helper

### U

U : uracile

UNG : Uracyle N- glycosylase

### V

V : segments de variabilité

### W

Wt : wild type

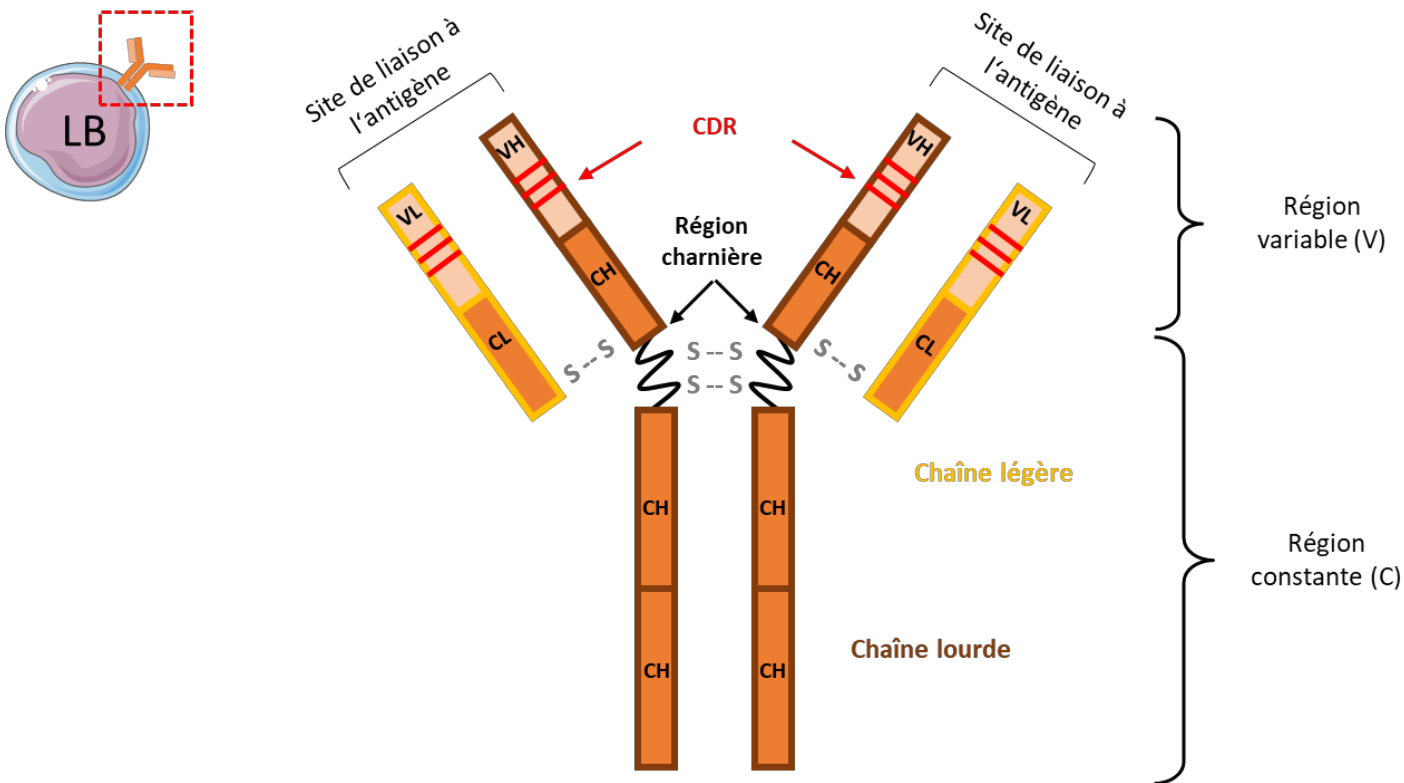
### Z

ZC : zone claire

ZS : zone sombre



## *Synthèse bibliographique*



**Figure 1 : Structure générale d'une immunoglobuline, exemple d'une IgG.**

Une immunoglobuline est formée par la réunion de 4 chaînes polypeptidiques : 2 chaînes lourdes (encadrées en marron) et 2 chaînes légères (encadrées en jaune) reliées entre elles par des ponts disulfures (symbolisés « S -- S » sur ce schéma). Les chaînes lourdes sont composées d'un domaine variable ( $V_H$ ) et de plusieurs domaines constants ( $C_H$ ) ; les chaînes légères sont composées d'un domaine variable ( $V_L$ ) et d'un domaine constant ( $C_L$ ). Les régions variables (en orange pâle) sont composées des domaines variables des chaînes lourdes et légères (respectivement  $V_H$  et  $V_L$ ). Chacun de ces domaines contient 3 régions hypervariables (CDR : complementary determining regions, représentées en rouges).

La région constante (en orange foncée) est composée des domaines constants des chaînes lourdes et légères (respectivement  $C_H$  et  $C_L$ ).

La région charnière (symbolisée entre les domaines  $C_H$ ), confère sa flexibilité à l'Ig.



## Introduction.

---

Le système immunitaire (ou SI) regroupe un ensemble de cellules, de tissus et d'organes spécialisés dans la défense de l'organisme. Le SI lutte contre les agressions extérieures (agents pathogènes) mais aussi contre certains dysfonctionnements cellulaires pouvant conduire à la formation de tumeurs. Pour accomplir sa fonction, le SI est doté de capacité d'apprentissage permettant une discrimination des antigènes (Ag) du soi (tolérance) et du non soi (élimination).

Il existe 2 types d'immunité : l'immunité innée non spécifique et l'immunité adaptative spécifique. L'immunité innée constitue la première ligne de défense de l'organisme et repose sur la présence de barrières physiques (cutanée, muqueuse) et sur l'action de cellules de l'immunité innée (macrophages, cellules NK natural killer) capables de discerner des motifs communs aux pathogènes. Quant à l'immunité adaptative, elle constitue la seconde ligne de défense de l'organisme et repose sur l'action dirigée des cellules lymphocytaires capables d'assurer la reconnaissance spécifique des Ag. Contrairement à l'immunité innée, l'immunité adaptative est caractérisée par une mémoire immunitaire permettant la mise en place d'une réponse plus rapide lors d'une seconde rencontre avec l'antigène. Les cellules lymphocytaires sont de 2 types : les lymphocytes T et B, respectivement acteurs de l'immunité adaptative à médiation cellulaire et humorale.

Nous nous intéresserons plus particulièrement aux lymphocytes B (LB) : capables de reconnaître de façon spécifique un grand nombre d'Ag différents grâce à leur récepteur de surface, le BCR (B Cell Receptor). Ce dernier est composé d'une immunoglobuline (Ig) membranaire constituant le module de reconnaissance antigénique, associée à un hétérodimère  $Ig\alpha - Ig\beta$  constituant le module de signalisation. Chaque cellule B possède à sa surface plusieurs centaines de BCR spécifiques d'un déterminant antigénique donné. La collection complète des spécificités antigéniques portée par l'ensemble des cellules B constitue le répertoire antigénique. La liaison entre le BCR et l'Ag est déterminante pour induire l'activation des cellules B et leur différenciation jusqu'au stade ultime de différenciation B en cellules plasmocytaires sécrétrices d'Ig sous forme soluble : les anticorps (Ac). Ces Ac vont lier les Ag dans le but de les éliminer.

Une molécule d'Ig monomérique est formée par la réunion de 4 chaînes polypeptidiques. Il s'agit de 2 chaînes lourdes (Heavy, H) identiques entre elles et 2 chaînes légères (Light, L) également identiques entre elles (**Figure 1**). Les chaînes lourdes et légères sont associées entre elles par des ponts disulfures (Porter, 1973). Chaque chaîne lourde contient un domaine peptidique variable ( $V_H$ ) et 3 à 4 domaines peptiques constants ( $C_{H1}$ ,  $C_{H2}$ ,  $C_{H3}$  et  $C_{H4}$ ) et chaque chaîne légère contient 2 domaines ( $V_L$  et  $C_L$ ). Au niveau d'une molécule d'Ig, il faut distinguer 2 régions variables constituant 2 sites spécifiques de liaison à l'antigène et une région constante qui confère à la molécule ses fonctions effectrices. Une région variable est constituée du domaine variable  $V_H$  de la chaîne lourde et du domaine variable  $V_L$  de la chaîne légère. Quant aux autres domaines, ils forment la



région constante. A noter que la distribution de la variabilité n'est pas équivalente sur l'ensemble de la région variable d'une Ig. En effet, chaque domaine variable ( $V_H$  et  $V_L$ ) est caractérisé par la présence de 3 régions hypervariables portant la spécificité antigénique (nommées CDR pour complementary determining regions) et de 4 régions « charpente » de variabilité moindre (nommées FR pour framework) (Lara-Ochoa et al., 1994). Il existe 2 types de chaînes légères ( $\kappa$  ou  $\lambda$ ) et 5 types de chaînes lourdes :  $\mu$ ,  $\delta$ ,  $\gamma$ ,  $\alpha$  et  $\epsilon$  correspondant aux 5 classes (ou isotypes) d'Ig : IgM, IgD, IgG, IgA et IgE. Ces isotypes se trouvent sous forme monomérique, sauf les Ig de classe M et A qui existent sous forme pentamérique et dimérique respectivement. Les Ig de classe M sont les premières Ig à être produites lors d'une réponse immunitaire et possèdent un fort pouvoir agglutinant des micro-organismes. Les IgD sont impliquées dans les réponses anti-bactériennes. Les IgG sont les Ig majoritaires, et interviennent dans les réponses anti-virales, anti-parasitaires et anti-bactériennes selon les sous classes. Les IgA jouent un rôle dans l'immunité mucoale et enfin les IgE dans la réponse anti-parasitaire et allergique. Des variations de plusieurs types existent entre les différentes classes d'Ig, telles que la longueur de la région constante, la longueur de la région charnière, le nombre et la position des ponts disulfures et des sites de glycosylation (Schroeder and Cavacini, 2010).

Les processus de genèse des Ig sont très complexes et spécifiques aux cellules B. En effet, les gènes codant les Ig ne sont pas fonctionnels dans leur configuration germinale. Les loci des chaînes lourdes et légères d'Ig doivent subir d'importants remaniements qui permettront l'expression de chaînes fonctionnelles capables de s'associer pour former une Ig. Au cours de leur développement, les LB sont donc capables de remanier leur génome de façon à générer une immense diversité d'Ig, nécessaire pour une réponse immunitaire efficace et spécifique. Ces remaniements peuvent représenter une menace pour l'intégrité génomique et doivent donc être strictement contrôlés dans le temps et dans l'espace.

Mon travail de thèse concerne l'étude du rôle d'éléments *cis*-régulateurs, en particulier de la région  $E\mu$  située au sein du locus codant les chaînes lourdes d'immunoglobuline (locus *IgH*) et d'éléments *trans*-régulateurs associés. Dans un premier temps, nous avons montré l'importance de la région  $E\mu$  dans la régulation du processus d'hypermutation somatique des gènes d'Ig (événement de mutations introduites dans les régions variables des gènes d'Ig). Dans un second temps, nous avons évalué l'implication de cette même région régulatrice  $E\mu$  sur les « dommages collatéraux » (induits par le processus d'hypermutation) infligés aux « voisins » des gènes d'Ig que sont les cibles illégitimes (oncogènes) aussi appelées « AID off targets ».

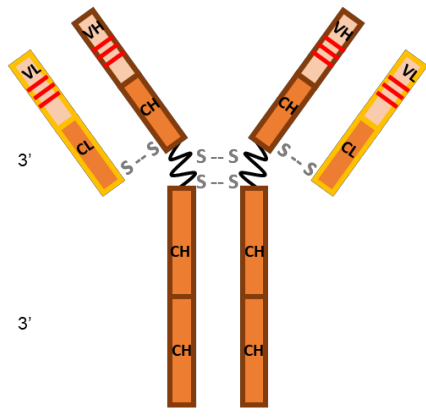
Dans ce manuscrit, je décrirai tout d'abord l'organisation des gènes d'Ig puis les différentes étapes du développement des lymphocytes B et les phénomènes de remaniements associés. Enfin, je détaillerai les éléments *cis*-régulateurs décrits au locus *IgH* impliqués dans le contrôle des modifications géniques complexes que subissent les gènes d'Ig et parfois, de façon accidentelle, les gènes voisins « AID off targets ».

**Gènes des chaînes légères**

a. Chromosome 6 : locus *Igk*

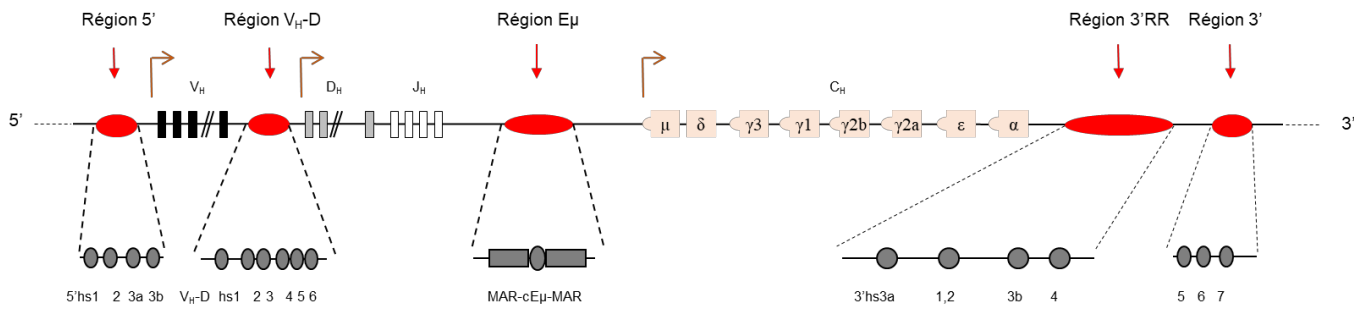


b. Chromosome 16 : locus *Igλ*



**Gènes des chaînes lourdes**

c. Chromosome 12 : locus *IgH*



**Figure 2 : Représentation schématique des loci des gènes des immunoglobulines.**

Les segments de variabilités (V), de diversité (D), de jonction (J) et constants (C) sont annotés et représentés par des rectangles. (Cette représentation n'est pas à l'échelle).

- a. Représentation schématique du locus de chaîne légère *Igk*.
- b. Représentation schématique du locus de chaîne légère *Igλ*.
- c. Représentation schématique du locus de chaîne lourde *IgH* et de ses éléments régulateurs. Les éléments régulateurs sont détaillés et symbolisés de couleur rouge. Les flèches de couleur marron indiquent la présence de promoteurs.

## Chapitre I. Organisation des gènes des immunoglobulines.

---

Chez l'homme et la souris, les gènes codant les Ig sont organisés en 3 loci localisés sur 3 chromosomes différents. Il existe 2 loci (*Igκ* et *Igλ*) codant respectivement les chaînes légères des Ig de type K et λ et un locus codant les chaînes lourdes d'Ig (locus *IgH*).

Une succession de segments de gènes compose les loci d'Ig : segments de variabilité (V), de diversité (D, uniquement au locus *IgH*), de jonction (J) et constants (C). Les régions variables et constantes sont codées à partir des segments V(D)J et C, respectivement.

Les loci murins et humains des immunoglobulines présentent une organisation assez proche, le modèle souris constitue donc un modèle de choix pour leur étude (Hirano et al., 2011). Mon travail de thèse s'est intéressé aux loci murins des immunoglobulines, je détaillerai donc uniquement leur organisation. Nous nous intéresserons plus particulièrement à l'organisation du locus des chaînes lourdes des Ig (locus *IgH*) et aux éléments *cis*-régulateurs qui le jalonnent.

### I.1. Les loci des chaînes légères des immunoglobulines.

#### I.1.1. Le locus kappa (*Igκ*).

Le locus *Igκ* est situé sur le chromosome 6 murin. Dans sa configuration germinale (non réarrangée), ce locus est constitué de 176 segments  $V_{\kappa}$  (dont plus de 90 fonctionnels) et 5 segments  $J_{\kappa}$  (dont 4 fonctionnels) codant la région variable de la chaîne légère. Il existe un seul segment  $C_{\kappa}$  codant la région constante (**Figure 2a**) (Giudicelli et al., 2005) (Schlissel, 2004).

#### I.1.2. Le locus lambda (*Igλ*).

Le locus *Igλ*, situé sur le chromosome 16 murin, possède une structure particulière organisée en 2 groupes de segments :  $V_{\lambda 2}$   $V_{\lambda 3}$   $J_{\lambda 2}$   $C_{\lambda 2}$   $J_{\lambda 4}$   $C_{\lambda 4}$  et  $V_{\lambda 1}$   $J_{\lambda 3}$   $C_{\lambda 3}$   $J_{\lambda 1}$   $C_{\lambda 1}$  (**Figure 2b**). Les segments  $V_{\lambda}$  s'associent préférentiellement aux segments  $J_{\lambda}$   $C_{\lambda}$  suivants. Ainsi les combinaisons de réarrangements observées sont généralement  $V_{\lambda 2}$  et 3 associés à  $J_{\lambda 2}$  –  $C_{\lambda 2}$  ( $J_{\lambda 4}$  et  $C_{\lambda 4}$  étant défectifs) et  $V_{\lambda 1}$  associé à  $J_{\lambda 3}$  –  $C_{\lambda 3}$  et  $J_{\lambda 1}$  –  $C_{\lambda 1}$  (Carson and Wu, 1989; Storb et al., 1989). Chez la souris, seulement 5% des lymphocytes B expriment une chaîne légère de type λ (Arakawa et al., 1996; Engel et al., 1999).

### I.2. Le locus des chaînes lourdes des immunoglobulines (*IgH*).

Situé sur la partie sub-télomérique du chromosome 12 murin, le locus *IgH* s'étend sur près de 3 Mb. Du côté 5' vers 3' (du télomère vers le centromère), le locus *IgH* comprend les segments de variabilité  $V_H$ , les segments de diversité  $D_H$  et les segments de jonction  $J_H$  suivis des gènes constants  $C_H$ .

La région variable de chaîne lourde des Ig est codée à partir de segments de gènes V, D et J. Les bases de données estiment que le locus *IgH* comprend environ 160 segments  $V_H$  (dont 100 fonctionnels) regroupés en 15 familles, 20 segments D regroupés en 4 familles (dont 14 fonctionnels) et 4 segments  $J_H$  (tous fonctionnels)

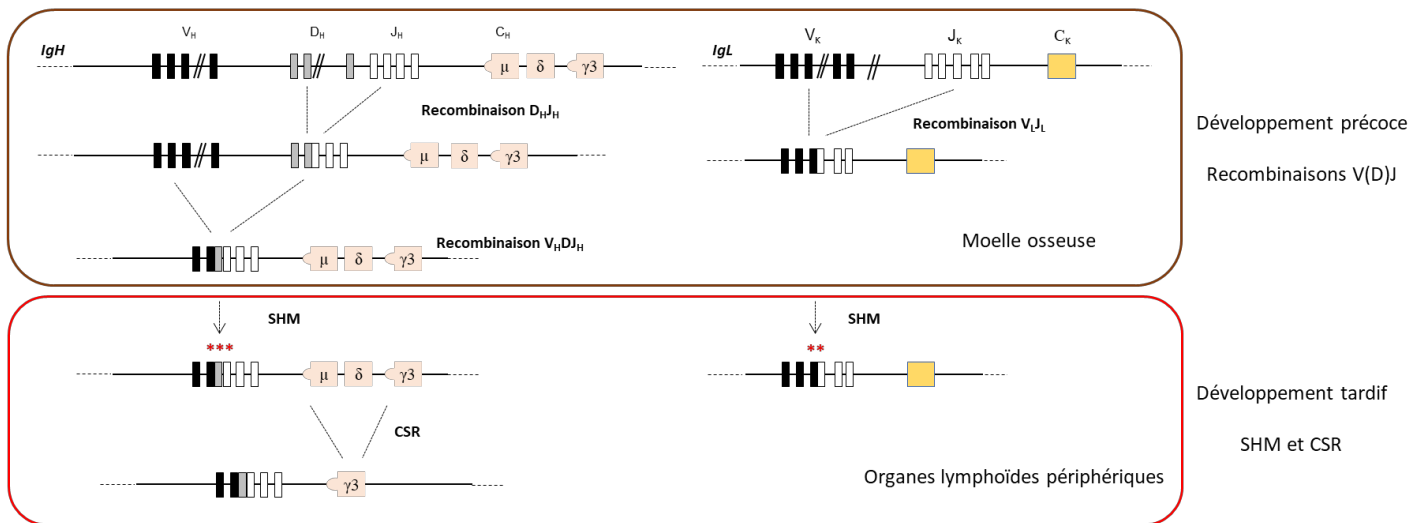


(Giudicelli et al., 2005) (Johnston et al., 2006). Ces derniers sont suivis des segments constants codant la région constante de la chaîne lourde d'Ig et sont au nombre de 8, correspondant aux 8 classes et sous classes des Ig :  $\mu$ ,  $\delta$ ,  $\gamma 3$ ,  $\gamma 1$ ,  $\gamma 2b$ ,  $\gamma 2a$  (ou  $2c$  selon les souches de souris),  $\epsilon$  et  $\alpha$  (Morgado et al., 1989). Les segments constants sont constitués de plusieurs exons codant les domaines constants, la région charnière et les domaines transmembranaires et intra-cytoplasmiques des Ig de surface.

Des éléments régulateurs ont été identifiés au sein du locus *IgH*, parmi lesquels des promoteurs, des activateurs de la transcription (ou enhancers) et des isolateurs (ou insulators). De nombreux promoteurs sont répartis sur le locus entier : en 5' des segments  $V_H$  et  $D_H$ , et en amont des gènes constants. L'activateur transcriptionnel  $E_\mu$ , constitué d'un élément central ou « core  $E_\mu$  » flanqué de 2 régions d'attachement à la matrice nucléaire (matrix attachment regions,  $MAR_{E_\mu}$ ), est localisé entre le dernier segment  $J_H$  et le gène  $C_\mu$ . En aval du gène  $C\alpha$  se trouve la région régulatrice en 3' (3' regulatory region, 3'RR) composée de 4 activateurs transcriptionnels répartis sur près de 30Kb. Enfin, le locus *IgH* comprend plusieurs isolateurs répartis en position 5' du locus (région 5'), entre les segments  $V_H$  et  $D$  (région intergénique  $V_H$ - $D$ ) et en aval de la région 3'RR (**Figure 2c**).

Ces différents éléments *cis*-régulateurs au locus *IgH* participent au contrôle du développement B et des événements de recombinaison associés. Leur fonction sera décrite dans un prochain chapitre (chapitre IV. : « Les régions régulatrices au locus IgH. »).

Développement B et remaniements :



**Figure 3 : Evènements de remaniement s’opérant aux loci des chaînes lourdes (*IgH*) et légères (*IgL*) d’Ig.**

Le développement précoce est associé au processus de recombinaison V(D)J. Le développement tardif est marqué par les évènements de SHM et de CSR.

SHM : somatic hypermutation ou hypermutation somatique, CSR : classe switch recombination ou recombinaison de classe.



## Chapitre II. Développement des lymphocytes B.

---

Le développement B se déroule en 2 phases successives : une phase précoce de développement se déroulant dans la moelle osseuse indépendamment de la présence de l'Ag, et une phase tardive de développement prenant place dans les organes lymphoïdes périphériques suite à la rencontre avec l'Ag (**Figure 3**).

La phase précoce de développement B se caractérise par plusieurs réarrangements entre les segments de gènes codant les chaînes lourdes et légères des Ig, appelées recombinaisons V(D)J. Les réarrangements fonctionnels conduisent à l'expression d'une Ig à la surface des cellules B. Ces recombinaisons V(D)J assurent la diversité des Ig qui est essentielle à la reconnaissance spécifique des innombrables motifs antigéniques.

En périphérie, les gènes d'Ig subissent de nouveaux remaniements : les mécanismes d'hypermutation somatique (SHM, somatic hypermutation) et de commutation de classe (CSR, class switch recombination) permettent respectivement la modification de l'affinité de l'Ig pour l'Ag et le changement de classe d'Ig. Ces remaniements tardifs augmentent la diversité des Ig (le répertoire antigénique) de façon à optimiser l'efficacité de la réponse immune.

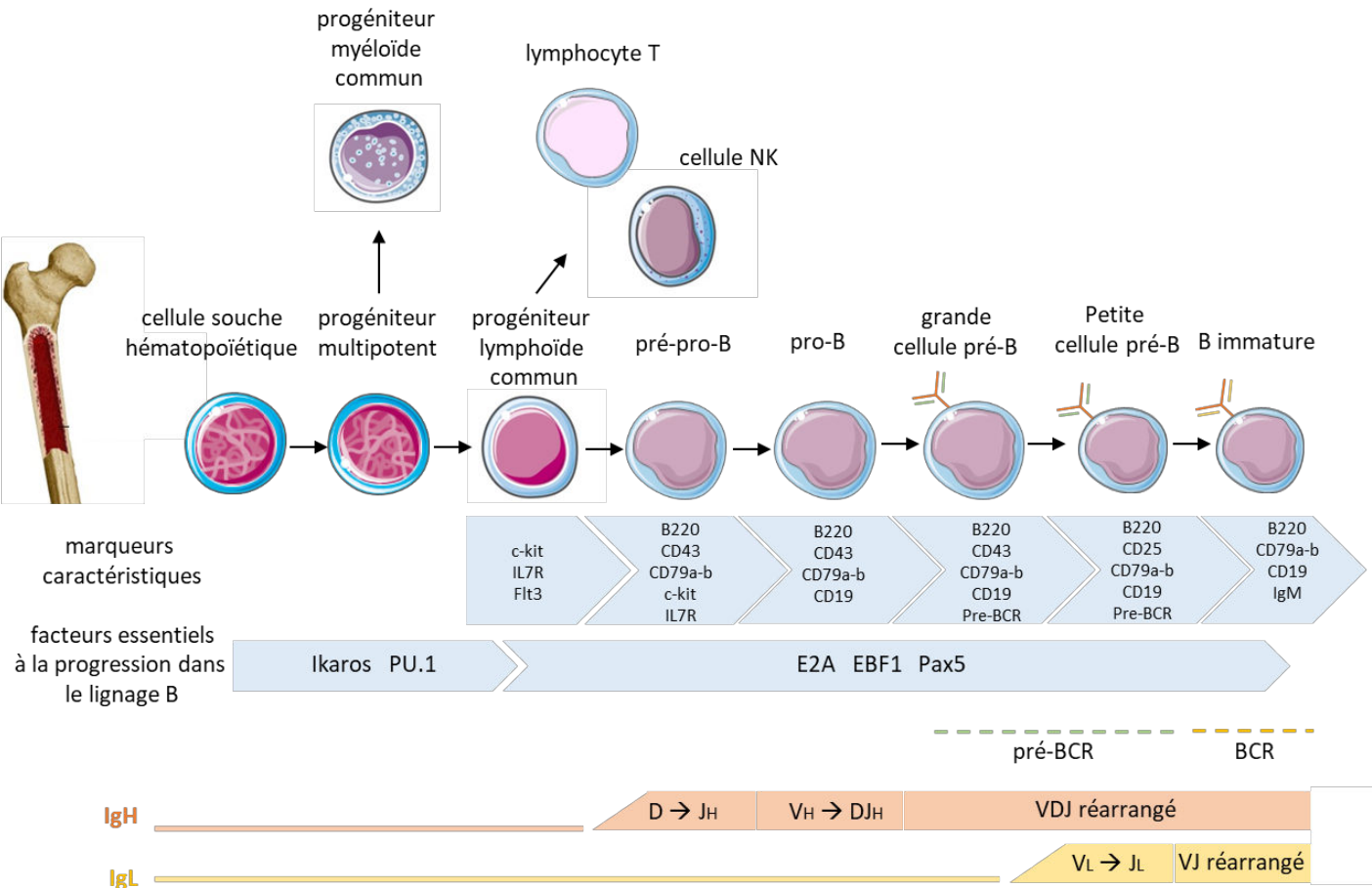
### II.1. Phase précoce du développement B et recombinaisons V(D)J.

Plusieurs événements de recombinaison génique vont avoir lieu dans le but de remodeler les gènes des Ig et de permettre l'expression d'immunoglobulines fonctionnelles. La phase précoce de développement B est marquée par les événements de recombinaison V(D)J permettant de générer des exons VDJ et VJ fonctionnels codant respectivement la partie variable des chaînes lourdes et légères des Ig.

Chez l'animal adulte, le développement précoce des lymphocytes B débute dans la moelle osseuse à partir de **cellules souches hématopoïétiques** (CSH). Le développement précoce est ordonné et caractérisé par plusieurs stades, classiquement définis en fonction de la configuration des gènes des Ig (germinale ou recombinée), de l'expression de marqueurs de surface et de facteurs de transcription spécifiques (**Figure 4**).

#### II.1.1. Stades précoces de développement B : statut des gènes d'immunoglobulines.

Les CSH, cellules multipotentes (à l'origine des différentes lignées cellulaires sanguines) et à haute capacité d'auto-renouvellement, vont se différencier en **progéniteurs multipotents** (MPP) qui ont perdu la capacité d'auto-renouvellement. Les MPP sont, quant à eux, à l'origine de progéniteurs dits oligopotents parmi lesquels on compte les **progéniteurs lymphoïdes communs** (CLP). Les cellules lymphocytaires B, T, les cellules NK (natural killer) et certaines cellules dendritiques sont générées à partir des CLP. La différenciation des CLP en **cellules pré-pro-B**: (stade de développement le plus précoce de la lignée lymphocytaire B) marque un tournant et se caractérise par des remaniements géniques séquentiels et finement contrôlés (Hardy et al., 2007). Les facteurs protéiques essentiels guidant la progression dans le lignage B seront évoqués dans le prochain paragraphe.



**Figure 4 : Représentation schématique des différents stades associés au développement B précoce.**

Ce schéma reprend les principaux stades de développement B précoce avec les marqueurs caractéristiques des différents stades. Les facteurs essentiels à la progression dans le lignage B et la configuration des gènes d'Ig sont également indiqués.

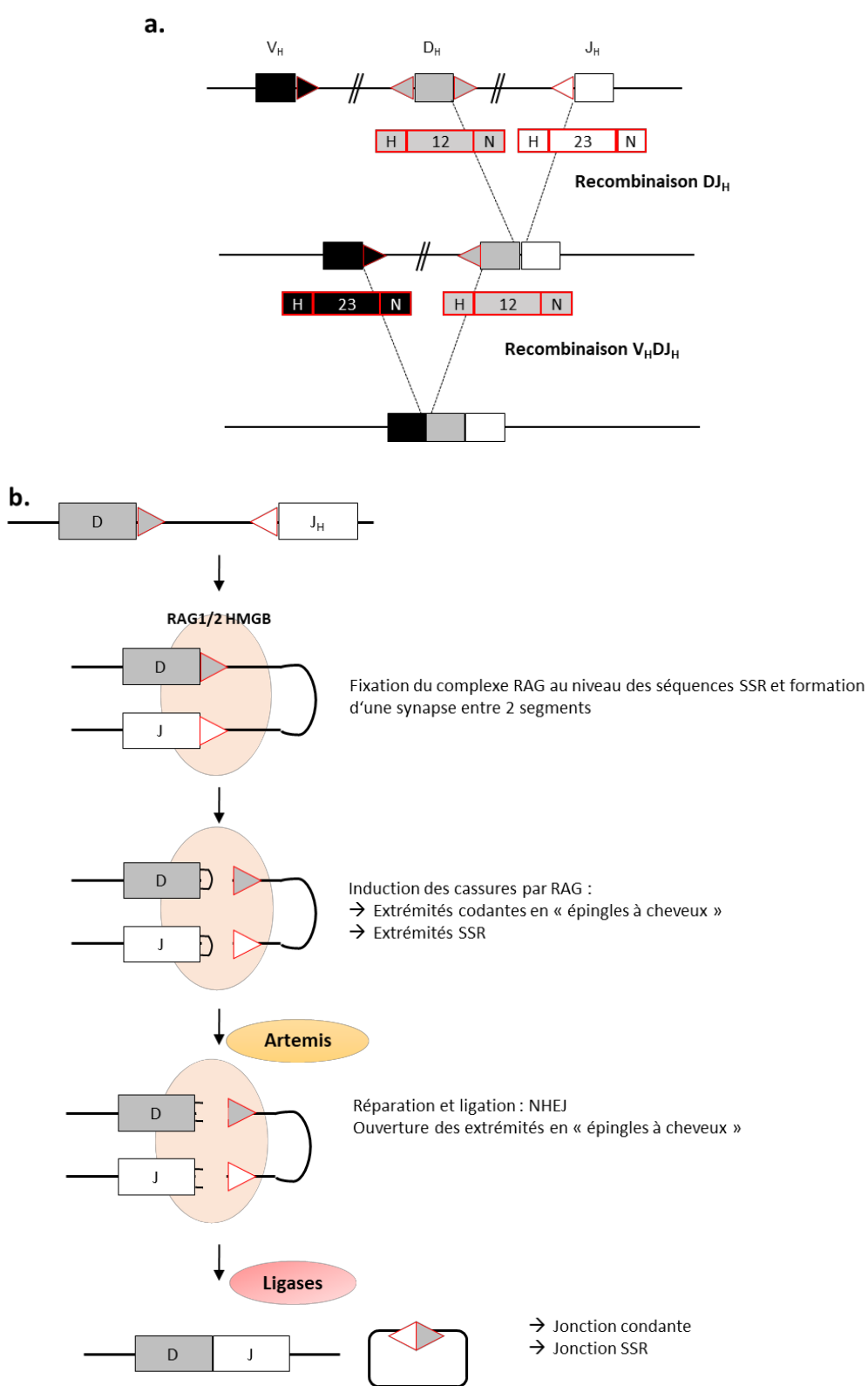
Les recombinaisons V(D)J vont débiter au locus des chaînes lourdes des Ig (locus *IgH*) puis se poursuivent aux loci des chaînes légères (*IgL*) (Alt et al., 1992; Jung et al., 2006). Dès le stade pré-pro-B, un premier type de recombinaison « DJ » va avoir lieu au locus *IgH* : un segment D va être rabouté à un segment J<sub>H</sub>. Cette recombinaison « DJ » se produit sur les 2 allèles. Puis, au **stade pro-B**, un segment V<sub>H</sub> va être rabouté sur le segment DJ<sub>H</sub> préalablement réarrangé. Environ 1/3 des réarrangements V<sub>H</sub> vers DJ<sub>H</sub> sont productifs (cadre de lecture respecté) et ne vont alors concerner qu'un seul allèle (Alt et al., 1984). Si la recombinaison est non productive, alors le second allèle va subir un réarrangement du même type « VDJ ». Les réarrangements fonctionnels permettent l'expression d'une chaîne lourde « μ » à la surface des cellules B qui peuvent alors poursuivre leur développement. Dans le cas où les réarrangements V<sub>H</sub> vers DJ<sub>H</sub> sont non productifs pour les 2 allèles, la cellule B est éliminée par apoptose.

Une chaîne lourde « μ » fonctionnelle produite (issue d'un réarrangement VDJ productif) va s'associer à une pseudo chaîne légère composée des produits des gènes *V<sub>pré-B</sub>* et *λ5* (Mårtensson and Ceredig, 2000; Melchers et al., 1993). Cet ensemble (chaîne lourde « μ » et pseudo chaîne légère), en combinaison avec les molécules de transduction du signal nommées Igα (ou CD79a) et Igβ (ou CD79b), constitue le récepteur pré-BCR exprimé à la surface des cellules B et définit le **stade pré-B**. La signalisation par le pré-BCR induit la prolifération des cellules B et inhibe le réarrangement éventuel du second allèle (Mårtensson et al., 2002). Ce phénomène, d'exclusion allélique, empêche la cellule B de produire 2 chaînes lourdes différentes à partir des 2 loci *IgH* : à la surface d'une cellule B, une seule chaîne lourde est exprimée à partir d'un seul locus *IgH* réarrangé de façon fonctionnelle.

Le stade de développement pré-B est caractérisé par 2 sous populations cellulaires : les grandes cellules pré-B et les petites cellules pré-B. Les **grandes cellules pré-B** subissent plusieurs cycles de division puis sont à l'origine des **petites cellules pré-B**. C'est à partir de ce stade (petites cellules pré-B) que les réarrangements des chaînes légères vont avoir lieu. Ces réarrangements de type V<sub>L</sub> vers J<sub>L</sub> sont également ordonnés et vont tout d'abord concerner les loci des chaînes légères κ puis, en cas de réarrangements non productifs, les loci des chaînes légères λ.

Un réarrangement productif permet à la cellule B de produire une chaîne légère fonctionnelle qui va s'associer à la chaîne lourde formant ainsi une molécule complète d'immunoglobuline de classe M (IgM). Une recombinaison de type V<sub>Lκ</sub>-J<sub>Lκ</sub> productive va entraîner un blocage des recombinaisons au niveau du second allèle κ (exclusion allélique) et au niveau des loci λ (exclusion isotypique).

Dès l'expression d'une Ig fonctionnelle à leur surface, les cellules B sont dites **immatures** et expriment à leur surface le récepteur BCR (B Cell Receptor) constitué par une IgM en combinaison avec les molécules de transduction du signal précédemment évoquées. Avant de quitter la moelle osseuse pour rejoindre les organes lymphoïdes périphériques, les cellules B immatures vont se soumettre à une sélection négative aboutissant à



**Figure 5 : Règle dite « 12/23 » et recombinaison VDJ au locus *IgH*.**

Les séquences signal de recombinaison (SSR) sont schématisées par des triangles rouges.

**a.** La règle dite « 12-23 » va guider la recombinaison entre une séquence SSR possédant un espaceur de 12 nucléotides et une séquence SSR possédant un espaceur de 23 nucléotides. Cette règle permet une recombinaison ordonnée autorisant un segment D à se réarranger sur un segment  $J_H$  puis un segment  $V_H$  à s'associer sur le segment  $DJ_H$  préalablement réarrangé. (H : motif de 9 nucléotides et N : motif de 7 nucléotides)

**b.** Représentation schématique du mécanisme de recombinaison. Le complexe RAG va initier des cassures de l'ADN créant ainsi des extrémités codantes et SSR. Le complexe de réparation NHEJ est ensuite recruté et fait notamment intervenir la protéine Artémis qui ouvre les structures en épingles à cheveux. Les ligases vont ensuite intervenir pour générer les jonctions codantes et SSR.

l'élimination par apoptose des cellules possédant un BCR auto-réactif (capable de reconnaître les antigènes du soi). Cependant un mécanisme de « sauvetage » des cellules B auto-réactives, appelé édition du BCR (ou BCR editing), peut être mis en place (Halverson et al., 2004). Ce mécanisme est à l'origine de nouvelles recombinaisons (en particulier aux loci des chaînes légères) et aboutit à la production de nouvelles chaînes et donc d'un nouveau BCR avec une spécificité antigénique modifiée et potentiellement plus compatible avec les antigènes du soi. Si ce processus d'édition du BCR ne permet pas de produire une cellule avec un BCR non auto-réactif, la cellule B va être rendue anergique ou éliminée par apoptose (Norvell et al., 1995; Pike et al., 1982). Les cellules B non auto-réactives vont poursuivre leur développement en périphérie où elles vont subir de nouvelles étapes de maturation.

Durant le développement B précoce, les remaniements des gènes des Ig sont finement régulés. Les réarrangements entre segments de gènes V(D)J sont dirigés par des séquences non codantes spécifiques adjacentes aux segments V, D et J et appelées séquences signal de recombinaison (SSR) (**Figure 5a**). Les SSR sont constituées par 2 motifs hautement conservés : un motif de 7 nucléotides et un motif de 9 nucléotides espacés par des séquences de 12 ou 23 nucléotides. La règle dite « 12-23 » va guider la recombinaison entre une séquence SSR possédant un espaceur de 12 nucléotides et une séquence SSR possédant un espaceur de 23 nucléotides. Cette règle permet une recombinaison ordonnée autorisant un segment D à se réarranger sur un segment J<sub>H</sub> puis un segment V<sub>H</sub> à s'associer sur le segment DJ<sub>H</sub> préalablement réarrangé. Aux loci des chaînes légères, la recombinaison selon la règle « 12-23 » s'applique également (Schatz and Ji, 2011).

Le mécanisme de recombinaison V(D)J se déroule en 2 étapes successives : (1) génération de cassures double brin de l'ADN et (2) ligature/réparation des cassures (**Figure 5b**). Lors de la 1<sup>ère</sup> étape : un complexe enzymatique, comprenant les endonucléases RAG-1 et RAG-2 (recombination activating genes) (Mombaerts et al., 1992; Shinkai et al., 1992) et les protéines HMGB (high mobility group), va lier l'ADN au niveau des séquences SSR des segments à recombinaison permettant leur rapprochement sous forme d'une synapse (Lovely et al., 2015; Pavlicek et al., 2008). Les enzymes RAG vont ensuite catalyser des cassures double brin de l'ADN. Les enzymes RAG, accompagnées des facteurs KU70/KU80, vont rester associés à l'ADN après clivage de façon à protéger les extrémités générées. La présence de cassures double brin va activer la voie de réponse aux dommages à l'ADN dépendante de la protéine ATM, et recruter le système de réparation NHEJ (Non Homologous End Joining) faisant intervenir différents acteurs protéiques (dont Artemis, DNA-PKs, DNA ligase IV et XRCC4) (Boboila et al., 2010). Les extrémités vont être raboutées aboutissant au rapprochement d'un segment D vers J ou V vers DJ, avec élimination des séquences SSR. La jonction entre les segments de gènes (dite « jonction codante ») fait intervenir l'enzyme TdT (Terminale Deoxynucleotidyl Transferase) qui introduit aléatoirement des nucléotides aux extrémités 3' augmentant ainsi la diversité du répertoire antigénique (Desiderio et al., 1984).



### II.1.2. Stades précoces de développement B : facteurs essentiels à la progression dans le lignage B.

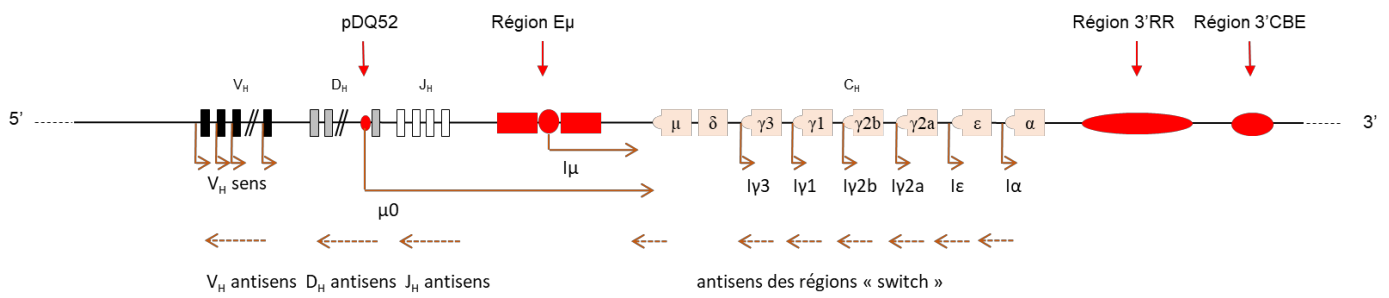
De nombreux facteurs interviennent dans la régulation de la lymphopoïèse B, tels que les facteurs Ikaros, PU.1 (Purine box factor -1), E2A (E box protein binding 2A), EBF1 (Early 1b cell Factor 1) et Pax5 (Paired box 5) (Hardy et al., 2007; Monroe and Dorshkind, 2007; Ye and Graf, 2007) (**Figure 4**). Ces facteurs sont d'une importance capitale car ils guident la progression dans le lignage B tout en bloquant l'engagement vers un lignage non B (Nutt and Kee, 2007).

La transition de CSH vers MPP puis vers CLP dépend notamment de l'expression des facteurs de transcription Ikaros et PU.1. En absence de Ikaros, les MPP ne sont plus capables de se différencier en CLP (Yoshida et al., 2006). De plus, Ikaros intervient dans le processus de recombinaison VDJ au locus *IgH* en régulant l'expression des gènes codant les enzymes RAG et l'accessibilité des segments  $V_H$  (Reynaud et al., 2008). Des études plus récentes, montrent que Ikaros est essentiel pour la transition pro-B vers pré-B (Schwickert et al., 2014). Quant au facteur de transcription PU.1, son action semble être dose dépendante : un niveau intermédiaire de PU.1 est requis pour une différenciation vers le lignage B (DeKoter and Singh, 2000; Ye and Graf, 2007). Sous l'influence des facteurs de transcription Ikaros et PU.1, les MPP se différencient progressivement en CLP caractérisés par l'expression des enzymes RAG et du récepteur à l'IL7 (IL7-R) (DeKoter et al., 2002; Igarashi et al., 2002).

La progression CLP vers les stades pro puis pré-B fait intervenir les facteurs de transcription E2A (E12 et E47, codés par le gène *E2a*). Un blocage complet du développement B (dès le stade CLP) est observé dans des modèles murins n'exprimant pas E2A (Zhuang et al., 1994). E2A est un facteur régulateur essentiel, impliqué dans l'initiation de l'expression de gènes spécifiques du lignage B (*Rag1*, *Rag2*, *Mb1*...) mais également dans la survie des lymphocytes B matures (Lazorchak et al., 2006; O'Riordan and Grosschedl, 1999).

L'action combinée du facteur E2A et de la signalisation induite par l'IL7 va induire l'expression d'un autre facteur clé du développement B : EBF1. L'expression ectopique des facteurs E2A ou EBF1 (en combinaison avec les enzymes RAG) peut induire le processus de recombinaison V(D)J dans une lignée non lymphoïde. Ces facteurs semblent donc favoriser l'accès aux régions D et  $J_H$  (Goebel et al., 2001; Romanow et al., 2000).

EBF1 semble également impliqué dans le contrôle de l'expression de PAX5 : un autre facteur de transcription indispensable au développement B (Maier and Hagman, 2002; O'Riordan and Grosschedl, 1999). Des cellules pro-B issues des souris déficientes pour PAX5 ne poursuivront pas leur progression dans le lignage B (Mikkola et al., 2002; Nutt et al., 1999). PAX5 réprime l'expression d'un ensemble de gènes spécifiques aux lignages non B, tel que *Notch1* impliqué dans la différenciation lymphocytaire T (Souabni et al., 2002) et permet l'expression de protéines spécifiques du lignage B, telles que la protéine costimulatrice CD19 ou encore la protéine CD79a entrant dans la composition BCR (Cobaleda et al., 2007; Kozmik et al., 1992).



**Figure 6 : Transcrits germinaux sens et antisens initiés au locus *IgH*.**



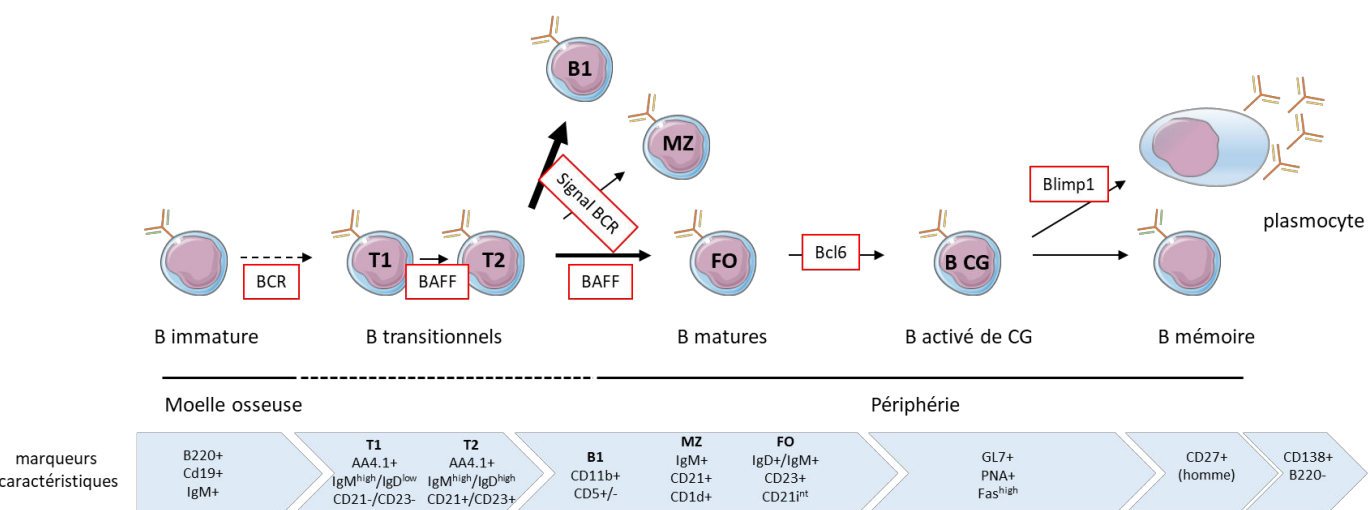
L'ensemble de ces facteurs indispensables au développement B, sont interconnectés et agissent ensemble pour guider la progression dans le lignage B (Medina et al., 2004). La transition entre les différents stades de développement dépend également du bon déroulement des recombinaisons V(D)J et de l'expression d'Ig fonctionnelles à la surface des cellules B.

### **II.1.3. Régulation de l'accessibilité au locus *IgH* lors des recombinaisons VDJ.**

Le mécanisme de recombinaison VDJ se déroule selon des étapes bien ordonnées, suggérant une accessibilité séquentielle des segments D, J<sub>H</sub> et V<sub>H</sub> aux endonucléases du complexe RAG. Plusieurs transcrits germinaux (ne codant pas de protéine) sont initiés transitoirement au locus *IgH*. Leur détection et leur localisation corrént avec le déroulement séquentiel des évènements de recombinaison s'opérant au locus *IgH*. Il a donc été supposé que la transcription germinale soit associée à l'accessibilité des régions à recombiner (Yancopoulos and Alt, 1985). Cette hypothèse est soutenue par des données plus récentes faisant le lien entre déroulement ordonné des recombinaisons VDJ et modifications dynamiques de la chromatine accompagnant la transcription germinale.

#### **II.1.3.1. La transcription germinale.**

La transcription germinale sens est initiée avant les remaniements de type D vers J<sub>H</sub> à partir de l'activateur intronique E<sub>μ</sub> et du promoteur pDQ52 (promoteur associé au segment D<sub>H</sub> localisé le plus en 3'). Ces transcrits non codants sont nommés, respectivement, I<sub>μ</sub> et μ0 et s'étendent vers C<sub>μ</sub> (Alessandrini and Desiderio, 1991; Lennon and Perry, 1985; Schlissel et al., 1991). Des transcrits antisens D<sub>H</sub> ont également été identifiés et précèdent les réarrangements D<sub>H</sub> vers J<sub>H</sub> (Bolland et al., 2007a). Après les recombinaisons D<sub>H</sub> vers J<sub>H</sub>, la transcription sens est initiée en amont des segments V<sub>H</sub> à partir de leur promoteur (pV<sub>H</sub>). Ces promoteurs, connus pour être bidirectionnels, sont également à l'origine de transcrits antisens identifiés dans cette même région V<sub>H</sub> (Bolland et al., 2004a) (**Figure 6**). Ces transcrits antisens sont observés au cours des réarrangements D – J<sub>H</sub> et ne sont plus détectés à la suite des réarrangements V<sub>H</sub> – DJ<sub>H</sub>. Il a donc été suggéré que ces transcrits antisens contribuent à l'accessibilité de la région V<sub>H</sub> constituant un prérequis indispensable pour l'étape de recombinaison de type V<sub>H</sub> vers DJ<sub>H</sub> (Bolland et al., 2004a). Il a également été suggéré que des transcrits pourraient être initiés à partir du segment DJ<sub>H</sub> réarrangé et épissés sur le gène constant C<sub>μ</sub>, appelés transcrits D<sub>μ</sub> (Reth and Alt, 1984). Cependant, les études récentes ne font pas mention de tels transcrits (Bolland et al., 2007a). Par ailleurs il est intéressant de noter la présence de 2 sites CBE (CTCF binding element liant la protéine CTCF ou CCCTC-binding factor) dans la très grande région intergénique V<sub>H</sub>-D (Chaumeil and Skok, 2012; Degner et al., 2009). Ces éléments semblent agir telles que des barrières qui empêchent la transcription antisens de s'étendre à la région V<sub>H</sub> durant les 1<sup>ères</sup> étapes de recombinaison D vers J<sub>H</sub>, assurant ainsi une régulation ordonnée des évènements de remaniement VDJ (Featherstone et al., 2010). Des études montrent que la délétion de la région intergénique V<sub>H</sub>-D (contenant les sites CBE) provoque une augmentation de la transcription



**Figure 7 : Représentation schématique des différents stades associés au développement B tardif.**

Ce schéma reprend les principaux stades de développement B tardif avec les marqueurs caractéristiques des différents stades. Les facteurs impliqués dans la maturation B tardive sont également indiqués.

antisens dans la région D qui se propage vers la région V<sub>H</sub> (Giallourakis et al., 2010). La région intergénique V<sub>H</sub>-D semble également être impliquée dans la conformation du locus *IgH* par sa capacité à fixer le facteur CTCF et la cohésine (Degner et al., 2011a; Guo et al., 2011a).

La transcription germinale est donc corrélée au déroulement ordonné des différents remaniements et semble séquentiellement favoriser l'accessibilité des différents segments à la machinerie de recombinaison.

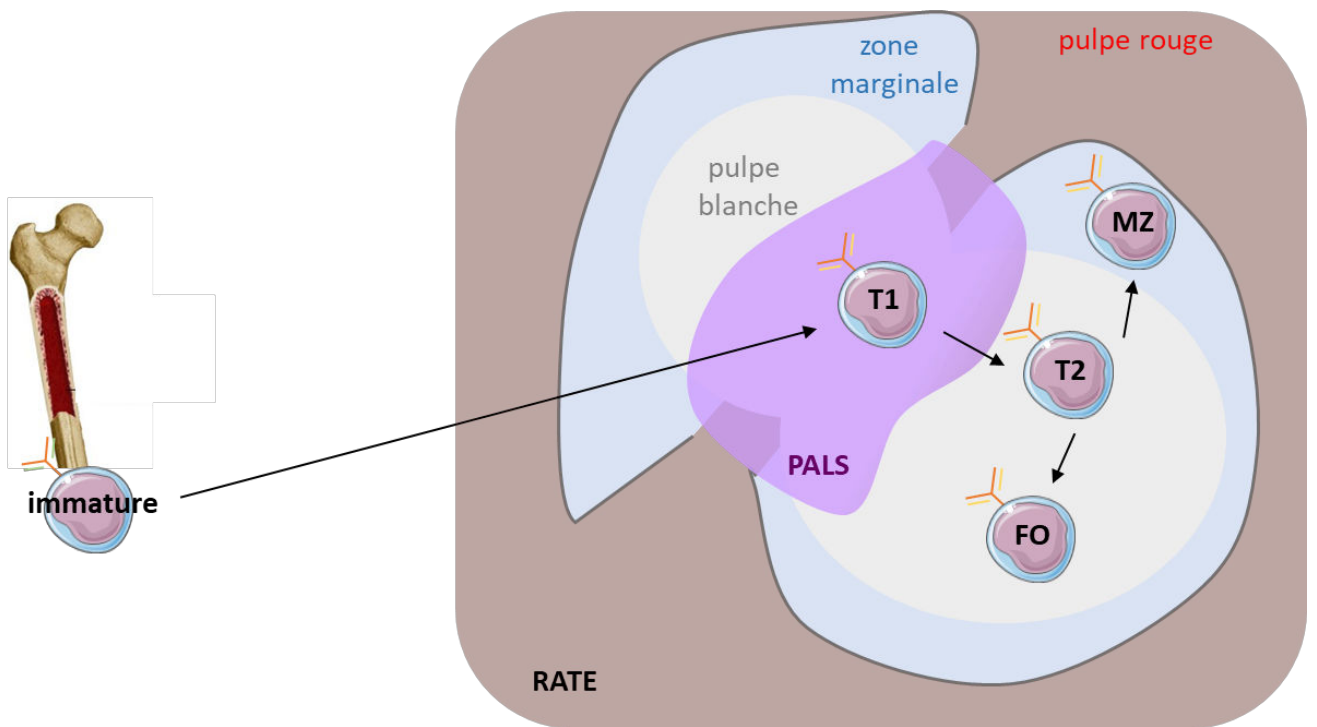
### **II.1.3.2. Les modifications de la chromatine.**

Dans le noyau des cellules eucaryotes, l'ADN génomique est organisé sous forme de nucléosome. Cette structure est constituée par un cœur protéique d'histones autour duquel s'enroule l'ADN. L'extrémité aminoterminal des histones est la cible de nombreuses modifications post-traductionnelles (acétylation, méthylation, phosphorylation...) influençant l'état de compaction de la chromatine (Jenuwein, 2001). La configuration dite « ouverte » de la chromatine, c'est-à-dire transcriptionnellement accessible, est associée à un faible niveau de compaction.

Des modifications dynamiques de la chromatine vont accompagner le processus de recombinaison VDJ (Sen and Oltz, 2006). Les 1<sup>ers</sup> évènements de recombinaison de type D vers J<sub>H</sub> vont avoir lieu dès le stade pré-pro B. A ce stade, les segments D (plus particulièrement les segments localisés en 3' et en 5' de la région D, utilisés de façon plus fréquente) ainsi que les segments J<sub>H</sub> sont caractérisés par des marques actives de la chromatine (Chakraborty et al., 2007; Tsukada et al., 1990). Alors que la région englobant les segments V<sub>H</sub> est associée à des marques inactives de la chromatine (Chakraborty et al., 2007; Morshead et al., 2003). Par ailleurs, il est intéressant de noter que l'endonucléase RAG2 par son domaine PHD (plant homeo-domain) est capable de reconnaître et de lier la marque de la chromatine H3K4me3 (triméthylation de la lysine K4 de l'histone H3) (Liu et al., 2007; Matthews et al., 2007). Cette marque, particulièrement enrichie au niveau des segments D et J<sub>H</sub>, est importante pour le recrutement de l'enzyme RAG2 (Ji et al., 2010). Au stade pro-B, après les réarrangements DJ<sub>H</sub>, les histones des régions d'ADN correspondant aux segments V<sub>H</sub> vont être hyperacétylées : modification associée à une configuration accessible de la chromatine. Dans les cellules pré-B exprimant une chaîne lourde fonctionnelle, les histones de l'ADN constituant la région V<sub>H</sub> sont de nouveau hypoacétylées ; cette région se trouve donc dans une configuration peu accessible limitant les recombinaisons à ce stade (Chowdhury and Sen, 2001, 2003).

## **II.2. Phase tardive du développement B et recombinaisons secondaires.**

Les lymphocytes B immatures quittent la moelle osseuse et poursuivent leur développement en périphérie jusqu'au stade ultime de différenciation des cellules B en plasmocytes (cellules sécrétrices d'Ig sous forme solubles ou anticorps) (**Figure 7**). Le développement tardif des lymphocytes B dépend de la rencontre avec l'antigène et est associé aux évènements d'hypermutation somatique (SHM) et de recombinaison de classe (CSR) permettent respectivement la modification de l'affinité de l'Ig pour l'Ag et le changement de classe d'Ig.



**Figure 8 : Localisation et devenir des cellules B transitionnelles de la rate.**

PALS : periarteriolar lymphoid sheath ou manchons lymphoïdes péri-artériolaires.

Les organes lymphoïdes en périphérie rassemblent des structures dites encapsulées telles que la rate (sur le trajet sanguin) et les ganglions lymphatiques (sur le trajet lymphatique). Il existe également des structures non encapsulées, appelées MALT (mucosae associated lymphoid tissue), associées aux tissus épithéliaux et aux muqueuses. Les organes lymphoïdes périphériques (ou secondaires) sont constitués de structures cellulaires temporaires appelées follicules : lieux de développement de la réponse immunitaire. A un stade avancé de la réponse immune, les follicules contiennent des centres germinatifs qui sont le siège des événements de SHM et de CSR (voir paragraphe II.2.3.3. « La formation du centre germinatif. »).

### II.2.1. Le stade B transitionnel.

Les cellules B immatures quittant la moelle osseuse, pour rejoindre les organes lymphoïdes périphériques sont nommées cellules B transitionnelles. Ces cellules B sont caractérisées par l'expression des marqueurs suivants : AA4.1 (ou CD93), HSA (ou CD24) et une forte expression d'IgM membranaire. Trois sous populations de cellules composent le compartiment B transitionnel et diffèrent selon l'expression de marqueur et leur fonction (Chung et al., 2003). Les cellules B transitionnelles T1 sont des cellules transitionnelles précoces et présentent un phénotype proche des cellules B immatures (Allman et al., 1992). Les cellules B T1 se différencient en cellules B transitionnelles tardives ou T2. Ces cellules B T2 vont poursuivre leur développement et sont à l'origine des cellules B matures. Selon la signalisation induite par le BCR, les cellules B matures vont avoir différents devenir. Il existe une 3<sup>ème</sup> sous population de cellules B transitionnelles T3 (Allman et al., 2001). Ces cellules ne semblent pas être à l'origine des cellules B matures, mais constitueraient un pool de cellules anergiques et majoritairement auto-réactives (Liubchenko et al., 2012; Teague et al., 2007).

Les **cellules B transitionnelles T1**, notamment définies comme étant  $IgM^{high}/IgD^{low}/CD21^{-}/CD23^{-}$ , transitent via la circulation sanguine depuis la moelle osseuse vers la rate. Les cellules B T1 vont majoritairement se localiser au niveau des manchons lymphoïdes péri-artériolaires (ou PALS periarteriolar lymphoid sheath) situés dans la pulpe blanche (**Figure 8**). Pour ces cellules, une étape de sélection s'opère également en périphérie. La stimulation antigénique du BCR de ces cellules B T1 conduit à une sélection négative des cellules auto-réactives par apoptose (Petro et al., 2002).

Les **cellules B transitionnelles T2** ont migré vers les follicules (lieu de présentation des Ag exogènes) et constituent un stade de développement intermédiaire entre les stades B T1 et B matures (**Figure 8**). Ces cellules B de type T2 sont toujours caractérisées par une forte expression d'IgM et expriment également les marqueurs IgD, CD21 et CD23. Plusieurs études montrent que les cellules B T2 surexpriment des protéines favorisant la survie et sont moins sensibles à l'apoptose. Les cellules B T2 sont également capables de proliférer suite à une stimulation antigénique (Chung et al., 2002). La signalisation induite par le BCR va conditionner la suite du développement B et réguler le devenir des cellules B au stade mature (Niiri and Clark, 2002).



### II.2.2. Signalisation induite par le BCR et devenir des cellules B.

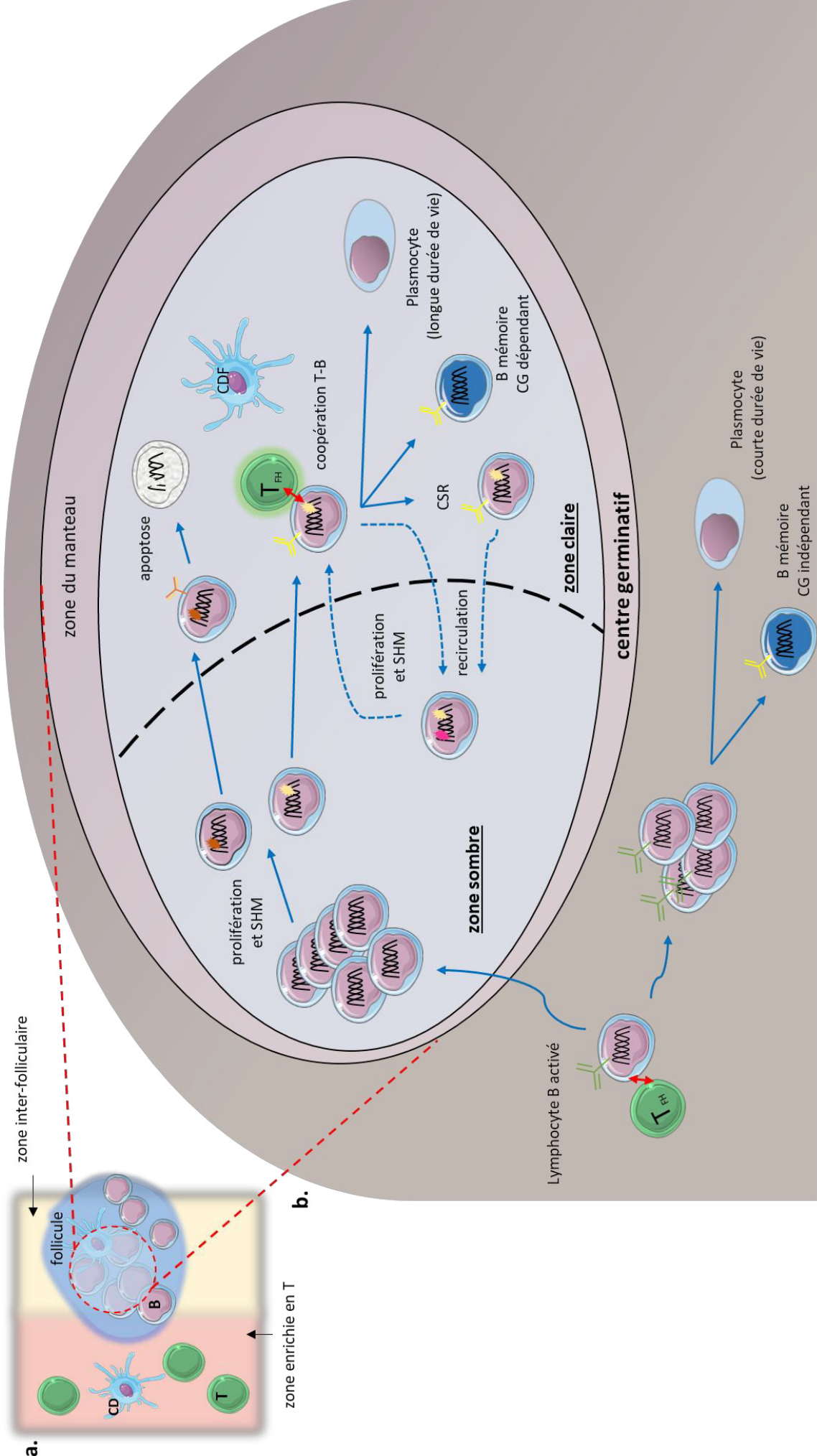
Le devenir de la cellule B ou « B cell fate » dépend de l'expression et de la signalisation induite par le BCR. Le BCR a notamment pour rôle la transmission du signal assurant la continuité du développement B ; et la capture de l'Ag qui sera ensuite intégré et présenté au sein du complexe CMH II (complexe majeur d'histocompatibilité II), étape nécessaire à l'activation complète des cellules B.

Les cellules B immatures exprimant un BCR quittent la moelle osseuse et se dirigent vers la rate pour poursuivre leur différenciation en cellules B matures de type B1 ou B2 (B2 de la zone marginale MZ ou folliculaire FO). Cette différenciation est guidée en fonction de la force du signal induit par le BCR. Pour exemple, une signalisation forte est préférentiellement associée au développement des cellules B de type FO alors qu'une signalisation plus faible va favoriser la différenciation en cellules B de type MZ (Pillai and Cariappa, 2009).

Comme mentionné précédemment, le BCR est composé d'une Ig de surface constituant le module de reconnaissance antigénique associée à un hétérodimère Ig $\alpha$  (ou CD79a) - Ig $\beta$  (ou CD79b) constituant le module de signalisation. La liaison d'un Ag au BCR va initier une série de phosphorylations permettant la transduction du signal via différents seconds messagers. Cette cascade de signalisation aboutira à une réponse cellulaire orientant la différenciation des cellules B (Packard and Cambier, 2013). Une altération de signalisation peut être associée à un défaut de répartition des différentes sous populations de cellules B (Allman et al., 2004). D'autres signalisations sont impliquées dans le devenir des cellules B. La liaison du facteur BAFF à son récepteur, notamment présent à la surface des cellules B transitionnelles, favorise la différenciation en cellules B matures (Khan, 2009; Sasaki et al., 2004). Les cellules B expriment également à leur surface des récepteurs de type TLR (toll like receptor) capables de reconnaître des motifs associés à différents pathogènes (tels que les lipopolysaccharides associés à la membrane de certaines bactéries). Une double stimulation des cellules B transitionnelles T2, par le BCR et le TLR va favoriser une différenciation en cellules B FO ; alors que la seule stimulation via le TLR va plutôt être à l'origine d'une différenciation en cellules B MZ (Su et al., 2004). L'orientation vers les diverses sous populations B résulte donc d'un équilibre complexe entre différentes voies de signalisation interconnectées.

### II.2.3. Le stade B mature.

Il existe 3 sous populations majeures de cellules B matures : les cellules de type B1 et les cellules de type B2 regroupant les cellules B de la zone marginale (MZ) et les cellules B folliculaires (FO). Ces différentes populations de cellules B matures diffèrent selon leur localisation, leur fonction et leur expression phénotypique (**Figures 7 et 8**).



**Figure 9 : Représentation schématique de la structure d'un centre germinatif.**

a. Les lymphocytes B de type folliculaire sont localisés à l'intérieur des follicules des organes lymphoïdes secondaires. Les follicules sont séparés par des régions dites inter-folliculaires et se trouvent à proximité de régions enrichies en lymphocytes T.

b. Zoom sur l'organisation d'un centre germinatif.

CG : centre germinatif, SHM : somatic hypermutation ou hypermutation somatique, CSR : class switch recombination ou recombinaison de classe.



### II.2.3.1. Les cellules B1.

Les **cellules B1** sont retrouvées au niveau des cavités pleurales et péritonéales et sont de 2 types selon l'expression du marqueur CD5 (Hayakawa et al., 1985). Alors que les cellules B1-a sont CD5 positives, les cellules B1-b n'expriment pas ce marqueur. Des travaux récents proposent que les cellules B1 soient issues de précurseurs différents de ceux des cellules B2 (Esplin et al., 2009).

Les cellules B1 sont des cellules sécrétrices d'Ac naturels (majoritairement IgM) poly-spécifiques et poly-réactifs reconnaissant des Ag partagés par des pathogènes et des Ag du soi (tels que ceux exprimés par les cellules apoptotiques). Les cellules B1 assurent ainsi une 1<sup>ère</sup> ligne de défense face à différents pathogènes et semblent contribuer au contrôle de la réponse inflammatoire (Baumgarth, 2011, 2017).

### II.2.3.2. Les cellules B2.

Les cellules B matures de type B2 se divisent en 2 catégories : les **cellules B de la zone marginale (MZ)** et les **cellules B folliculaires (FO)**.

Les **cellules B MZ** sont localisées dans la rate à l'interface entre la pulpe rouge et la pulpe blanche, cette zone est appelée zone marginale (**Figure 8**). Les cellules B MZ possèdent des caractéristiques proches de celles des cellules B1. Ces cellules expriment un BCR peu diversifié capable de reconnaître des motifs communs partagés par différents micro-organismes (Pillai et al., 2005). Les B MZ sont impliquées dans des réponses immunitaires T-indépendantes et sont capables de se différencier rapidement en plasmocytes à courte durée de vie sécrétant d'Ig de faible affinité de type IgM (Balázs et al., 2002; Martin et al., 2001). D'autres études montrent que les cellules B MZ sont capables de répondre à une stimulation T dépendante à l'origine de leur différenciation en plasmocytes à longue durée de vie produisant des Ig de haute affinité de type IgG ou IgA (MacLennan et al., 2003; Song and Cerny, 2003). Enfin, des études proposent un rôle des cellules B MZ dans le transport des Ag vers les follicules (Cinamon et al., 2008; Ferguson et al., 2004).

La majorité des cellules B spléniques sont des **cellules B FO**, localisées au niveau de la pulpe blanche, dans les follicules (**Figure 8**). Suite à la rencontre avec un Ag, les cellules B folliculaires sont à l'origine des réponses immunitaires T-dépendantes et vont participer à la formation des centres germinatifs (CG). Au sein de ces structures, les cellules B poursuivent leur développement et subissent de nouvelles modifications de leurs gènes d'Ig de façon à fournir une réponse immunitaire la plus adaptée possible.

### II.2.3.3. La formation du centre germinatif.

Les lymphocytes B naïfs de types folliculaires sont localisés à l'intérieur des follicules des organes lymphoïdes secondaires. Les follicules sont séparés par des régions dites inter-folliculaires et se trouvent à proximité de régions enrichies en lymphocytes T (**Figure 9a**). A la différence des lymphocytes B activés (ayant rencontrés l'antigène), les lymphocytes B dits naïfs (n'ayant pas encore rencontrés l'antigène) sont caractérisés par



l'absence d'expression des marqueurs PNA (lectines détectées par la peanut agglutinin), FAS (ou CD95) et GL7 (2,6-N-acetyl-neuraminique).

Les lymphocytes B naïfs vont être activés suite à leur rencontre avec un antigène sous sa forme soluble ou présenté par l'intermédiaire de cellules présentatrices d'antigène (CPA), également localisées à l'intérieur des follicules : les cellules dendritiques folliculaires (CDF). Les lymphocytes B activés migrent alors vers la région inter-folliculaire où ils vont rencontrer les lymphocytes T activés de même spécificité antigénique. L'interaction entre les lymphocytes B et T permet de fournir les signaux de co-stimulation nécessaires à la complète activation des cellules B. Les lymphocytes T deviennent alors des cellules TFH ou cellules T « follicular helper » qui migreront depuis la zone inter-folliculaire vers le centre des follicules.

A ce stade les cellules B activées peuvent avoir plusieurs devenir et se différencier : (1) en cellules à courte durée de vie sécrétrices d'anticorps de faible affinité, (2) ou directement en cellules B mémoires, (3) ou migrer vers le centre des follicules pour former des cellules B de centre germinatif (CG) (Jacob and Kelsoe, 1992; Kurosaki et al., 2015; Taylor et al., 2012) (**Figure 9b**).

Suite à la prolifération des cellules B activées au centre des follicules, les lymphocytes B au repos se trouvent refoulés en périphérie du CG et forment la zone du manteau. Le CG va évoluer et passer du stade précoce au stade de CG mature polarisé présentant 2 zones distinctes : la sombre (ZS) et la zone claire (ZC). Les cellules B de ces 2 zones peuvent être différenciées sur la base de l'expression de marqueurs caractéristiques : CXCR4<sup>high</sup>/CD83<sup>low</sup>/CD86<sup>low</sup> (cellules B de la ZS) et CXCR4<sup>low</sup>/CD83<sup>high</sup>/CD86<sup>high</sup> (cellules B de la ZC) (Allen et al., 2004; Victora et al., 2010). Les techniques de coloration histologique permettent de distinguer nettement les 2 zones composant les CG en fonction de la densité cellulaire : très forte dans la ZS et plus faible dans la ZC.

Les cellules B de la ZS sont caractérisées par la perte de l'expression d'Ig membranaire et sont appelées centroblastes. Ces cellules prolifèrent activement et subissent des modifications de leur gènes d'Ig réarrangés dans le but d'améliorer l'affinité des Ig : c'est le processus de maturation d'affinité ou d'hypermutation somatique (SHM). Ce processus consiste en l'introduction de mutations ponctuelles au niveau des régions variables des chaînes lourdes et légères d'Ig, constituant le site de reconnaissance à l'antigène. Les mutations accumulées sont préférentiellement observées au niveau des régions dites hypervariables ou CDR porteuses de la spécificité antigénique. Ces mutations vont conduire à des changements d'acides aminés dans la séquence protéique des chaînes lourdes et légères et sont donc responsables de la production de nouvelles molécules d'Ig dont l'affinité a été modifiée.

Après leur passage dans la ZS et une étape de SHM, les cellules B migrent vers la ZC du centre germinatif et expriment de nouveau une Ig de surface, ces cellules sont appelées centrocytes. Ces centrocytes se trouvent à proximité d'autres cellules entrant dans la composition de la ZC telles que les cellules TFH et les CFD et vont alors subir une sélection basée sur la capacité de l'Ig constituant leur BCR (potentiellement muté) à



reconnaître l'antigène. Les lymphocytes B possédant un BCR de meilleure affinité sont capables de capturer efficacement l'antigène présenté par les CFD.

Les cellules B pourront alors se comporter comme des CPA et à leur tour présenter l'antigène, en association avec le complexe CMH de classe II (complexe majeur d'histocompatibilité), aux lymphocytes T. Cette interaction B-T fournie aux lymphocytes B des signaux de survie essentiels à la poursuite de leur développement.

A l'inverse les cellules B exprimant un BCR de plus faible affinité sont dans l'incapacité de présenter efficacement l'antigène et ne recevront pas les signaux nécessaires à leur survie. Dans ce cas précis, les cellules sont éliminées par apoptose.

Certaines cellules positivement sélectionnées peuvent circuler à nouveau vers la ZS où elles subissent un nouveau cycle de prolifération et d'hypermutation afin d'augmenter davantage l'affinité du récepteur à l'antigène (Victoria et al., 2010). Ainsi les CG sont constitués en majorité de cellules B exprimant un récepteur à l'antigène de très haute affinité qui seront à l'origine des plasmocytes et des cellules B mémoires (De Silva and Klein, 2015). C'est au sein de la ZC que les lymphocytes B vont subir le processus de recombinaison de classe (CSR), permettant à la cellule B d'exprimer une Ig d'une nouvelle classe (IgM vers IgG/IgE/IgA).

L'accessibilité au locus *IgH* durant les processus d'hypermutation et de recombinaison de classe sera détaillée dans le chapitre suivant (chapitre III. : « AID : enzyme clé du développement B tardif. »).

#### **II.2.3.4. Facteurs essentiels à la régulation du centre germinatif.**

Le facteur BCL6 (B cell lymphoma 6) est un répresseur transcriptionnel fortement exprimé dans les cellules B du CG (Allman et al., 1996; Cattoretti et al., 1995). Des études menées à partir de modèles murins déficients pour ce facteur, montrent que BCL6 est nécessaire à la migration des cellules B à l'intérieur des follicules (Huang et al., 2013; Kitano et al., 2011). De plus, BCL6 est capable de moduler différents processus cellulaires dans le but de favoriser un environnement propice aux événements de recombinaison associés au développement des CG (Klein and Dalla-Favera, 2008). L'action de BCL6 confère une tolérance vis-à-vis des dommages causés à l'ADN induits dans les cellules B du CG, tout en assurant une certaine vigilance en maintenant ces cellules B du CG dans un état pro-apoptotique. Pour exemples, BCL6 inhibe l'expression de protéines impliquées dans la réponse DDR (DNA damage response) telles que p53 et ATR (Phan and Dalla-Favera, 2004; Ranuncolo et al., 2007). BCL6 est également responsable de l'inhibition de l'expression de protéines anti-apoptotiques telle que BCL2 (Saito et al., 2009). Par ailleurs, en modulant l'expression de gènes impliqués dans l'activation complète des cellules B, tel que CD80, BCL6 semble réguler l'interaction B-T et empêcher une activation prématurée des cellules B (Niu et al., 2003).

Une autre cible de BCL6 est la protéine BLIMP1 (B lymphocyte induced maturation protein 1). En effet, BCL6 est capable de réprimer l'expression du gène *Prdm1* (positive regulatory domain containing 1) codant BLIMP1 nécessaire à la différenciation plasmocytaire (Shaffer et al., 2000). Ainsi BCL6 participe au maintien de



l'identité des cellules B du CG. Réciproquement, BLIMP1 est capable d'inhiber l'expression de BCL6 (Calame et al., 2003; Crotty et al., 2010).

D'autres facteurs sont capables de moduler l'expression de BCL6, tels que MEF2B ou IRF4. Dans les cellules B activées, l'expression de MEF2B précède l'expression de BCL6 (Ying et al., 2013). Quant au facteur IRF4, il peut à la fois réguler positivement et négativement la transcription du gène codant BCL6 (Ochiai et al., 2013; Saito et al., 2007). Une étude plus récente met en évidence l'importance du facteur mTORC1 dans le développement des CG, et montre notamment que en absence de mTORC1 l'expression de BCL6 est réduite (Raybuck et al., 2018).

Le facteur de transcription C-MYC est également impliqué dans la formation des CG. C-MYC est exprimé de façon précoce lors de l'initiation des CG, puis dans certaines cellules de la ZC destinées à réentrer dans la ZS (Calado et al., 2012; Dominguez-Sola et al., 2012).

Par ailleurs, une forte expression de l'histone méthyl-transférase EZH2 est observée dans les cellules B de CG. Cette protéine est impliquée dans la triméthylation de la lysine 27 de l'histone H3 : marque épigénétique associée à un état « fermé » (répressif) de la chromatine. Des études montrent que EZH2 est nécessaire à la formation des CG, notamment en favorisant la prolifération cellulaire via la répression d'inhibiteurs de CDK (cyclin dependent kinase) tel que l'inhibiteur CDKN1a (Béguelin et al., 2013, 2017).

Les dernières étapes de maturation des cellules B permettent de générer un pool de cellules B mémoires et plasmocytaires.

#### **II.2.4. Les cellules B mémoires.**

A l'issue de la différenciation B et lors d'une réponse primaire vis-à-vis d'un Ag, des cellules B mémoires sont générées (Kurosaki et al., 2015). Lors d'une nouvelle rencontre avec l'Ag dont elles sont spécifiques (réponse secondaire), ces cellules seront capables de se différencier rapidement en cellules plasmocytaires ou d'entrer à nouveau dans un CG (Dogan et al., 2009). Les cellules B mémoires peuvent être générées suite à une réponse T dépendante à partir de cellules B activées en dehors du CG ou après maturation dans le CG. La balance d'expression entre facteurs anti et pro apoptotiques (tels que MCL1 et PUMA) contrôle la survie des cellules B mémoires (Clybouw et al., 2011). L'interaction T-B (CD40L/CD40) semble être un facteur influençant la différenciation en cellules B mémoires dites « CG indépendante » (engagement du CD40 important) et « CG dépendante » (engagement CD40 moindre) ( Taylor et al., 2012). Des cellules B mémoires peuvent également être générées en réponse à une stimulation T indépendante à partir de cellules B matures de type B1 (Yang et al., 2012).

#### **II.2.5. Les plasmocytes : cellules B sécrétrices d'anticorps.**

Les plasmocytes constituent le stade ultime de différenciation B, ce sont les effecteurs de l'immunité humorale assurant la sécrétion d'anticorps. Les plasmocytes sont caractérisés par l'expression du marqueur CD138 et la

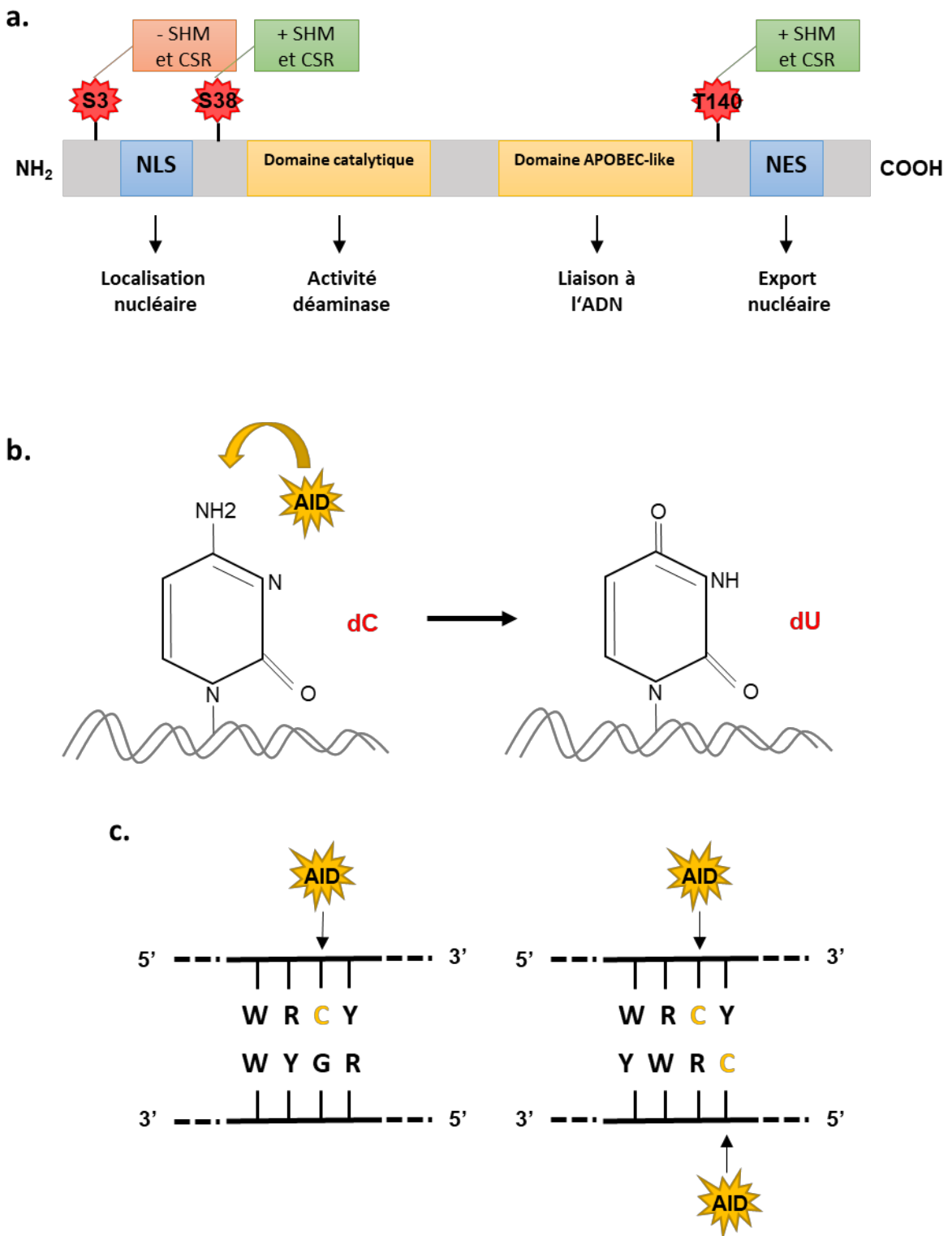




perte d'expression de marqueurs B spécifiques tels que le B220 et le CD19. La protéine BLIMP1 est le facteur clé de la différenciation plasmocytaire (Angelin-Duclos et al., 2000).

Il existe des plasmocytes à courte durée de vie tels que les plasmocytes générés à partir des cellules B1 et B1Z après une activation T-indépendante. Ces plasmocytes sécrètent principalement des IgM peu affins, et sont retrouvés particulièrement dans les cavités pleurale et péritonéale et dans la pulpe rouge de la rate. Les plasmocytes à courte durée de vie sont également générés lors de réponses T dépendantes et de façon rapide pour fournir un 1<sup>er</sup> pool d'anticorps en attendant la maturation de cellules sécrétrices d'anticorps de haute affinité. Il s'agit là de plasmocytes dits extra-folliculaires formés en dehors des CG (Jacob et al., 1991; Kurosaki et al., 2015; Smith et al., 1996).

En parallèle, suite à une activation T-dépendante, les cellules B FO engagées dans le CG sont à l'origine de plasmocytes à longue durée de vie capables de produire des anticorps d'une nouvelle classe et de haute affinité (Nutt et al., 2015). Les plasmocytes issus des CG peuvent également se relocaliser au niveau de la moelle osseuse (phénomène de « homing ») dans des niches spécifiques fournissant les facteurs essentiels à leur survie (Calame et al., 2003; Tangye, 2011).



**Figure 10 : Structure et activité de l'enzyme AID.**

**a.** Structure schématique de l'enzyme AID : principaux domaines et sites de phosphorylation. De l'extrémité N-terminal vers l'extrémité C-terminal, la protéine AID possède un signal de localisation nucléaire (NLS), un domaine catalytique, un domaine APOBEC-like et un signal d'exportation nucléaire (NES).

**b.** Activité de l'enzyme AID : déamination des cytidine de l'ADN en uridine.

**c.** Motifs préférentiels de déamination par AID, dans le cas de la SHM (à gauche) et de la CSR (à droite).

## Chapitre III. AID : enzyme clé du développement B tardif.

---

La diversité du répertoire antigénique repose à la fois sur les mécanismes de recombinaisons V(D)J et sur les mécanismes de SHM et de CSR permettant respectivement la mise en place d'un répertoire primaire puis secondaire (Pavri and Nussenzweig, 2011). L'enzyme AID, à l'origine des processus de SHM et de CSR, est donc un facteur enzymatique clé du développement lymphocytaire B, responsable de la production d'anticorps hautement spécifiques.

En 1999, l'équipe de Honjo met en évidence l'expression de l'enzyme AID dans les cellules B du CG et stimulées *in vitro* (Muramatsu et al., 1999) puis démontre son implication dans les événements de SHM et de CSR (Muramatsu et al., 2000). Simultanément, une étude menée par l'équipe de Durandy, montre que les anomalies observées chez les patients atteints du syndrome d'hyper-IgM sont caractéristiques d'un défaut de fonctionnement de l'enzyme AID (Revy et al., 2000).

### III.1. AID : structure, fonction et régulation.

#### III.1.1. Structure de AID.

L'enzyme AID est codée par le gène *Aicda* situé sur le chromosome 6 murin. AID appartient à la famille protéique des APOBEC (apolipoprotein B mRNA editing catalytic component), sa structure est proche de la protéine APOBEC-1. Les protéines appartenant à cette famille sont capables de déaminer les cytidines provenant d'une structure nucléotidique simple brin mais ont des fonctions distinctes (Conticello et al., 2005; Rebhandl et al., 2015).

AID est une petite protéine (198 acides aminés, 24kDa) caractérisée par plusieurs domaines (**Figure 10a**). De l'extrémité N-terminale vers l'extrémité C-terminale, la protéine AID possède un signal de localisation nucléaire, un domaine catalytique, un domaine APOBEC-like et un signal d'exportation nucléaire. Le domaine catalytique porte la fonction de déamination et le domaine APOBEC-like est impliqué dans la liaison à l'ADN (Ta et al., 2003; Vaidyanathan et al., 2014; Zan and Casali, 2013).

Des études conduites à partir de mutants de AID ont permis de mettre en évidence des rôles différents pour les domaines N- et C-terminaux. Des mutations dans la région N-terminale sont associées à un défaut des processus de SHM et de CSR alors que des mutations dans la région C-terminale ne vont affecter que le processus de CSR (Barreto et al., 2003; Bransteitter et al., 2004; Shinkura et al., 2004; Ta et al., 2003).

#### III.1.2. Fonction de AID.

Par homologie avec les protéines APOBECs, il a d'abord été supposé que AID était une déaminase de l'ARN (Muramatsu et al., 2000). Par la suite, plusieurs études ont montré que l'ADN simple brin est le substrat de l'enzyme AID (Bransteitter et al., 2003; Dickerson et al., 2003; Larijani and Martin, 2007; Pham et al., 2003). Néanmoins, des études menées par l'équipe de Honjo n'excluent pas que l'action de AID puisse s'exercer sur l'ARN. Dans une étude récente cette équipe montre que l'ARN viral du virus HBV peut être déaminé par AID



(Liang et al., 2013). Les études par séquençage haut débit des transcrits (RNA-seq) n'apportent quant à elles aucun élément en faveur du ciblage de l'ARN par AID (Fritz et al., 2013).

En considérant que AID agit sur l'ADN, il semblerait que le ciblage de cet enzyme soit influencé par la présence du groupement hydroxyle en position 2' du deoxyribose composant le deoxyribonucléotide (Nabel et al., 2013). L'enzyme AID va déaminer une cytosine (C) de l'ADN préférentiellement localisées dans des motifs WRCY (W = dA/dT ; R = dA/dG ; Y = dT/dC/dA) (**Figure 10 b et c**) (Álvarez-Prado et al., 2018; Basu et al., 2009; HWANG et al., 2015; Pham et al., 2003). Il en résulte un mésappariement de base dU:dG qui sera pris en charge par les voies de réparation de l'ADN BER (base excision repair) ou MMR (mismatch repair). Ce mésappariement de base, initié par AID, constitue la première étape des processus de SHM et de CSR.

Par ailleurs, AID pourrait intervenir dans la déméthylation de l'ADN en ciblant les 5-méthyl-cytosine (Fritz and Papavasiliou, 2010; Morgan et al., 2004). Dans un modèle d'hétérokaryons, AID favorise la déméthylation de promoteurs de gènes dont l'expression est typiquement associée au statut de cellules souches pluripotentes (tel que *nanog*) (Bhutani et al., 2010). La fonction de AID dans les cellules souches reste controversée : bien que l'induction de cellules souches soit augmentée à partir de fibroblastes surexprimant AID (Bhutani et al., 2013), la capacité à générer des cellules souches demeure cependant identique pour des fibroblastes directement issus d'animaux *wt* ou AID KO (Bhutani et al., 2013; Shimamoto et al., 2014). Une étude plus récente, montre que la perte de AID dans les cellules souches hématopoïétiques est associée à une hyperméthylation de l'ADN et à un défaut de différenciation en défaveur du lignage érythrocytaire (Kunimoto et al., 2017).

Du fait de la forte expression de AID dans les cellules B de CG, la déméthylation de l'ADN a été étudiée au sein de ce type cellulaire avec des résultats discordants. Si une première étude n'indique aucune différence entre les profils de méthylation des cellules B activées *ex vivo* issues de souris sauvages ou déficientes pour AID (Fritz et al., 2013), la seconde étude a montré des modifications du profil de méthylation dans des lymphocytes B de CG *in vivo* et ces modifications ne sont pas retrouvées en l'absence de AID (Dominguez et al., 2015).

### III.1.3. Régulation de l'activité de AID.

Etant donné sa capacité à modifier l'ADN, l'activité de l'enzyme AID doit être strictement régulée : restreinte aux gènes codant les immunoglobulines dans le noyau des lymphocytes B activés. Cette régulation va intervenir à plusieurs niveaux : (1) transcription et post transcription, (2) post traduction et (3) localisation (Keim et al., 2013).

(1) AID est exprimée dans les cellules B activées de CG (Cattoretti et al., 2006). Des études mettent également en évidence que cette enzyme est exprimée plus précocement et à des niveaux très faibles, dans les cellules pré-B de la moelle osseuse. AID pourrait alors être impliquée dans un processus de diversification antigène indépendant des gènes des immunoglobulines (Han et al., 2007). Le gène *Aicda* est caractérisé par la présence de 4 régions régulatrices permettant le contrôle de sa transcription. Ces régions possèdent des sites de fixation



pour des facteurs de transcription B spécifiques et pour des protéines ID (inhibitor differentiation) inhibitrices de la transcription (Gonda et al., 2003; Tran et al., 2010; Vaidyanathan et al., 2014).

L'expression de l'enzyme AID est aussi contrôlée au niveau post transcriptionnel via la liaison de micro-ARNs (ou miRNAs) sur la région 3' non traduite de son ARN messager (ARNm) (de Yébenes and Ramiro, 2010). Plusieurs micro-ARNs, intervenant dans la stabilité de l'ARNm de AID ont été décrits : miR-155, miR-181b, miR-361 et miR-93 (Borchert et al., 2011; Vaidyanathan et al., 2014). Parmi ces micro-ARNs, miR-155 est le plus étudié et semble avoir une importance particulière : miR-155 est surexprimé dans les cellules B stimulées effectuant la commutation de classe (Teng et al., 2008) et joue un rôle dans l'homéostasie cellulaire B en limitant les dommages induits par AID à l'ADN (Bouamar et al., 2015). En l'absence de liaison de miR-155, le niveau d'ARNm et de la protéine AID est augmenté conduisant à un taux élevé de translocation IgH;c-Myc (Dorsett et al., 2008). Par ailleurs, le facteur de transcription BCL6, requis pour la formation des CG, est capable de réprimer l'action de miR-155 (Basso et al., 2012).

Contrairement à miR-155, miR-181b est principalement exprimé dans les cellules B au repos. L'expression ectopique de miR-181b dans les cellules B stimulées effectuant la commutation de classe conduit à une réduction du taux d'ARNm de AID et à un défaut de ce processus (de Yébenes et al., 2008).

(2) La régulation de l'activité de l'enzyme AID va aussi s'effectuer au niveau post traductionnel. En effet, AID possède plusieurs sites de phosphorylation : sur les résidus sérine en position 3 et 38 (S3, S38), thréonine en position 27 et 140 (T27, T140) et tyrosine en position 184 (Y184) (Pasqualucci et al., 2006; Vaidyanathan et al., 2014; Vuong et al., 2009). L'impact des phosphorylations S3 et S38 et T140, sur le fonctionnement de AID a été plus largement étudié (**Figure 10a**).

La phosphorylation S3 est sous la dépendance de la protéine kinase PKC et a un rôle inhibiteur de l'activité de AID. La mutation de ce résidu va entraîner une augmentation des processus de CSR et de SHM (Gazumyan et al., 2011). La protéine kinase C est également responsable de la phosphorylation T140 et la mutation de ce résidu affecte plus particulièrement le processus de SHM (McBride et al., 2008). Enfin, la phosphorylation S38, par la protéine kinase PKA, augmente l'activité de AID et la mutation de ce résidu entraîne un défaut de SHM et de CSR (McBride et al., 2006). Il a été également montré que la phosphorylation S38 est essentielle pour l'interaction entre AID et un de ses partenaires : la protéine RPA (protéine de réplication A) (Basu et al., 2005) (Cheng et al., 2009). Enfin, une étude récente montre que la structure tridimensionnelle de la poche catalytique de l'enzyme AID semble être impliquée dans la restriction de son activité (King et al., 2015).

(3) Le contrôle de la localisation de AID intervient également dans la régulation de son activité. AID est principalement localisée dans le cytoplasme des cellules (>90%), et doit être importée dans le noyau pour exercer son activité de déamination sur l'ADN (Ito et al., 2004; Rada et al., 2002a). La localisation de AID dans le noyau est temporaire : l'enzyme est rapidement exportée vers le cytoplasme et peut être prise en charge par des mécanismes de dégradation par le protéasome. La navette de AID entre les compartiments





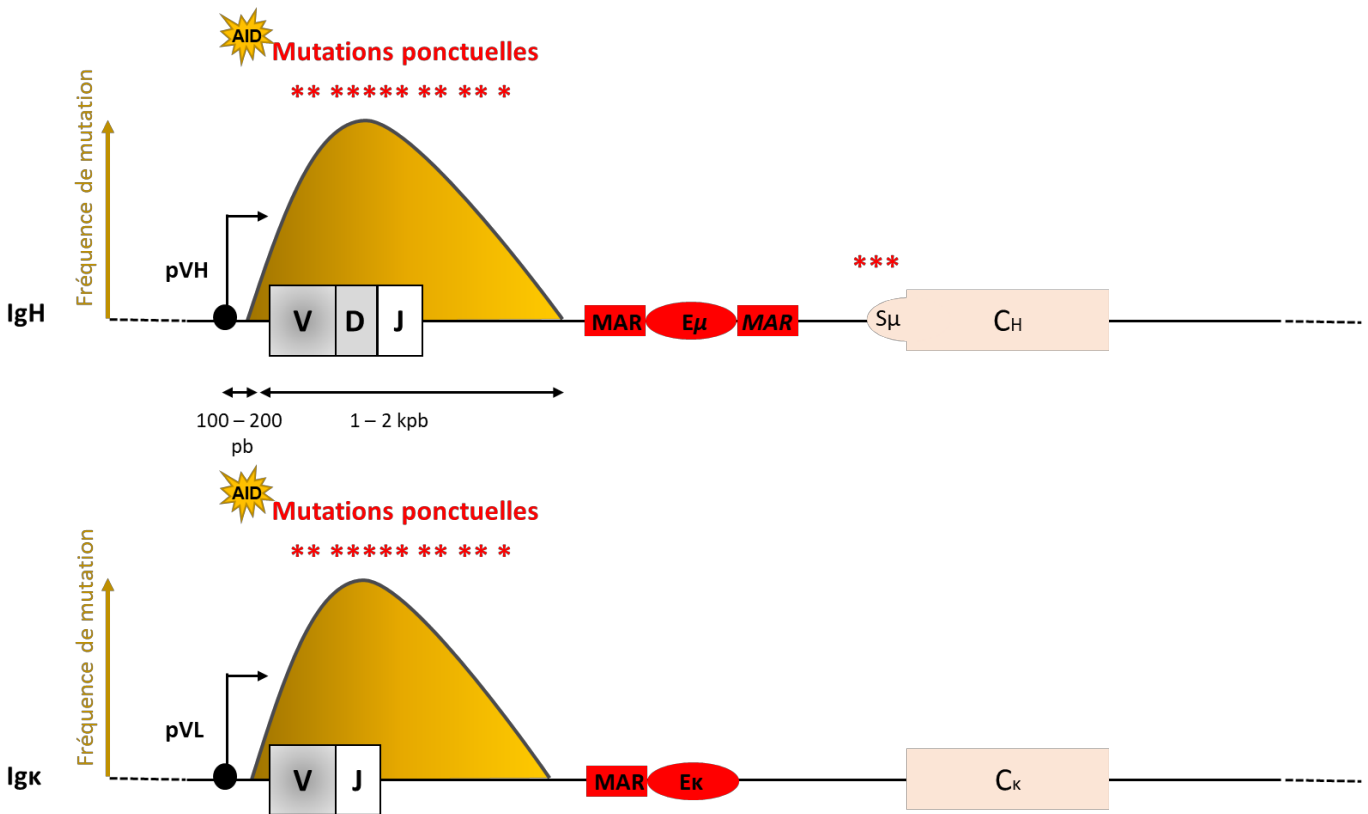
cytoplasmique et nucléaire repose sur un mécanisme actif faisant intervenir les séquences signal de localisation nucléaire (NLS) et d'export nucléaire (NES) ainsi que des protéines de type importine-3 et exportine CRM1 (Orthwein and Di Noia, 2012; Patenaude et al., 2009). Deux voies de dégradation de l'enzyme AID via le protéasome ont été décrites : une voie impliquant une étape de poly-ubiquitination de AID dans le noyau et une voie indépendante de l'étape de poly-ubiquitination (Aoufouchi et al., 2008; Uchimura et al., 2011).

### **III.2. Recrutement de AID aux loci des gènes d'immunoglobulines.**

Le caractère mutagène de AID impose un recrutement ciblé de l'enzyme vers les loci codant les immunoglobulines. Ce recrutement précis de AID sur les structures d'ADN simple brin est associé à l'étape de transcription (1) et dépend également des caractéristiques propres aux séquences ciblées (2).

(1) Le lien entre ciblage par AID et transcription a été montré à la fin des années 1990, avant même la découverte de l'enzyme AID. De nombreuses études montrent que la transcription est nécessaire au processus de SHM : le nombre de mutation diminue lorsque le promoteur en amont de la région variable est supprimé (Fukita et al., 1998) et le déplacement du promoteur induit également un déplacement du spectre de mutation (Tumas-Brundage and Manser, 1997). Plusieurs équipes se sont efforcées de mettre en évidence des interactions entre AID et d'éventuels partenaires parmi les acteurs composant la machinerie de transcription. Le complexe de l'ARN polymérase II (ou ARN pol II) et le facteur SPT5 (suppressor of ty5 homolog) ont été identifiés comme étant des partenaires de AID : l'ARN pol II facilite le recrutement de AID vers l'ADN en cours de transcription et SPT5 facilite l'association entre AID et l'ARN pol II (Pavri et al., 2010). Les sites de recrutement de AID se superposent avec ceux de la forme dite « en pause » du complexe de l'ARN pol II, phosphorylé au niveau du résidu serine 5 localisé à l'extrémité C terminale de la grande sous unité (Kenter, 2012 ; Nambu et al., 2003; Pavri et al., 2010). Ainsi dans les cellules B de CG, le complexe de l'ARN pol II et SPT5 sont accumulés au niveau des régions variables et switch ciblées par AID (Maul et al., 2014). Les régions constantes au locus *IgH*, non ciblées par AID, ne présentent pas d'accumulation de l'ARN pol II en pause (Wang et al., 2009).

L'étape de transcription permet de fournir à AID son substrat (Storb, 2014) : l'ADN exposé sous forme simple brin (via la formation de « boucles » ou de « patches » lors de la transcription) semble être une cible préférentielle de la machinerie de mutation (Chaudhuri et al., 2003; Ronai et al., 2007; Wang et al., 2014). Le super-enroulement négatif généré à l'arrière de la bulle de transcription favoriserait l'exposition de l'ADN sous forme simple brin (Parsa et al., 2012; Shen and Storb, 2004). La stabilisation de l'ADN sous forme simple brin peut être favorisée par :



**Figure 11 : Caractéristiques générales du processus de SHM.**

Le processus de SHM consiste en l'introduction de mutations ponctuelles dans les régions variables réarrangées des loci des chaînes lourdes et légères d'Ig. L'introduction des mutations débute à environ 120 pb en aval du site d'initiation de la transcription et s'étend sur plus de 1 kpb.

- la composition en base de la séquence, pour exemple les régions switch sont riches en purines et forment des structures de type R-loops au niveau desquelles l'ARN néo-synthétisé est associé au brin d'ADN matrice formant un hybride ADN/ARN et laissant le brin non matrice exposé sous forme simple brin ;

- le recrutement de protéines stabilisatrices comme le complexe RPA (replication protein A). Une étude *in vitro* montrant une association de AID et du complexe RPA propose un rôle stabilisateur de RPA sur l'ADN simple brin lors de la transcription (Chaudhuri et al., 2004). Il semblerait que le recrutement de RPA suit celui de AID (Yamane et al., 2011).

Bien que le brin non matrice ait été proposé comme étant immédiatement accessible à AID car exposé lors de la transcription (Chaudhuri et al. 2003) ; il a cependant été montré que les 2 brins (matrice et non matrice) sont efficacement ciblés par AID (Milstein et al., 1998; Shen and Storb, 2004). Plusieurs travaux tentent d'expliquer l'accessibilité du brin matrice à AID dans le contexte de la transcription (Haddad et al., 2011; Roa et al., 2008a) :

- la transcription « antisens » est une hypothèse plausible car plusieurs sites d'initiation de la transcription antisens ont été décrits au locus *IgH* dans les régions ciblées par AID (en aval des exons variables réarrangés et en aval des régions switch) (**Figure 6**) (Perlot et al., 2008). La découverte, par l'équipe de Alt, du recrutement préférentiel de AID dans des régions présentant une transcription convergente constitue un nouvel argument en faveur de la transcription des deux brins d'ADN (Meng et al., 2015).

- une autre hypothèse pouvant expliquer l'accessibilité du brin matrice (sous forme simple brin) est la dégradation et/ou le déplacement, par le complexe du RNA exosome, de la molécule d'ARN qui lui est associée (Pefanis and Basu, 2015; Pefanis et al., 2014). Le complexe du RNA exosome reconnaît un substrat ARN par son extrémité 3' libre qui devient accessible lorsque le complexe de l'ARN pol II en pause fait « marche arrière » (phénomène de backtracking) (Basu et al., 2011; Cogné, 2013) ou après une terminaison prématurée de la transcription (Wang et al., 2014). Une étude plus récente (à laquelle j'ai contribué) révèle une coopération entre le RNA exosome et des hélicases à ARN (telle que MTR4) dans le but de promouvoir une distribution symétrique des mutations (sur les 2 brins) (Lim et al., 2017).

(2) Au-delà de la nécessité de la transcription, le recrutement de AID dépend également des caractéristiques propres aux séquences ciblées. Outre la présence des « hotspots » de AID, des études montrent que des séquences mutées par AID possèdent des sites de fixation pour certains facteurs de transcription, notamment, des sites E-box liant les facteurs E2A. Des études dans des modèles transgéniques proposent que la seule présence des E-box soit suffisante pour induire des mutations (Michael et al., 2003; Tanaka et al., 2010). Cependant, le lien entre la présence de motifs de liaison pour certains facteurs de transcription et le ciblage par AID doit être considéré avec prudence, de surcroît avec la description de nouvelles cibles de AID (cibles dites « AID off targets » qui seront décrites dans un prochain paragraphe) (Staszewski et al., 2011).

<b>Paramètres</b>	<b>SHM</b>	<b>CSR</b>
Cible	Régions variables Chaînes lourde + légère	Régions switch Chaîne lourde
Structures associées à la transcription	Boucles de transcription + transcription antisens ?	R-Loop + G quadruplex + transcription antisens ?
Activité de AID	Cassure simple brin - mutation	Cassure double brin

**Table 1 : Principales différences entre les processus de SHM et de CSR.**

Le ciblage de AID reste un processus complexe et mal défini à l'heure actuelle. Au-delà de l'indispensable transcription des motifs *cis* préférentiels, s'ajoute une coopération entre différents complexes protéiques partenaires de AID (complexe de l'ARN pol II, du RNA exosome...). Bien que la transcription soit une étape commune pour l'initiation de la SHM et de la CSR par AID, les régions cibles de ces deux mécanismes possèdent des spécificités qui leur sont propres et qui seront abordées dans les prochains paragraphes (**table 1**).

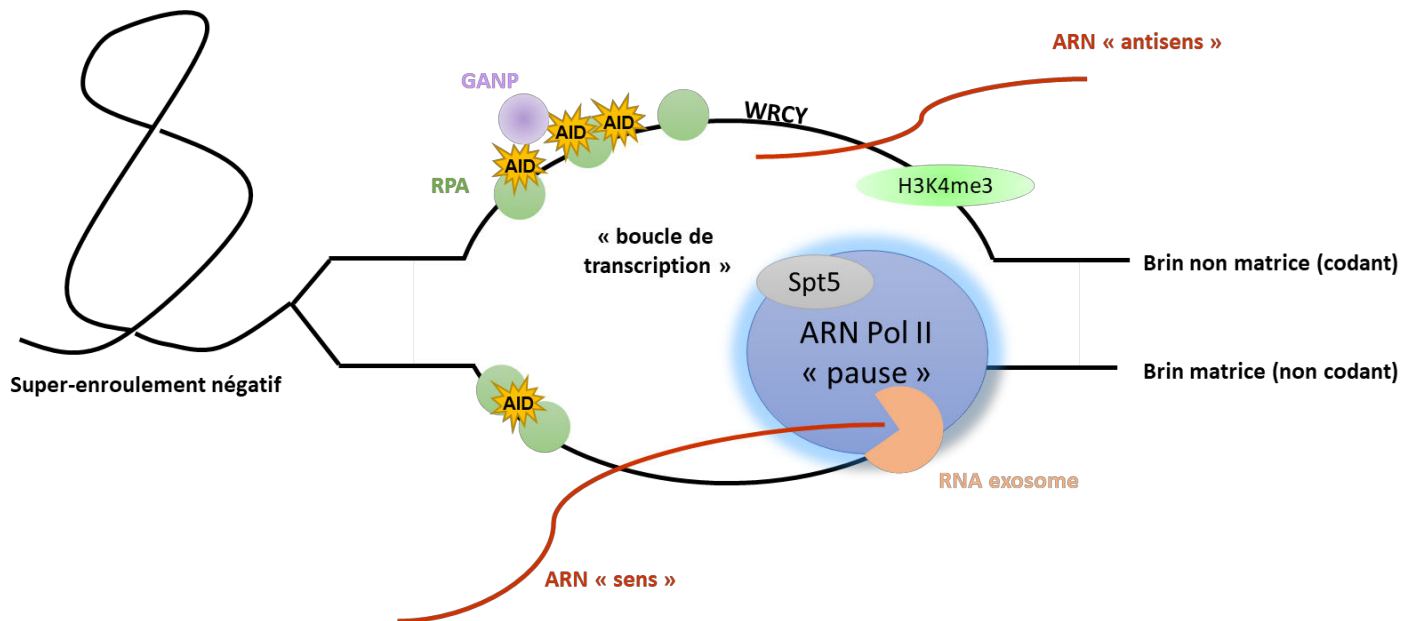
### **III.2.1. Ciblage de AID lors de la SHM et mécanisme moléculaire de la SHM.**

#### **III.2.1.1. Ciblage de AID lors de la SHM.**

Lors du processus de SHM, AID va initier l'introduction de mutations ponctuelles dans les régions variables des chaînes lourdes et légères des Ig (IgV). L'introduction des mutations débute environ 120pb en aval du site d'initiation de la transcription et s'étend sur plus de 1kb (**Figure 11**). Le paragraphe précédent souligne l'importance de la transcription pour le recrutement de AID sur ses cibles ; dans le cas de la SHM la transcription est initiée aux promoteurs situés en amont de chaque segment V réarrangé ( $V_H$  ou  $V_L$ ). Des structures secondaires adoptées lors de la transcription de la région variable semblent être en faveur d'un ciblage par AID (Wright et al., 2008). Cependant, la transcription à elle seule ne suffit pas à induire un taux optimal de mutation. Pour exemple, le remplacement du promoteur ( $pV_L$ ) au locus *IgL* dans le modèle DT40 (cellules B de poulet) par un promoteur fort capable d'induire la transcription au locus *IgL* n'est pas suffisant pour supporter le processus de mutation (Yang et al., 2006).

Des recherches se sont également portées sur l'identification de séquences d'ADN (en dehors des régions activatrices de la transcription) présentes dans les gènes d'Ig qui pourraient favoriser le recrutement de la machinerie de mutation. De telles séquences nommées MEE (pour mutational enhancer element) ou DIVAC (pour diversification activator) ont alors été identifiées au locus *IgL* dans les cellules B DT40 (Batrak et al., 2011; Blagodatski et al., 2009; Kothapalli et al., 2011). Le mode d'action des DIVAC reste à définir, mais leur présence semble associée à une augmentation du recrutement du complexe de l'ARN pol II en pause (Kohler et al., 2012). Il n'est pas exclu que de tels éléments puissent recruter des cofacteurs protéiques (tels que PU.1 et IRF4) favorisant l'activité de la machinerie de mutation (Luo and Tian, 2010).

L'action de AID sur les régions variables réarrangées semble dépendre de la coopération avec ses partenaires, notamment RPA et GANP (germinal center associated protein). En effet, le complexe AID/RPA se lie à l'ADN simple brin exposé par les bulles de transcription préférentiellement sur les motifs « hotspots » de la SHM (Chaudhuri et al., 2004) ; la protéine GANT semble quant à elle impliquée à la fois dans le transport de AID vers le noyau et dans le ciblage de AID vers les régions IgV activement transcrites (Maeda et al., 2010). Si l'ensemble des facteurs associés au recrutement de AID sur les régions IgV ne sont pas connus ; plusieurs études récentes proposent un modèle pour le ciblage de AID lors de la SHM. Dans ce modèle, l'accumulation de l'ARN pol II en pause favoriserait dans un premier temps l'exposition du brin non matrice à la machinerie



**Figure 12 : Ciblage de AID vers les régions d'intérêt lors du processus de SHM.**

Le ciblage de AID est un processus complexe multifactoriel lié à la transcription et nécessitant une coopération étroite entre différents éléments schématisés ici : ADN sous forme simple brin, partenaire de AID, ARN pol II en pause, RNA exosome, séquences en *cis* consensus ...

de mutations, alors que le brin matrice (inclus dans l'hybride ADN/ARN) serait protégé. Dans un second temps, l'accumulation d'effecteurs protéiques sur le brin non matrice protégerait ce dernier de « l'attaque » de AID alors que le brin matrice serait à son tour exposé à AID suite à la dissociation de l'hybride ADN/ARN par le complexe du RNA exosome (Basu et al., 2011; Lim et al., 2017; Pefanis and Basu, 2015; Wang et al., 2014). Dans ce modèle, le rôle éventuel de la transcription antisens n'a pas été pris en compte.

Par ailleurs, des études se sont intéressées au potentiel rôle des topoisomérases I (TOPI) dans le processus de SHM. La topoisomérase I est une enzyme qui relâche les contraintes d'enroulement de l'ADN rencontrées lors de la transcription. TOPI génère une cassure transitoire sur l'un des 2 brins de l'ADN ce qui permet de relâcher la torsion associée à l'avancement du complexe de l'ARN pol (Baranello et al., 2013; Pommier et al., 2016). La diminution de l'expression de TOPI (induit par siRNA, dans la lignée BL2) entraîne une augmentation de SHM (Kobayashi et al., 2011). Cette étude suggère que l'absence ou la réduction de l'activité de TOPI favoriserait l'action de AID pour la SHM : le mécanisme impliqué pourrait être une augmentation de l'accessibilité aux mutations via le maintien de structures d'ADN simple brin en aval du complexe de transcription. Bien qu'une 1<sup>ère</sup> étude avait proposé que AID puisse induire une diminution de l'expression de TOPI dans le but de promouvoir l'accessibilité à la machinerie de mutation, d'autres études ont cependant montré que le niveau d'expression de TOPI est identique entre des cellules B issues de souris *wt* ou déficientes pour AID (Kobayashi et al., 2009; Maul et al., 2015). La régulation de l'expression de TOPI par AID reste controversée mais il est établi que l'augmentation des mutations observée suite à une diminution de l'expression de TOPI résulte de l'accumulation du complexe de l'ARN pol II favorisant le recrutement de AID (Maul et al., 2015).

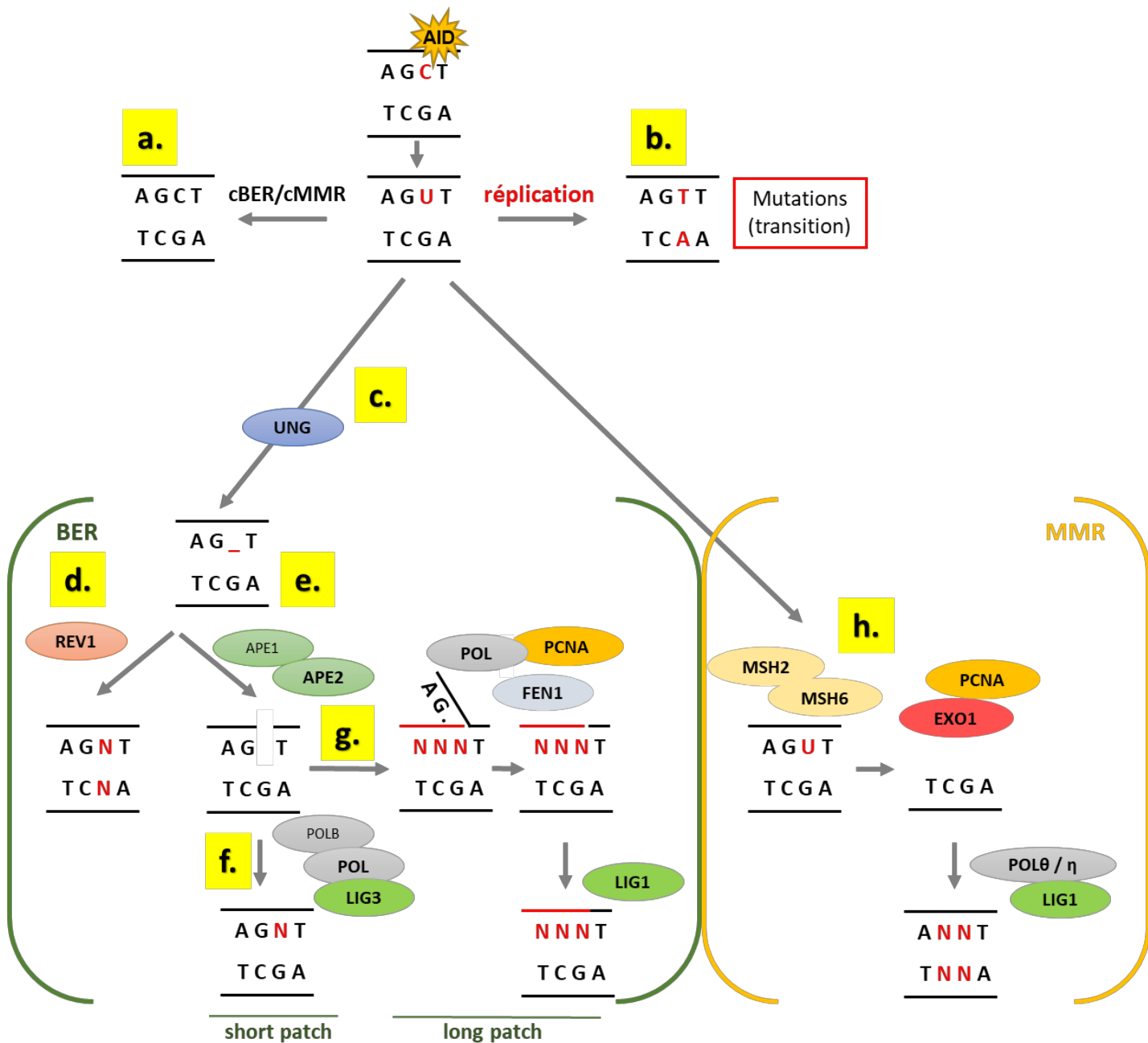
Des modifications épigénétiques ont également été décrites comme associées au processus de maturation d'affinité (Sheppard et al., 2018; Wu et al., 2018). Bien que non spécifiques des régions hypermutées, la présence de la marque H3K4me3 semble favoriser le bon déroulement du processus de SHM : la diminution de H3K4me3 corrèle avec une diminution de la fréquence de mutation (Begum et al., 2012). Les régions ciblées par la SHM sont également enrichies en histones hyperacétylées (Woo et al., 2003). De plus, dans les cellules B de CG, l'ADN se trouve préférentiellement sous forme hypométhylée (Shaknovich et al., 2011). Dans un modèle de souris portant 2 allèles *Igκ* préalablement réarrangés : seul l'allèle hypométhylé est muté (Fraenkel et al., 2007). Le variant d'histone H3.3 semble également favoriser l'exposition de l'ADN simple brin au sein des bulles de transcription et pourrait promouvoir le ciblage par AID (Aida et al., 2013; Romanello et al., 2016).

La **figure 12** reprend quelques éléments essentiels impliqués dans le ciblage de AID lors de la SHM.

### **III.2.1.2. Mécanisme moléculaire de la SHM.**

Le processus de SHM se déroule en 2 phases successives :

- (1) déamination d'une cytosine de l'ADN par AID conduisant à un mésappariement de base dU:dG ;
- (2) prise en charge du mésappariement par les voies de réparation de l'ADN.



**Figure 13 : Devenir du mésappariement initié par AID lors du processus de SHM.**

- a.** Les voies conventionnelles BER et MMR corrigent le mésappariement de façon fidèle.
- b.** L'étape de réplication va générer des mutations de type transition.
- c.** La voie BER (non conventionnelle) fait intervenir l'enzyme UNG qui excise la base dU, créant un site abasique. **d.** Des polymérase infidèles (notamment REV1) vont combler le site abasique ou, **e.** celui-ci peut conduire à l'apparition d'une cassure simple brin *via* l'action des endonucléases APE. La cassure est résolue selon la voie « short patch » (agissant sur un seul nucléotide **f.**) ou « long patch » (qui implique le retrait et remplacement de 2 à 13 nucléotides **g.**).
- h.** Lors d'une prise en charge par la voie MMR, interviennent le complexe MSH2/MSH6 et l'exonucléase EXO1 qui excise une portion du brin contenant le mésappariement. Des polymérase infidèles synthétisent alors un nouveau brin.



Les mutations introduites sont la résultante de la prise en charge du mésappariement. Avant même la découverte de AID, l'étude des modèles mutants déficients pour les voies de réparation de l'ADN proposait dès lors un mécanisme de SHM en 2 phases distinctes (Rada et al., 1998).

Les deux voies de réparation de l'ADN interviennent à la suite de la déamination des cytidines de l'ADN par AID: la voie BER (base excision repair) et la voie MMR (mismatch repair). Ces 2 voies de réparation ne sont pas spécifiques des lymphocytes B et réparent fidèlement les dommages causés à l'ADN dans la majorité des types cellulaires. Cependant, dans les lymphocytes B activés de CG qui subissent le processus de SHM, ces voies de réparation sont volontairement infidèles et introduisent des mutations à une forte fréquence (jusqu'à 1 mutation pour 100 pb) dans le but d'augmenter la spécificité des Ig pour l'Ag d'intérêt. Il existe des voies de réparation BER et MMR dites conventionnelles (fidèles) impliquées dans le maintien de l'intégrité de l'ADN et les voies non conventionnelles (infidèles) nécessaires pour le déroulement du processus de maturation d'affinité des Ig.

Lors du processus de SHM, la seule base déaminée et/ou les bases voisines peuvent être modifiées selon la voie de réparation impliqué et selon le type de polymérase infidèle recrutée (**Figure 13**). Le mésappariement peut alors avoir plusieurs devenir :

- ce dernier est pris en charge par les voies conventionnelles de réparation : aucune mutation ne sera générée (**Figure 13a**) ;
- le mésappariement n'est pas pris en charge avant l'étape de réplication de l'ADN : l'ADN polymérase considère le dU comme un dT générant ainsi une mutation de type transition dans l'une des 2 cellules filles (dC vers dT sur le brin déaminé et dG vers dA sur le brin complémentaire) (**Figure 13b**) (Neuberger et al., 2005; Peled et al., 2008).
- le mésappariement peut être pris en charge par les voies de réparation BER ou MMR.

La voie BER fait intervenir une DNA glycosylase dont la fonction est d'exciser la base mutée créant ainsi un site abasique. L'enzyme UNG (uracil DNA glycosylase) va éliminer le dU initié par AID (**Figure 13c**). Des polymérases infidèles vont intervenir pour combler le site abasique. Il en résulte des mutations de type transition ou transversion aux bases C/G (**Figure 13d**) (Bahjat and Guikema, 2017). L'inactivation de l'enzyme UNG entraîne un défaut du profil de mutation associé à une quasi absence de transversions depuis les bases G/C et une nette augmentation des transitions depuis les bases G/C (Di Noia and Neuberger, 2002; Rada et al., 2002b; Shen et al., 2006).

Le site abasique généré par l'action de UNG peut être reconnu par des endonucléases APE1 ou APE2 (apuriniques/aprimidiques endonucléase) qui clivent l'ADN en 5' du site abasique créant ainsi une cassure simple brin (**Figure 13e**) (HWANG et al., 2015; Masani et al., 2013). Les cellules B de CG n'expriment que très peu APE1 par rapport à APE2 et l'intervention de APE2 promeut davantage le recrutement de polymérases infidèles (Stavnezer et al., 2014).



La résolution de la cassure simple brin peut se faire selon 2 mécanismes : le « short patch repair » (action au niveau d'une seule base) ou le « long patch repair » (impliquant le retrait de 2 à 13 nucléotides et leur remplacement) (Almeida and Sobol, 2007; Fortini and Dogliotti, 2007). Lors du « short patch repair », une ADN polymérase sera recrutée pour combler le site de cassure simple brin qui sera ensuite ligaturé par des ADN ligases III (**Figure 13f**). Lors du « long patch repair », un brin d'ADN va être néosynthétisé par les DNA polymérases infidèles associées notamment à la protéine PCNA (proliferating cell nuclear antigen) (Fortini et al., 1998). L'endonucléase FEN1 (flap endonuclease 1) va exciser le brin d'ADN portant le site apurinique, les ADN polymérases assureront le comblement puis des ADN ligases I vont assurer la ligature du brin néosynthétisé (**Figure 13g**) (Fortini and Dogliotti, 2007; Kim and Wilson, 2012; Krokan and Bjørås, 2013; Robertson et al., 2009; Stucki et al., 1998).

Lorsque la voie de réparation MMR est impliquée, le mésappariement est reconnu par l'hétérodimère protéique MSH2/MSH6 (hétérodimère MutS $\alpha$ ) qui va recruter l'hétérodimère protéique MLH1/PMS2 (hétérodimère MutL $\alpha$ ). L'hétérodimère MLH1/PMS2 clive l'ADN en 5' du site déaminé (notamment grâce à l'activité endonucléasique de PMS2) et permet le recrutement de l'exonucléase EXO1 associée à PCNA qui va assurer l'excision d'une portion du brin d'ADN contenant le mésappariement (sur quelques dizaines voire plusieurs centaines de bases). Des ADN polymérases infidèles sont ensuite recrutées et assurent la synthèse du brin excisé, enfin l'ADN ligase I procède à la ligature du brin néosynthétisé (**Figure 13h**) (Jiricny, 2006; Saribasak and Gearhart, 2012). Des études suggèrent que la protéine PCNA sous sa forme mono-ubiquitinylée est responsable du recrutement de polymérases infidèles (Arakawa et al., 2006; Langerak et al., 2007; Roa et al., 2008b). L'inactivation de la protéine MSH2 (ou MSH6) de la voie MMR seule ou combinée à l'inactivation de la protéine UNG de la voie BER provoque une baisse drastique (au moins 80%) des mutations aux bases A/T, les mutations observées sont quasi exclusivement de type transition aux bases C/G (Rada et al., 2004; Shen et al., 2006).

Ainsi les mutations introduites lors de la SHM sont la conséquence d'une prise en charge par des voies de réparation recrutant préférentiellement des polymérases « infidèles ».

Plusieurs études se sont alors efforcées de déterminer quelles polymérases sont impliquées lors de la SHM. Si une polymérase en particulier intervient lors du processus de SHM, alors son inactivation va conduire à une modification du profil de mutation reflétant l'activité de cette enzyme (Weill and Reynaud, 2008).

Pour exemple, le profil de mutation observé chez des animaux déficients pour l'ADN polymérase  $\eta$  est essentiellement caractérisé par des substitutions depuis les bases G et C (85%). L'importante diminution des substitutions depuis les bases A et T est similaire à celle observée dans un contexte déficient pour la voie MMR (Delbos et al., 2005). Par ailleurs, la délétion combinée de l'ADN polymérase  $\eta$  et de MSH2 est associée à une complète absence de mutations depuis les bases A et T, indiquant que les mutations résiduelles aux bases A et T observées en l'absence de MSH2 dépendent de l'ADN polymérase  $\eta$  probablement recruté par la voie du



BER. Ces études soulignent le rôle de la polymérase  $\eta$  dans l'introduction de mutations à partir des bases A/T (Delbos et al., 2005, 2007).

La polymérase REV1 semble aussi intervenir de façon significative lors du processus de SHM et favorise les mutations de type transversion C vers G. Une étude menée dans des modèles murins déficients pour REV1 montre en effet une quasi absence des mutations de type C vers G suggérant que le site abasique généré par UNG constitue un substrat pour REV1 (Jansen et al., 2006).

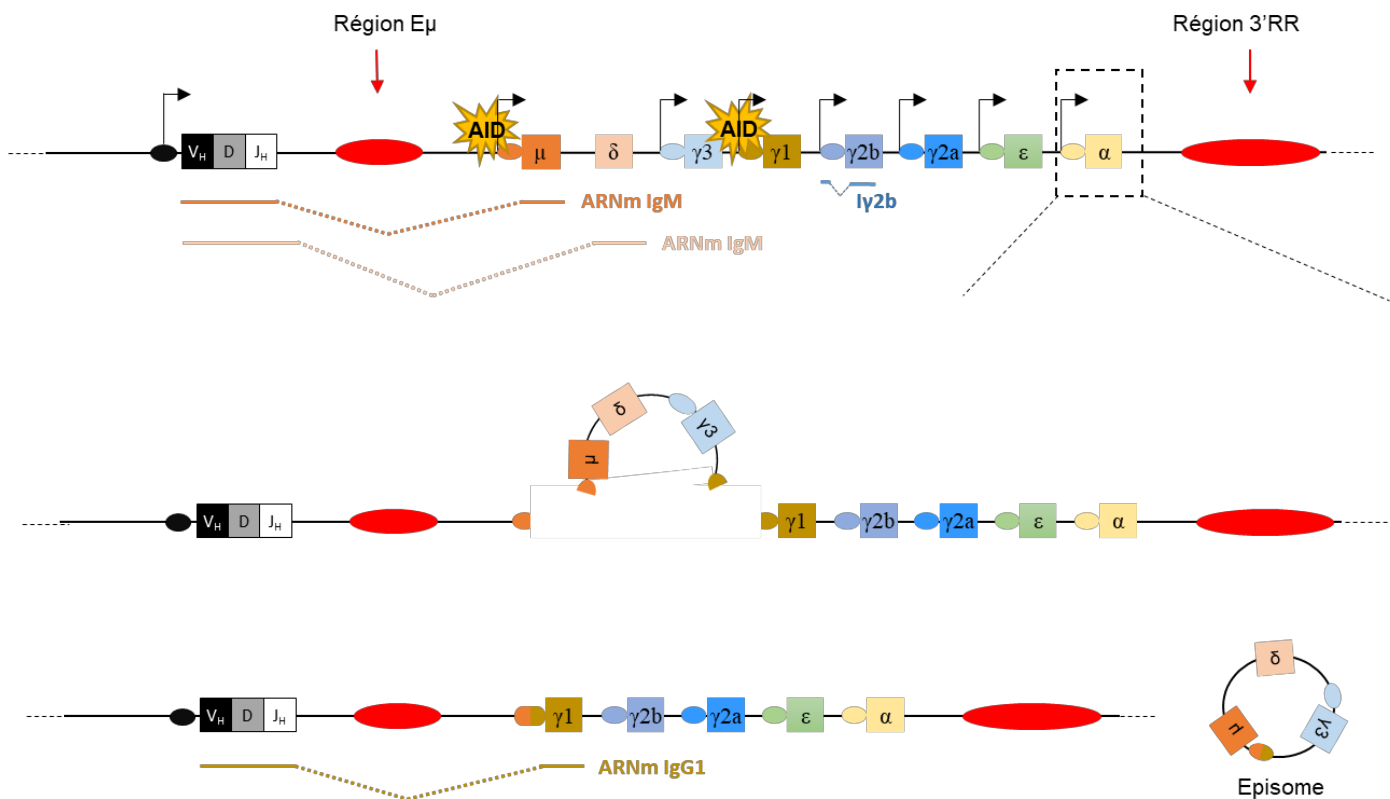
Des études se sont également intéressées au rôle de l'ADN polymérase  $\iota$  puisque cette enzyme semble capable d'interagir avec la polymérase  $\eta$  (McIntyre et al., 2013). Une première étude, basée sur la lignée de souris 129 considérée comme déficiente pour l'ADN poly  $\iota$  (causée par la présence d'un codon stop), semblait invalider son implication de le processus de SHM (McDonald et al., 2003) ; mais une étude approfondie invalide le modèle en montrant, dans cette souche, que la stabilité et non l'activité de l'enzyme pouvait être mise en cause (Aoufouchi et al., 2015). La multiplicité et la redondance des polymérases infidèles rendent leur étude complexe mais leur implication au cours de la SHM demeure très vraisemblable. D'autres études ont en effet montré l'implication des ADN polymérases  $\zeta$  et  $\iota$  dans la génération de mutations en tandem (Maul et al., 2016; Saribasak et al., 2012). A l'inverse, une étude montre que l'expression de l'ADN polymérase  $\beta$  (fidèle) est diminuée dans les cellules subissant la SHM (cellules de la lignée BL2) (Poltoratsky et al., 2007).

Environ 40% des mutations totales sont des mutations aux bases C/G et 60% interviennent aux bases A/T. Les voies non conventionnelles ncBER et ncMMR ne semblent donc pas contribuer de la même façon au processus de SHM. La voie « short patch » BER prend en charge un seul nucléotide et ne peut pas aboutir à l'apparition de mutations aux bases A/T. Les voies « long patch » BER et MMR agissent sur plusieurs nucléotides répartis de part et d'autre du site de déamination et pourront induire des mutations aux bases A/T (Bahjat and Guikema, 2017; HWANG et al., 2015; Peled et al., 2008). Cependant, il n'est pas exclu que les 2 voies de réparation BER et MMR agissent en coopération (Girelli Zubani et al., 2017; Krijger et al., 2009) : la cassure générée par l'action de APE lors du BER pourrait initier le recrutement de la voie MMR (Stavnezer et al., 2014). Les lymphocytes B sont donc capables de détourner les voies de réparation de façon à induire des mutations très fréquentes aux loci des immunoglobulines.

### **III.2.2. Ciblage de AID lors de la CSR et mécanisme moléculaire de la CSR.**

#### **III.2.2.1. Ciblage de AID lors de la CSR.**

La CSR permet à la cellule B d'exprimer à sa surface un nouvel isotype d'immunoglobuline qui diffère de par la nature de la région constante de la chaîne lourde. Ce processus constitue en un évènement de recombinaison entre une région switch « donneuse » (la régions  $S\mu$ ) et une région switch « acceptrice »



**Figure 14 : Schématisation du processus de CSR.**

Les rectangles représentent les différents segments de gènes composant le locus *IgH*. Les 2 régions régulatrices (Ep et 3'RR) au locus *IgH* sont représentées en rouge.

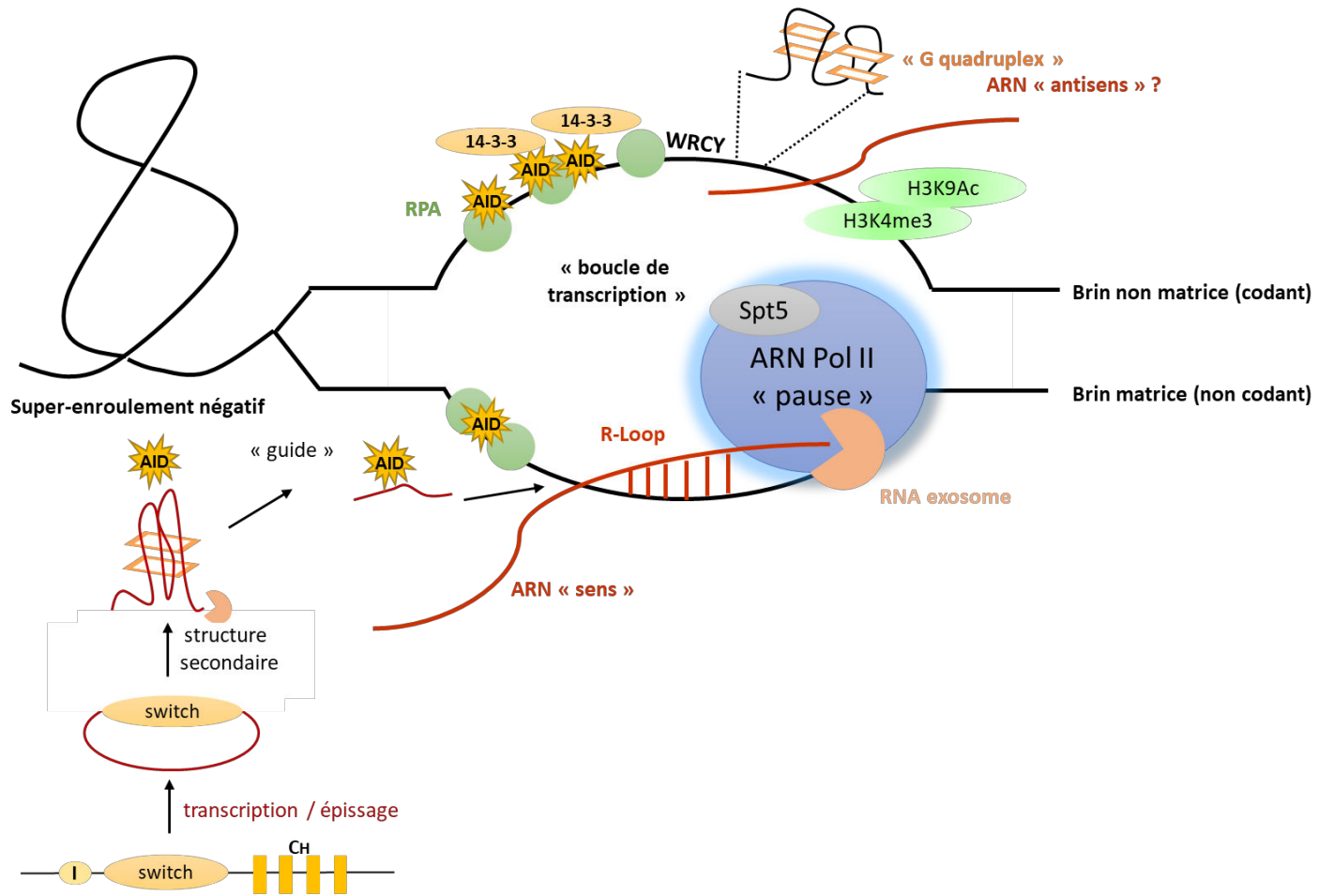
Chaque gène constant (sauf  $C\delta$ ) est précédé d'une région switch, symbolisée par un ovale. Un exemple de l'organisation d'une unité transcriptionnelle constante est schématisé au niveau de  $C\alpha$ .

L'enzyme AID cible les régions switch  $\mu$  (donneuse) et  $\gamma_1$  (acceptrice) induisant des cassures de l'ADN. Ces 2 régions switchs sont ensuite rapprochées puis religuées. La région comprise entre les points de cassure est éliminée sous forme d'un episome.

(correspondant à des régions S localisées en amont de chaque gène constant à l'exception de C $\delta$ ) (Xu et al., 2012). Chaque gène constant (sauf  $\delta$ ) constitue une unité transcriptionnelle comprenant : un promoteur et un exon « I » (intervening), une région S (switch) et les exons C $x$  (C $x$ , x représentant n'importe quel gène constant) (**Figure 14**). Le transcrit généré à partir de cette unité de transcription est épissé entre l'exon I et le premier exon constant, produisant un transcrit de type I $x$  – C $x$  (Chaudhuri and Alt, 2004). La transcription de la région « donneuse », initiée au promoteur I $\mu$ , est considérée comme active de façon constitutive alors que la transcription des autres régions « acceptrices » (I $x$ , correspondant à un autre isotype) sera induite en réponse à une stimulation appropriée. Pour exemple, une stimulation par le lipopolysaccharide bactérien (LPS) active la transcription des promoteurs I $\gamma$ 3 et I $\gamma$ 2b et donc un évènement de CSR vers IgG3 et IgG2b. Le LPS combiné à l'interleukine 4 (IL4) active la transcription des promoteurs I $\gamma$ 1 et I $\gamma$ E et donc un évènement de CSR vers IgG1 et IgE (HWANG et al., 2015; Stavnezer and Schrader, 2014). L'importance des promoteurs et exons I a été démontré il y a plusieurs années : la suppression des exons I $x$  entraîne un blocage de la transcription germinale de l'isotype correspondant empêchant ainsi la recombinaison vers le gène constant (Jung et al., 1993; Zhang et al., 1993). Les transcrits initiés aux promoteurs I sont des transcrits dits germinaux ou stériles : ces derniers ne codent aucune protéine. Cette transcription germinale facilite l'accessibilité des régions cibles (switch) à la machinerie de CSR (Nambu et al., 2003). Un modèle récent propose que les régions introniques épissées pourraient jouer le rôle d'ARN « guide » de AID en formant des structures particulières en « G quadruplex » capables de s'associer avec l'enzyme AID (DiMenna and Chaudhuri, 2016; Zheng et al., 2015). Une étude plus récente montre que les hélicases ARN sont impliquées dans le bon déroulement du processus de CSR, en favorisant la conversion des structures « G quadruplex » en R-Loop dans les régions switch (Ribeiro de Almeida et al., 2018).

L'importance des régions switch (position, nature, taille, etc..) a été mise en évidence au travers de nombreuses études. La délétion des motifs répétés dans S $\mu$  entraîne une diminution de CSR (Luby et al., 2001a), une délétion plus large de la région S $\mu$  réduit drastiquement la CSR vers tous les autres isotypes (Khamlichi et al., 2004; Luby et al., 2001b). Alors que la délétion de S $\gamma$ 1 entraîne l'abolition de la CSR vers l'isotype IgG1 (Shinkura et al., 2003). Les régions switches sont constituées de séquences répétitives enrichies en purines et en motifs « hotspots » de AID sur le brin non matrice ; elles s'étendent sur 1 à 12kpb selon le gène constant associé. Les régions S $\mu$ , S $\alpha$ , S $\epsilon$  sont composées de séquences pentamériques répétées et les régions S $\gamma$ 1, S $\gamma$ 2a et S $\gamma$ 2b et S $\gamma$ 3 contiennent des motifs répétés sur une cinquantaine de base (Chaudhuri and Alt, 2004). De plus, les régions switch contiennent des motifs « AGCT » reconnus par les protéines adaptatrices 14-3-3 impliquées dans le recrutement de AID (Lam et al., 2013; Xu et al., 2010).

L'enrichissement en base G sur le brin non matrice peut également mener à la formation de structure particulière en « G-quadruplex » (Pucella and Chaudhuri, 2017) sur lesquelles AID se lie de façon robuste (en comparaison à un substrat linéaire) (Qiao et al., 2017). Par ailleurs, le caractère répétitif et riche en purines du



**Figure 15 : Ciblage de AID vers les régions d'intérêt lors du processus de CSR.**

Le ciblage de AID est un processus complexe multifactoriel lié à la transcription et nécessitant une coopération étroite entre différents éléments schématisés ici : ADN sous forme simple brin (R-Loops, « G quadruplex »), partenaire de AID, ARN pol II en pause, RNA exosome, séquences en *cis* consensus ...



brin codant des régions switch permet la formation de R-Loops durant la transcription (Yu et al., 2003). L'efficacité d'une région switch à induire la CSR dépend de sa capacité à former des R-Loops (Zhang et al., 2014). Les structures en R-Loops favoriseraient également l'accumulation de l'ARN pol II en pause sur les régions switch.

De la même façon que pour la SHM, l'accessibilité du brin matrice semble être due à l'action du RNA exosome (Laffleur et al., 2017) et pourrait être facilitée par la transcription orientée en antisens (détectée selon un profil comparable à celui de la transcription sens) (Perlot et al., 2008). Cependant l'implication de la transcription antisens reste à démontrer ; une étude montre que la transcription sens est suffisante pour la CSR (Haddad et al., 2011).

Des modifications des histones sont également associées au processus de CSR (Wang et al., 2009). La région  $S_{\mu}$  est enrichie en marques actives de la chromatine (favorisant un état « ouvert ») telles que H3K9 acétylée, H3K27 acétylée, H3K4 tri-méthylée et H3K36 tri-méthylée (Sheppard et al., 2018; Wang et al., 2009; Wu et al., 2018). L'histone H3K9 acétylée et phosphorylée sur le résidu sérine 10 est une marque retrouvée séquentiellement au niveau des régions S donneuses et acceptrices : dans un premier temps un enrichissement est observé sur la région donneuse, puis après activation des cellules B, l'enrichissement s'observe sur la région acceptrice d'intérêt (Li et al., 2013).

La **figure 15** reprend quelques éléments essentiels impliqués dans le ciblage de AID lors de la CSR.

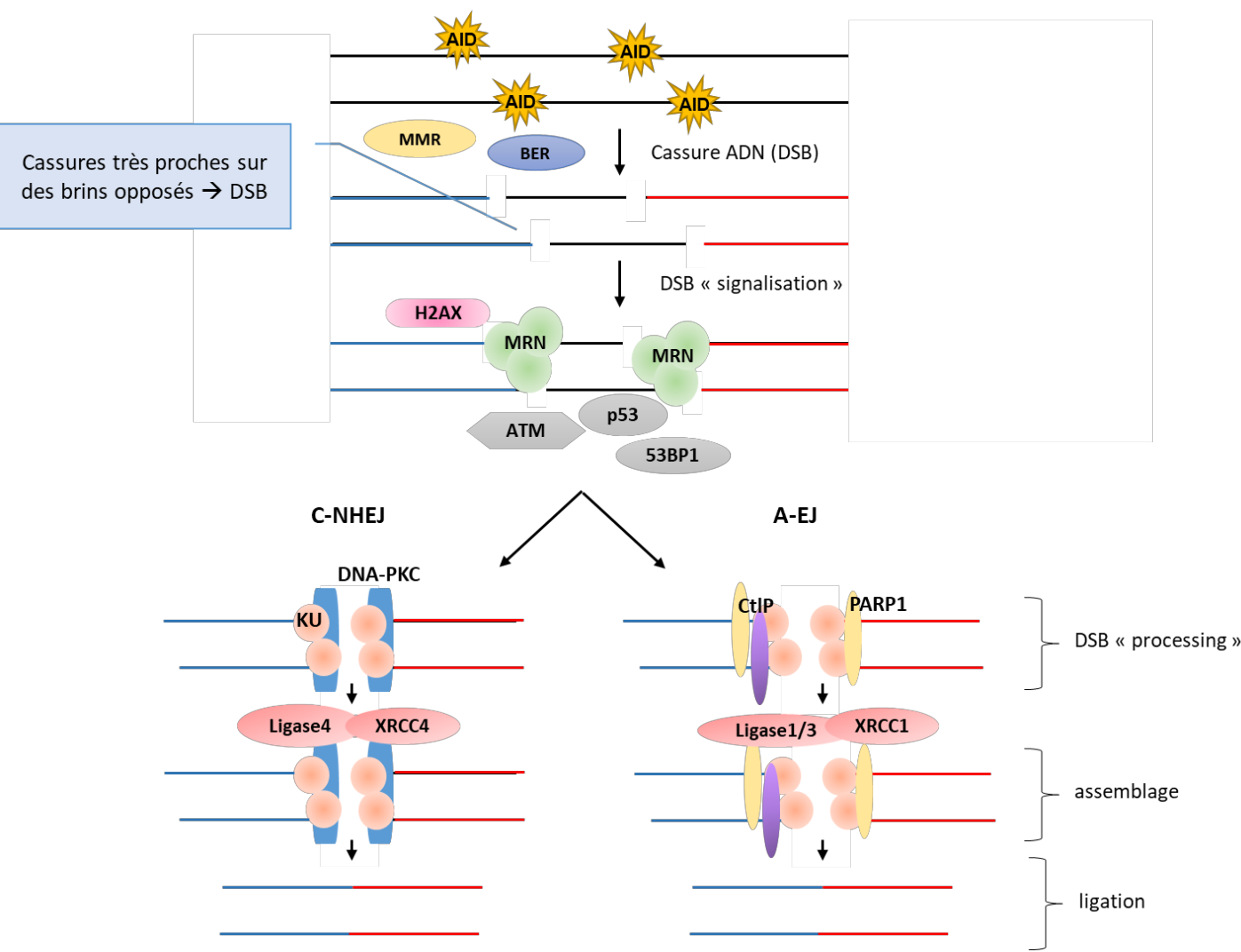
### III.2.2.2. Mécanisme moléculaire de la CSR.

Mon travail de thèse ayant consisté en l'étude de la fonction d'éléments *cis*-régulateurs au locus *IgH* sur le processus de SHM, je décrirai que brièvement le mécanisme de CSR qui se déroule en 2 phases successives :

(1) des cassures double brin de l'ADN (ou DSB pour double strand break) sont initiées (à partir du mésappariement initié par AID) dans une région switch donneuses ( $S_{\mu}$ ) et une région acceptrice ( $S_x$ ) ;

(2) une recombinaison par l'aboutage intervient entre les deux DSB initialement générées qui aboutira au remplacement du gène constant  $C_{\mu}$  par le gène constant  $C_x$  (l'ADN compris entre les 2 points de cassure sera éliminé) (**Figure 14**).

La première étape de CSR est similaire à celle décrite précédemment concernant la SHM. En effet, le mésappariement dU:dG initié par AID (dans les régions switch) va être pris en charge par la protéine UNG qui va éliminer la base dU. Le site abasique alors généré est reconnu par les endonucléases APE qui clive l'ADN créant ainsi une cassure simple brin. La richesse en hotspots de AID répétés en tandem dans les régions switch (Hackney et al., 2009) favorise l'apparition de cassures simples brins simultanées sur les 2 brins et suffisamment proches pour générer des DSB (Han et al., 2011; Stavnezer and Schrader, 2014; Xu et al., 2012). Les DSB sont reconnues par le complexe MRN (constitué par les protéines MRE11, RAD50 et NBS1) et activent la voie de réponse aux dommages à l'ADN dépendante de la protéine ATM (ataxia aelangiectasia mutated). La



**Figure 16 : Voies de réparation des cassures de l'ADN induites lors de la CSR.**

Des cassures double brin, initiées par AID, sont induites lors de la CSR. La résolution des cassures peut faire intervenir la voie du C-NHEJ ou de l'AEJ. La voie C-NHEJ fait intervenir plusieurs effecteurs protéiques, les principaux sont : Ku70, Ku80, DNA-PKCs, XRCC4 et la DNA ligase IV. D'autres effecteurs protéiques sont impliqués dans la voie AEJ, notamment les protéines : PARP-1, CtIP, XRCC1 et les DNA ligase I et III.

DSB : double strand break ou cassure double brin.

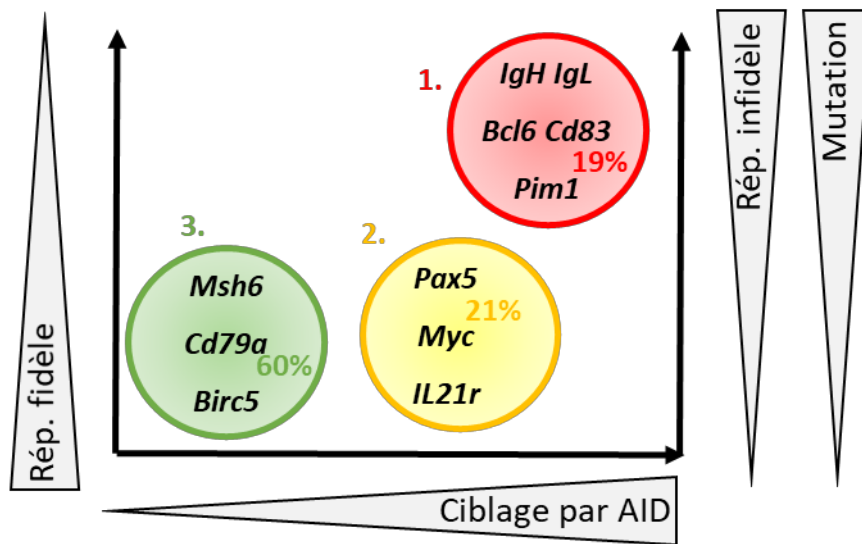
kinase ATM va alors recruter des effecteurs protéiques impliqués dans la prise en charge des DSB (p53, 53BP1, H2AX) (Lee and Paull, 2007). Il se forme alors un foyer de réparation recouvrant les sites de DSB des régions switch donneuses et acceptrices (**Figure 16**) (Bassing and Alt, 2004; Rogakou et al., 1999). Le facteur 53BP1 semble avoir une importance particulière, sa délétion entraîne un défaut de CSR (Manis et al., 2004; Ward et al., 2004).

Les régions switch destinées à être recombinées (région S $\mu$  avec une autre région S $x$  situés 60 à 160kpb en aval) vont ensuite se trouver rapprochées au sein d'une « synapse ». La jonction entre les 2 régions switch peut être effectuées par 2 voies : la voie C-NHEJ (classical non homologous end joining) et la voie A-EJ (alternative end joining) (**Figure 16**). La voie C-NHEJ (voie majeure) fait intervenir plusieurs effecteurs protéiques, les principaux sont : KU70, KU80, DNA-PKCs, XRCC4 et l'ADN ligase IV (Lieber, 2010). L'absence d'un de ces facteurs n'entraîne qu'une diminution partielle de CSR (environ 60%) mettant en évidence l'existence d'une autre voie de recombinaison (Boboila et al., 2010; Yan et al., 2007). D'autres effecteurs protéiques sont impliqués dans la voie A-EJ, notamment les protéines : PARP-1, CTLP, XRCC1 et les ADN ligase I et III (Boboila et al., 2012; HWANG et al., 2015). L'analyse de la composition des jonctions switch permet d'identifier des stigmata caractéristiques de la voie de réparation impliquée, pour exemple des jonctions directes ou présentant peu de micro homologies de bases (en général inférieur à 5 bases) résultent d'une prise en charge par la voie C-NHEJ (Boyer et al., 2017; Cortizas et al., 2013; Panchakshari et al., 2018).

### III.3. Cibles illégitimes de AID.

Pour assurer son rôle physiologique de diversification du répertoire antigénique, AID doit agir de façon restreinte sur les gènes des Ig. Etant à l'origine de mésappariements pouvant conduire à des mutations et des cassures de l'ADN, l'action de AID peut constituer une menace pour l'intégrité du reste du génome de la cellule B. Cette menace est amplifiée par le fait que certains oncogènes sont localisés au voisinage des gènes d'Ig dans les noyaux des lymphocytes B activés (Park et al., 2014; Rocha et al., 2012).

Malgré les multiples niveaux de restriction de l'activité de AID, des études menées sur les cancers du lignage B démontrent l'implication de AID (Jankovic et al., 2010; Okazaki et al., 2007; Ramiro et al., 2006) en particulier dans les lymphomes dérivant des centres germinatifs (Pasqualucci et al., 2008). Bien que la fixation de AID sur une cible ne conduise pas systématiquement à des mutations ou des cassures (Matthews et al., 2014), les signatures (ou « kataegis ») associées à des cancers du lignage B présentent des similitudes avec le mode d'action de AID (Casellas et al., 2016). Une étude récente montre que, dans un modèle déficient pour l'enzyme TET2 (ten-eleven-translocation 2), le pattern de mutation observé dans les tumeurs B développées par ces animaux est caractéristique du mode d'action de AID (avec une surreprésentation des mutations de type C vers T et G vers A) (Mouly et al., 2018). En dehors des cancers du lignage B, l'expression constitutive de AID dans des modèles de souris transgéniques entraîne des lésions au niveau des épithéliums respiratoires et



**Figure 17 : Les cibles illégitimes de AID sont réparties en 3 groupes.**

1 : les mésappariements induits par AID sont pris en charge par les effecteurs des voies de réparation de l'ADN dont l'activité va favoriser l'apparition de mutations ;

2 : les mésappariements induits par AID sont réparés par les voies de réparation de l'ADN de façon fidèle ;

3 : aucune mutation n'est détectée.

intestinaux (de type adénocarcinomes) (Matsumoto et al., 2010; Okazaki et al., 2003). Ces résultats suggèrent que l'expression aberrante de AID pourrait être à l'origine de cancers (autres que lymphoïdes) chez l'homme (Kou et al., 2007; Okazaki et al., 2003; Rebhandl et al., 2015).

De nombreux exemples montrent que l'enzyme AID est capable de cibler des gènes autre que ceux codant les Ig ; ces cibles dites illégitimes sont appelées « AID off targets ». Suite à la mise en évidence, dans des lymphomes B, de mutations dans la région 5' de l'oncogène *Bcl6* ; des études ont recherché à déterminer si ces mutations se produisaient uniquement dans un contexte de lymphomagenèse. Il s'est avéré que *Bcl6* subit des mutations dans les lymphocytes B normaux de CG suggérant que le processus de SHM n'est pas exclusivement limité aux gènes des Ig (Jardin and Sahota, 2005; Pasqualucci et al., 1998). De nombreuses études ont par la suite identifié de nouveaux gènes (autres que les gènes d'Ig) ciblés par la machinerie de mutation ; certains de ces gènes (incluant *Bcl6*, *Myc*, *Pim1*, *Pax5*...) sont impliqués dans la lymphomagenèse (Gordon et al., 2003; Müschen et al., 2000; Pasqualucci et al., 2001; Storb et al., 2001). Par la suite, l'équipe de Schatz s'est efforcée d'identifier chez la souris les cibles « AID off targets » (Liu et al., 2008). Cette étude compare par séquençage une centaine de gènes exprimés dans les cellules B de CG dans un contexte sauvage et déficient pour AID. Elle révèle que 40% des gènes sont ciblés par la machinerie de mutation et bien que la fréquence de mutation soit au moins 10 fois inférieure à celle observée au locus *IgH*, elle reste significativement supérieure à celle obtenue dans un contexte déficient pour AID. Le séquençage de ces mêmes gènes dans des modèles de souris déficientes pour les voies de réparation de l'ADN (BER et MMR), a permis d'établir une classification en 3 groupes des gènes étudiés (cibles potentielles de AID) (**Figure 17**) :

- le groupe 1 rassemble les gènes (19%) dont les mésappariements induits par AID sont pris en charge par les effecteurs des voies de réparation de l'ADN dont l'activité va favoriser l'apparition de mutations ;
- le groupe 2 concerne les gènes (21%) déaminés par AID mais réparés de façon fidèle ;
- enfin les gènes (60%) pour lesquels les mutations sont indétectables (même dans un fond déficient pour les voies de réparation de l'ADN) constituent le 3<sup>ème</sup> groupe.

Les mécanismes impliqués dans la prise en charge des déaminations induites par AID au niveau de ses cibles illégitimes ne sont pas totalement élucidés. Il semblerait qu'une balance entre la réparation dite « error-prone » et fidèle détermine le devenir du mésappariement de base initié par AID (Gu et al., 2016; Liu and Schatz, 2009).

Au-delà des événements de mutation infligés à certains oncogènes ayant pour conséquence le dérèglement de leur expression ou de la fonction des protéines qu'il codent (Migliazza et al., 1995), certains types de cancer sont caractérisés par des événements de recombinaison inter-chromosomique impliquant des cassures double brin initiées par AID (Pannunzio and Lieber, 2018). Pour exemple, un dérèglement du processus de CSR peut conduire à une translocation entre les loci *IgH* et *c-Myc*, caractéristique du lymphome de Burkitt (Li et



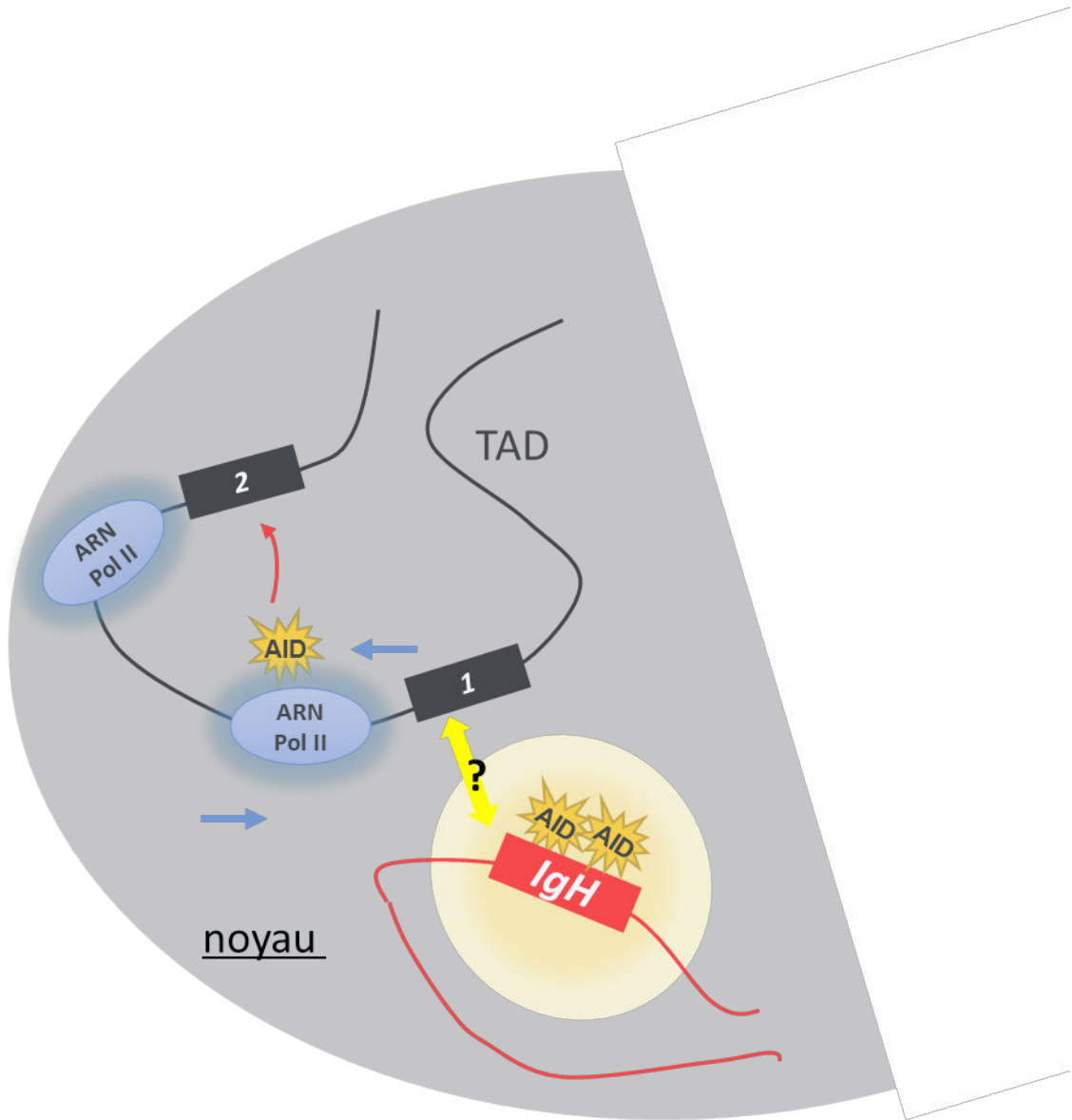
al., 2018; Oduor et al., 2017; Seifert et al., 2013) ou à une translocation entre les loci *IgH* et *c-Maf* retrouvée dans les myélomes multiples (Chesi et al., 1998).

La capacité de AID à muter des gènes non Ig (dont des oncogènes) est associée au processus de lymphomagenèse (Robbiani and Nussenzweig, 2013) mais de façon surprenante, une étude montre que ce caractère mutagène pourrait se révéler bénéfique. L'infection de cellules pré-B de moelle osseuse par le virus AMuLV (ou Abelson-murine leukemia virus) induit l'expression de AID et des mutations dans les gènes d'Ig et les cibles « off targets » : ces cellules acquièrent alors une meilleure résistance au phénomène de transformation virale (en comparaison à un contrôle déficient pour AID) (Gourzi et al., 2006).

Plusieurs équipes ont tenté d'identifier d'éventuelles caractéristiques communes entre les gènes cibles illégitimes de AID. Des analyses de sites de liaison à des facteurs de transcription identifient 3 sites potentiellement impliqués dans le recrutement de AID (E-box, YY1 et C/EBP- $\beta$ ) (Duke et al., 2013). Par ailleurs, il semble que les cibles illégitimes de AID ne soit pas enrichies en complexe RPA alors que le déroulement optimal des mécanismes de SHM et de CSR aux loci des gènes d'Ig nécessite le recrutement de cette protéine (Yamane et al., 2011). La protéine NBS1, capable de se lier aux DSB, est quant à elle enrichie sur des gènes d'Ig et sur certains « AID off targets » (Khair et al., 2015; Staszewski et al., 2011). D'autres données concernant les marques épigénétiques indiquent, qu'une fréquence de mutation élevée est corrélée à la marque H3K4me3 (Casellas et al., 2016; Kato et al., 2012).

Par ailleurs, des études menées dans un contexte déficient pour les composants du RNA exosome, révèlent la présence d'ARN non codants (substrats du RNA exosome) et préférentiellement antisens au niveau de cibles illégitimes de AID (telles que *Cd83*, *Pim1* ou encore *C-myc*). Le recrutement de AID au niveau des cibles illégitimes est également sévèrement impacté (Pefanis et al., 2014). La prise en charge des ARN non codants pour le RNA exosome semble donc être un facteur déterminant pour le recrutement de AID.

La liste des « AID off targets » dressée par l'équipe de Schatz a récemment été confirmée et étendue grâce à des techniques de haut débit (Álvarez-Prado et al., 2018). Cette étude confirme également l'enrichissement de l'ARN pol II et de SPT5 sur les cibles de AID et précise, en accord avec l'étude de Schatz, que la majorité des lésions induites par AID sur ses cibles illégitimes est prise en charge par les voies de réparation de l'ADN. Par ailleurs cette étude met en évidence que les cytosines appartenant à un motif de type AGCTNT sont plus significativement mutées que celles localisées dans un motif AGCTNV (avec V = A, C ou G), complétant ainsi les connaissances actuelles sur les motifs « hotspots » de AID. Enfin, l'analyse des « AID off targets » menée à partir des cellules B de CG issues de souris déficientes pour les voies de réparation de l'ADN révèle que les mutations détectées sont identiques à celles observées dans des échantillons de lymphomes humains (Álvarez-Prado et al., 2018).



**Figure 18 : Intervention de AID sur les cibles « off targets ».**

Dans le noyau des lymphocytes B activés de centre germinatif, les loci d'Ig et cibles « off targets » semblent être à proximité et AID semble préférentiellement agir sur des gènes cibles inclus dans un même TAD.



Il faut également noter que AID semble préférentiellement agir sur des gènes cibles inclus dans un même TAD (topologically associated domain) (Alinikula and Schatz, 2014; Qian et al., 2014) (**Figure 18**). Les TAD représentent une organisation structurale des chromosomes sous forme de boucle réalisant des interactions « longues distances » entre des promoteurs et des enhanceurs ; ces structures associent de façon topologique des gènes co-régulés par un même « super enhanceur ». Au sein de ces TAD, AID pourrait agir dans des zones de transcription convergente (Meng et al., 2014).

Enfin, il a été suggéré que l'organisation nucléaire puisse jouer un rôle important dans le ciblage de AID. Un évènement de translocation suppose que, à un instant donné, les chromosomes qui vont porter les cassures se trouvent à proximité pour effectuer la recombinaison (Robbiani and Nussenzweig, 2013; Strongin et al., 2014). Les translocations semblent plus fréquentes entre des gènes localisés en *cis* (sur le même chromosome) ou sur deux chromosomes différents situés physiquement proches (Klein et al., 2011). Il est donc envisageable que les loci des gènes Ig et les « off targets » se trouvent à proximité lors de leur déamination par AID, dans une même région nucléaire. Plusieurs études, utilisant des techniques d'hybridation *in situ* en fluorescence (fluorescence *in situ* hybridization, FISH), se sont intéressées au positionnement des cibles de AID dans les noyaux des lymphocytes. Pour exemple, des résultats obtenus par l'équipe de Fraser montre une relocalisation dynamique et transitoire des loci *IgH* et *c-Myc* dès les premières minutes suivant l'activation des lymphocytes B *in vitro* (Osborne et al., 2007).

D'autres techniques dites de « chromosome capture conformation » (3C, 4C, 5C) étudient les éventuelles interactions entre gènes. L'équipe de Skok montre alors une proximité des « AID off targets » avec le locus *IgH* dans des cellules B activées *in vitro* (Rocha et al., 2012). Cependant d'autres études par FISH 3D menées dans des cellules B triées de CG montre qu'il n'existe pas de corrélation entre la fréquence de mutation d'une cible illégitime de AID et sa proximité physique avec le locus *IgH* (Gramlich et al., 2012) (**Figure 18**).

Une partie de mon travail de thèse a concerné l'étude des cibles illégitimes de AID. Grâce à des techniques de séquençage haut débit, j'ai analysé la fréquence de mutation (reflet de l'activité de AID) de 4 gènes « off targets » (*Bcl6*, *Cd83*, *Pim1* et *Btg1*) suffisamment déaminés par AID pour être étudiés dans un contexte sauvage (compétent pour les voies de réparation MMR et BER). Cependant, la fréquence de mutation des cibles « off targets » reste faible et son étude, par les méthodes de séquençage nouvelle génération, a nécessité la mise au point d'un outil bio-informatique d'analyse de mutations rares (voir partie « Résultats : article 1 »). Afin de déterminer l'influence de l'organisation nucléaire sur le processus de ciblage illégitime des mutations somatiques, j'ai également étudié le positionnement relatif des gènes « off targets » (par DNA FISH 3D) dans des noyaux de cellules B de CG issues de différents modèles murins (voir partie « Résultats : article 3 »).



Les multiples évènements de mutation, cassure et réparation qui jalonnent le développement des lymphocytes B rendent les cellules de ce lignage particulièrement sensibles aux évènements de tumorigenèse. Un contrôle strict de ces remaniements est assuré par des éléments *cis*-régulateurs situés sur les loci des gènes d'Ig : nous aborderons dans le chapitre suivant les éléments régulateurs au locus *IgH*.



## Chapitre IV. Les régions régulatrices au locus *IgH*.

---

Le développement des lymphocytes B est associé à de nombreux événements de remaniement génique qui nécessitent une régulation spatio-temporelle stricte. Des éléments *cis*-régulateurs ont été identifiés au sein du locus *IgH*, parmi lesquels des promoteurs, des activateurs et des isolateurs (chapitre I : « Organisation des gènes des immunoglobulines. »). Grâce à l'étude de modèles murins, le laboratoire s'est employé à identifier et à préciser la fonction de certaines séquences régulatrices au locus *IgH* qui coordonnent les événements de transcription, de modification de la chromatine, de positionnement et d'organisation tridimensionnelle des gènes des immunoglobulines. Dans ce chapitre, je présenterai les éléments *cis*-régulateurs décrits au locus *IgH* et détaillerai plus précisément le rôle de région intronique E $\mu$  puisque mon travail de thèse s'est focalisé sur l'étude de cette région.

### IV.1. Les promoteurs au locus *IgH*.

De nombreux promoteurs sont répartis sur le locus entier : en 5' des segments V<sub>H</sub> et D, et en amont des gènes constants (**Figure 2**). Tout au long du développement B, des transcrits dits germinaux ou stériles (ne codant pas de protéine) sont initiés et favorisent le déroulement séquentiel des événements de recombinaison qui s'opèrent au locus *IgH*. L'importance de la transcription au cours du développement B, notamment durant les événements de remaniement (recombinaisons VDJ, SHM et CSR), a été souligné à plusieurs reprises dans ce manuscrit (chapitres II. « Développement des lymphocytes B. » et III. « AID : enzyme clé du développement B tardif. »).

#### IV.1.1. Les promoteurs pV<sub>H</sub>.

Chaque segment V<sub>H</sub> possède son propre promoteur (pV<sub>H</sub>) localisé environ 250 pb en 5' des segments V<sub>H</sub>. Ces promoteurs sont requis pour l'initiation d'une part des transcrits germinaux qui précèdent les recombinaisons de V<sub>H</sub> vers DJ<sub>H</sub> et d'autre part, après les événements de recombinaison, le promoteur pV<sub>H</sub> du segment VDJ<sub>H</sub> réarrangé est utilisé pour l'expression du transcrit de la chaîne lourde fonctionnelle dès le stade pré-B et jusqu'au stade plasmocyte. La grande majorité des promoteurs pV<sub>H</sub> est caractérisée par la présence d'une séquence « TATA box » (environ 30 pb en amont du site d'initiation de la transcription) et d'un élément initiateur (Inr, chevauchant le site d'initiation de la transcription) (Buchanan et al., 1997; Johnston et al., 2006). En amont de la « TATA box » se trouve une séquence octamérique conservée constituant un site de liaison pour les facteurs OCT appartenant à la famille des facteurs de transcription POU (Buchanan et al., 1997; Parslow et al., 1984; Roy et al., 2011). D'autres éléments caractéristiques de certains promoteurs pV<sub>H</sub> ont été décrits, tels que les éléments DICE (downstream immunoglobulin control element) et PAIR (pax- activated intergenic repeat). Les éléments DICE sont associés aux pV<sub>H</sub> des segments V<sub>H</sub> de la famille J558 et sont requis pour l'activité de leurs promoteurs (Tantin et al., 2004). Les éléments PAIR vont lier plusieurs facteurs (parmi



lesquels les facteurs PAX5, E2A et CTCF) et participent à la régulation des réarrangements de  $V_H$  vers  $DJ_H$  (Ebert et al., 2011).

#### **IV.1.2. Le promoteur pDQ52.**

Le promoteur pDQ52, localisé en amont du dernier segment  $D_H$  (nommé DQ52), est d'une importance particulière. Ce promoteur possède des séquences caractéristiques d'une activité largement supérieure à celle des promoteurs associés aux autres segments  $D_H$ , d'où son appellation de « promoteur / enhancer » (Alessandrini and Desiderio, 1991; Kottmann et al., 1994). Avant les événements de recombinaison de  $D$  vers  $J_H$ , ce promoteur est requis pour la synthèse du transcrit germinale «  $\mu 0$  » qui inclut la totalité des régions  $J_H$  et introniques en amont du gène constant  $C_\mu$  (Lennon and Perry, 1985) (**Figure 6**).

#### **IV.1.3. Les promoteurs I.**

Comme mentionné précédemment (chapitre III : « AID : enzyme clé du développement B tardif. » des lymphocytes B), chaque gène constant (sauf  $\delta$ ) constitue une unité transcriptionnelle comprenant : un promoteur et un exon « I » (ou intervening), une région S (ou switch) et les exons CH (**Figure 14**). Le transcrit initié à partir du promoteur I, de type  $I_x - C_x$ , ne code pas une protéine mais joue un rôle indispensable dans l'accessibilité des régions switch subissant des cassures double brin lors de la CSR (Nambu et al., 2003). Des transcrits antisens ont également été mis en évidence dans les régions S, mais leur rôle et leur régulation restent mal connus (Perlot et al., 2008) (**Figure 6**).

#### **IV.2. Les activateurs transcriptionnels ou « enhancers » au locus *IgH*.**

Les activateurs transcriptionnels, communément appelés « enhancers », correspondent à des régions définies d'ADN capables de stimuler la transcription d'un gène cible parfois distant. Les enhancers sont caractérisés par la présence de marques actives de la chromatine (H3K4me, H3K27Ac), de sites de fixation pour des facteurs de transcription et par leur sensibilité à la DNase. Il semblerait que les enhancers puissent être transcrits générant ainsi des ARN non codants appelés eRNA (enhancer RNA). La fonction des eRNA restent à préciser, cependant de tels transcrits pourraient être impliqués dans l'accessibilité de la chromatine et ou dans les interactions longues distances (Lam et al., 2014; Li et al., 2016a).

##### **IV.2.1. La région intronique $E_\mu$ .**

La région intronique  $E_\mu$  est située au locus *IgH* entre les segments  $J_{H4}$  et  $C_\mu$  (**Figure 2**). Une telle localisation permet à la région  $E_\mu$  d'être conservée au cours des différents événements de recombinaison s'opérant au locus *IgH*. Cette région est composée d'un activateur transcriptionnel « core  $E_\mu$  » ( $cE_\mu$ ) de 220 pb et flanqué par 2 régions d'attachement à la matrice nucléaire «  $MARS_{E_\mu}$  » (matrix attachment regions) de 310 et 350 pb. L'activateur intronique  $cE_\mu$  possède une activité spécifique au lignage B, et fut le 1er activateur transcriptionnel eucaryote décrit (Banerji et al., 1983; Gillies et al., 1983). Depuis, des modèles porteurs d'une

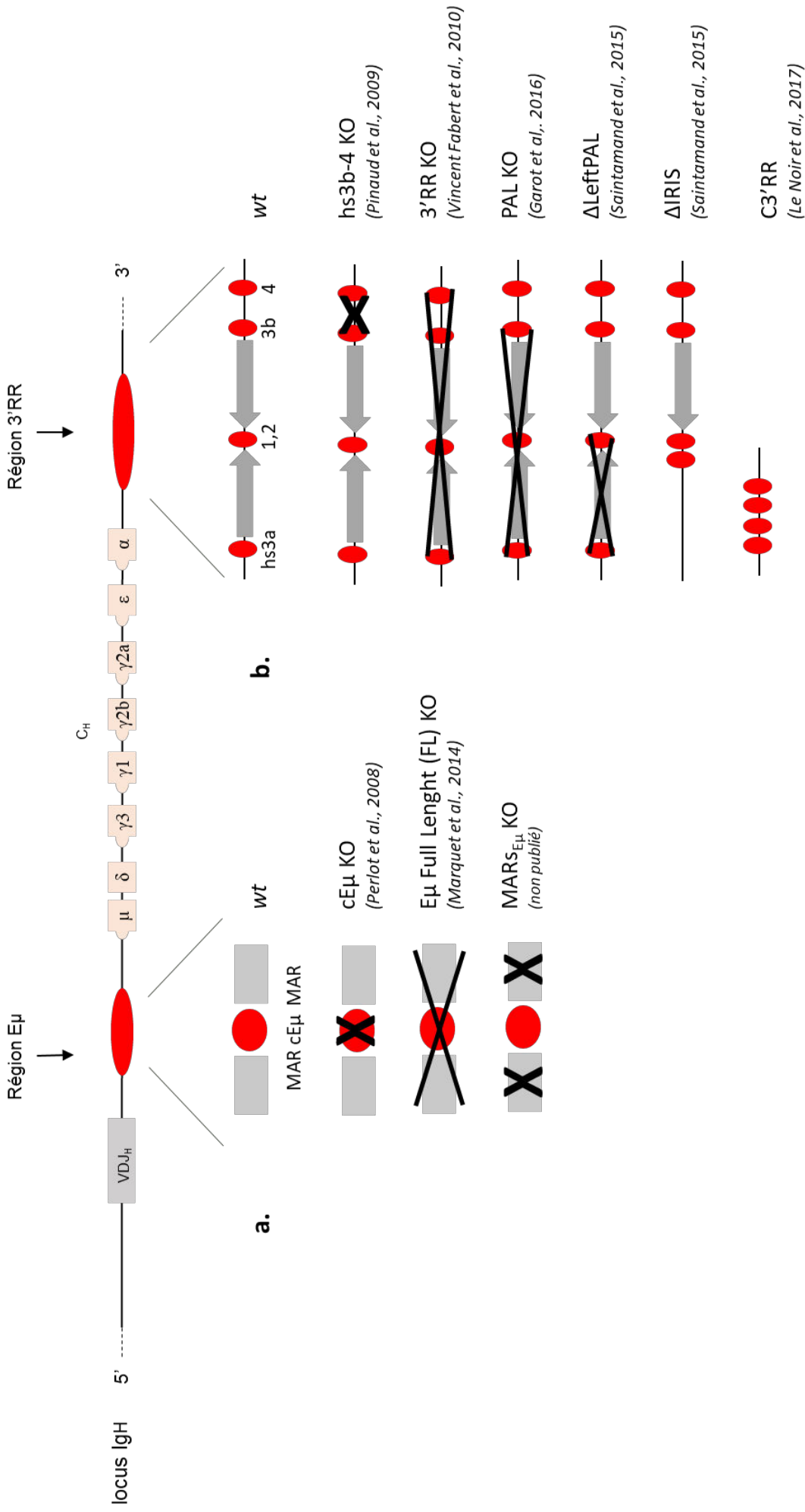


Figure 19 : Modèles murins disponibles au laboratoire pour étudier les rôles de la région Eμ (a) et 3'RR (b).



délétion totale ou partielle de la région  $E_{\mu}$  ont été développés afin de préciser les fonctions respectives du core  $E_{\mu}$  et des régions  $MARS_{E_{\mu}}$  (**Figure 19a et table 2**).

#### **IV.2.1.1. L'activateur intronique core $E_{\mu}$ ( $cE_{\mu}$ ).**

La région du core  $E_{\mu}$  contient plusieurs sites de fixation pour des facteurs de transcription (**figure 20**). Depuis l'extrémité 5' vers 3', se situent : un site C/EBP (CCAAT Enhancer Binding Protein),  $\mu E1$ ,  $\mu E5$ ,  $\mu E2$ ,  $\mu A$ ,  $\mu E3$ ,  $\mu B$ ,  $\mu E4$  et une séquence octamérique similaire à celle située au niveau des segments  $V_H$  (Ernst and Smale, 1995). L'activité restreinte de  $E_{\mu}$  au lignage B est assurée grâce à la fixation de facteurs activateurs ou inhibiteurs de la transcription, en particulier au niveau des sites  $\mu E4$  et  $\mu E5$ . Des protéines activatrices de la famille E2A fixent ces sites lors de l'activation de  $E_{\mu}$ , alors que dans des cellules non-B c'est la protéine inhibitrice ZEB qui occupe ces sites entraînant la répression de l'activité de  $E_{\mu}$  (Genetta et al., 1994).









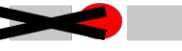


Plusieurs études menées à partir de modèles associés à des modifications (délétion, remplacement, insertion) de l'activateur  $E_{\mu}$ , montrent son implication dans les événements de remaniement précoces.

La présence de  $E_{\mu}$  au sein d'un transgène est suffisante pour initier les réarrangements  $D - J_H$  dans des cellules lymphoïdes (Ferrier et al., 1990).

Le remplacement de  $E_{\mu}$  par un oligonucléotide dans un modèle de souris chimériques entraîne un blocage (incomplet) des recombinaisons VDJ, les recombinaisons de type  $V_H$  vers  $DJ_H$  sont plus sévèrement affectées (Serwe and Sablitzky, 1993). Lorsque  $E_{\mu}$  est remplacé par une cassette pGK-Neo<sup>r</sup>, les réarrangements VDJ ( $D - J_H$  et  $V_H - DJ_H$ ) sont plus sévèrement affectés (Chen et al., 1993; Perlot et al., 2005; Sakai et al., 1999a). Le même défaut est observé lorsque la cassette est insérée dans la région située en 5' du core enhancer  $E_{\mu}$ , soulignant le fait que l'intégrité de la région en amont de  $E_{\mu}$  et de l'enhancer est un prérequis au mécanisme de recombinaisons VDJ (Chen et al., 1993; Sakai et al., 1999a).

La délétion de l'activateur  $E_{\mu}$  seul ou en combinaison avec les  $MARS_{E_{\mu}}$  entraîne une diminution significative des recombinaisons VDJ de type  $V_H - DJ_H$  (Sakai et al., 1999a). Des études menées à partir d'animaux KO homozygotes porteurs d'une délétion du core  $E_{\mu}$  ( $cE_{\mu}$  KO), confirment le défaut de recombinaison VDJ ( $D$  vers  $J_H$  et  $V_H$  vers  $DJ_H$ ). Ce défaut de recombinaison n'est cependant pas total dans ce modèle chez lequel on observe un blocage incomplet du développement B associé à une accumulation de précurseurs du stade pro-B et une diminution du nombre de cellules B matures en périphérie (Perlot et al., 2005).

Ce constat suggère l'existence d'un autre élément régulateur dont l'activité serait redondante ou complémentaire à celle de  $E_{\mu}$ . Parmi les candidats potentiels, l'activité du promoteur/enhancer pDQ52 ne semble pas être redondante à celle de  $E_{\mu}$  pour le processus de recombinaison VDJ. En effet, la seule délétion de pDQ52 dans la souris montre que celui-ci est dispensable pour le mécanisme de recombinaison ; les animaux présentant une délétion combinée de  $E_{\mu}$  et du promoteur/enhancer pDQ52 présentent quant à eux un défaut de recombinaison similaire à celui de la délétion de  $E_{\mu}$  (Afshar et al., 2006; Nitschke et al., 2001).

Modèles d'étude		Recb. VDJ	SHM	CSR	Réf.
<i>wt</i>	 MAR 5' cEμ MAR 3'				
Modèles chimériques :					
Eμ FL KO		↘	-	-	Sakai et al., 1999
cEμ KO			-	-	
MAR <sub>Eμ</sub> 5' KO		✓	-	-	
MAR <sub>Eμ</sub> 3' KO			-	-	
MAR <sub>S<sub>Eμ</sub></sub> KO			-	-	
Modèles KI :					
Eμ FL KO		-	↘ très mineure	✓	Li et al., 2010
Modèles KO :					
cEμ KO		↘	✓	✓	Perlot et al., 2005
MAR <sub>Eμ</sub> 5'-cEμ KO		↘	-	-	Afshar et al., 2006
Eμ FL KO		↘	↘	✓	Marquet et al., 2014
MAR <sub>S<sub>Eμ</sub></sub> KO		✓	↘	✓	non publié

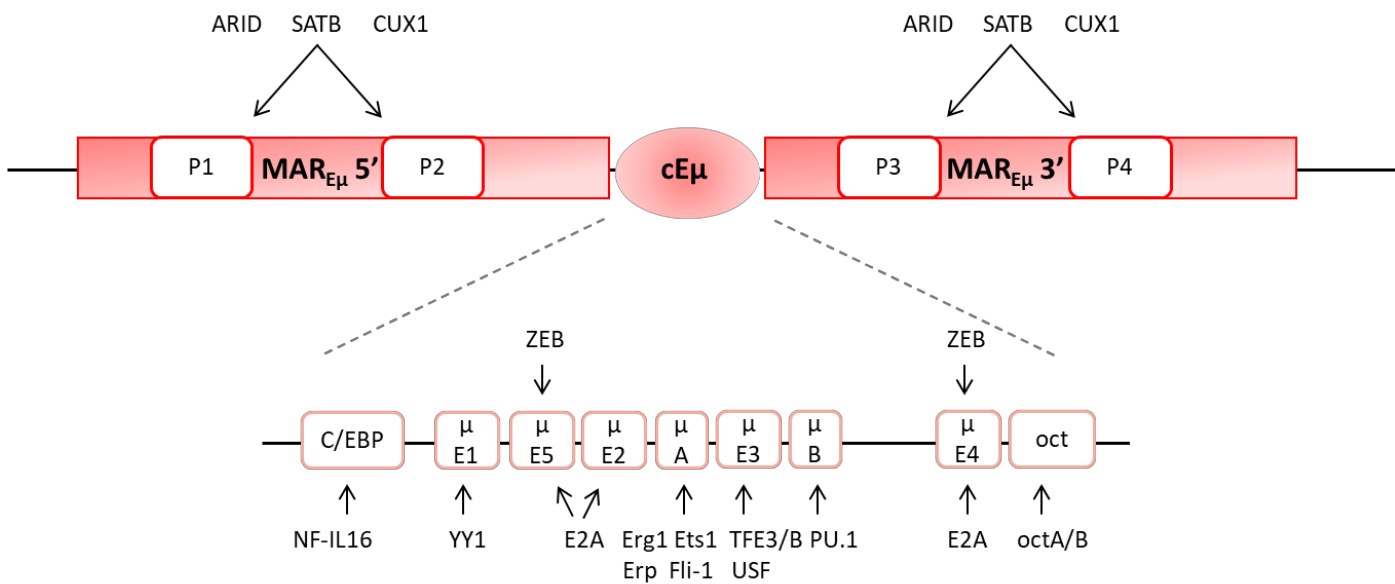
**Table 2 : Modèles murins de délétion de la région Eμ et conséquences sur les évènements de recombinaisons VDJ, d'hypermutation somatique (SHM) et de recombinaison de classe (CSR).**

Pour que le mécanisme de recombinaisons VDJ puisse se dérouler de façon efficace, les segments  $V_H$ , D et  $J_H$  doivent être accessibles aux endonucléases RAG. La transcription germinale et des modifications de la chromatine vont contribuer à cette accessibilité.

La transcription germinale (ou stérile) est initiée avant même les 1<sup>ers</sup> réarrangements de type D- $J_H$ . Des transcrits germinaux  $\mu 0$  et  $l\mu$  sont respectivement initiés à partir du promoteur pDQ52 et du core  $E\mu$  (Alessandrini and Desiderio, 1991; Lennon and Perry, 1985). Des transcrits antisens ont également été décrits en amont du core  $E\mu$  au niveau des segments  $J_H$  et  $D_H$  (Bolland et al., 2007a). Après les recombinaisons de D vers  $J_H$ , des transcrits germinaux sens et antisens sont détectés dans la région  $V_H$  (Bolland et al., 2004b, 2007a). La délétion du core  $E\mu$  est associée à une diminution significative des transcrits stériles  $\mu 0$  et  $l\mu$  mais également des transcrits antisens associés à l'accessibilité de la région  $D_H$ . Le défaut de transcription  $\mu 0$  observé en l'absence du core  $E\mu$  corrèle avec une diminution du recrutement de l'ARN polymérase II (Chakraborty et al., 2009a). En revanche, la délétion du core  $E\mu$  n'affecte pas la transcription germinale de la région  $V_H$  (Afshar et al., 2006; Bolland et al., 2007a; Perlot et al., 2005).

Des modifications de la chromatine sont également observées lors du processus de recombinaisons VDJ, et les régions à recombiner sont plutôt associées à une hyper acétylation des histones (Chowdhury and Sen, 2001). Dans des cellules non B, le locus *IgH* est associé à des marques inactives de la chromatine (H3K9me2). A l'opposé, dans des précurseurs du lignage B les marques inactives de la chromatine laissent place aux marques actives (H3K9ac). Une étude s'est intéressée à ces marques de la chromatine : en absence du  $cE\mu$  le profil de distribution de l'H3K9me2 n'est pas affecté ; alors que celui de l'H3K9ac est fortement impacté. Ces données suggèrent que la configuration active du locus *IgH* s'acquière de façon séquentielle selon une 1<sup>ère</sup> étape indépendante de  $E\mu$  et une 2<sup>nde</sup> phase dépendante de  $E\mu$  (Chakraborty et al., 2009a).

Plus récemment un modèle de souris KO dans la lignée germinale dépourvu de la totalité de la région  $E\mu$  (délétion du core  $E\mu$  et des 2 régions  $MAR_{E\mu}$  flanquantes : modèle  $E\mu$  Full Length ou  $E\mu$  FL KO) a été élaboré au sein du laboratoire (Marquet et al., 2014). En accord avec les données de la littérature, les animaux  $E\mu$  FL KO présentent un défaut de recombinaisons VDJ et un blocage incomplet du développement B au stade pro-B accompagné d'une diminution drastique des populations B en périphérie. Cette étude montre qu'en absence de la région  $E\mu$ , l'utilisation des différents segments  $V_H$  n'est pas affectée. Par ailleurs, ce modèle précise la fonction enhancer du  $cE\mu$  sur la production des chaînes lourdes d'Ig immédiatement après les réarrangements fonctionnels. La délétion de la région  $E\mu$  FL diminue l'expression de la chaîne lourde d'Ig aux stades précoces dans la moelle osseuse. Au stade pré-B les conséquences d'une diminution du récepteur (pré-BCR) limitent l'expansion clonale et induisent une sous-représentation de cette population. Aux stades suivants, B immatures et transitionnels, la diminution de l'expression du BCR modifie le destin des populations périphériques en favorisant la différenciation de la sous population de la zone marginale (B MZ). La sous-représentation de la population B folliculaire (B FO) est à mettre en relation avec la baisse du nombre total de



**Figure 20 : Représentation schématique de la région  $E\mu$  et des sites de fixation des facteurs *trans*-régulateurs.**

Chaque région  $MAR_{E\mu}$  contient 2 sites de fixation pour les MAR-BP (MAR-binding protein), nommés P1 à P4. Cette représentation n'est pas à l'échelle.

lymphocytes B transitionnels produits et leur faible capacité à exprimer leur BCR (Marquet et al., 2014). Aux stades matures, la délétion de la région  $E\mu$  FL n'impacte pas l'expression de la chaîne lourde d'Ig (Marquet et al., 2014). Ces dernières études montrent que le  $cE\mu$  est un élément *cis*-régulateur indispensable au bon déroulement des étapes de développement B précoce et que la fenêtre d'action du  $cE\mu$  s'étend depuis le stade pro-B jusqu'au stade B immatures transitionnels (Marquet et al., 2014; Peng and Eckhardt, 2013).

Les modèles animaux  $cE\mu$  KO ou  $E\mu$  FL KO présentent tous deux un blocage incomplet du développement B au stade pro-B accompagnés d'une réduction (d'au moins 50%) de la population de cellules B matures en périphérie (Marquet et al., 2014; Perlot et al., 2005). Une fois les cellules B matures générées en l'absence du  $cE\mu$ , cet élément ne semble plus jouer un rôle substantiel lors des étapes de développement B tardif. Néanmoins, en fonction du modèle d'étude employé, les conclusions sur l'implication de  $E\mu$  au cours des événements de CSR et de SHM divergent sensiblement.

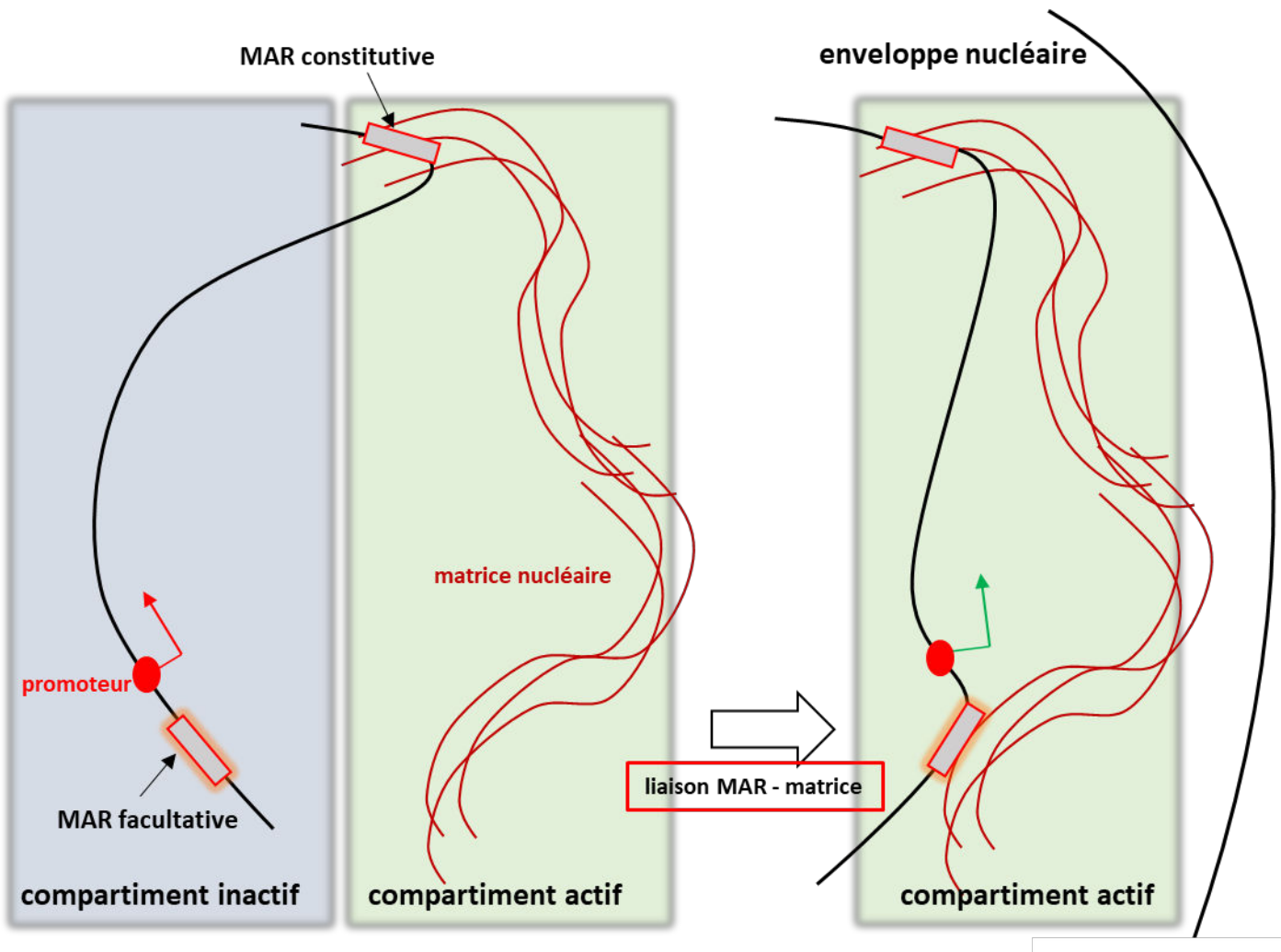
Dans des modèles d'hybridomes, la délétion totale de la région  $E\mu$  (core  $E\mu$  et  $MAR_{E\mu}$ ) est associée à une diminution de CSR de l'allèle non fonctionnel (Bottaro et al., 1998a). En revanche, les études menées dans les modèles de souris KO ( $cE\mu$  KO) ne révèlent pas de défaut significatif de CSR (évaluée par quantification ELISA des différentes classe d'Ig dans le sérum et par la capacité des cellules B à effectuer la CSR *in vitro*) (Perlot et al., 2005). Le mécanisme de CSR n'est également que modestement affecté dans des cellules B issues d'animaux dépourvus de la totalité de la région  $E\mu$  associée à un allèle *IgH* préalablement réarrangé (Li et al., 2010a).

L'impact des diverses délétions réalisées dans la région  $E\mu$  sur la SHM demeure à ce jour controversé. Dans les modèles d'hybridomes, la fréquence des mutations n'est pas modifiée en l'absence combinée du core  $E\mu$  et des  $MAR_{E\mu}$  alors que la délétion du core  $E\mu$  seul ou des  $MAR_{E\mu}$  seules entraîne une diminution significative de la fréquence de mutation (Ronai et al., 2005a). A l'inverse, la fréquence des mutations ainsi que le profil de répartition des mutations n'est pas affecté *in vivo*, au locus *IgH*, chez les d'animaux  $cE\mu$  KO ; bien que ce modèle présente une proportion de séquences non mutées plus importante (Perlot et al., 2005). Enfin, les modèles de souris porteurs d'un allèle *IgH* préalablement réarrangé associé à une délétion de la région  $E\mu$  ne présentent pas de défaut évident de SHM (fréquence et spectre de mutation) (Li et al., 2010a).

#### **IV.2.1.2. Les régions d'attachement à la matrice nucléaire ( $MAR_{E\mu}$ ).**

Le core enhancer  $cE\mu$  est flanqué par 2 régions d'attachement à la matrice nucléaire ( $MAR_{E\mu}$ ) qui ont été mises en évidence grâce à une technique *in vitro* de fixation d'ADN à des extraits solubles préparés à partir de noyaux de cellules B (Cockerill et al., 1987).

De telles régions d'attachement à la matrice nucléaire se trouvent réparties de façon non aléatoire au sein du génome et partagent des caractéristiques communes. Les régions MARs sont des séquences d'ADN longues de 200 à 1000 pb, enrichies en bases A/T et pouvant contenir des sites de liaisons pour les topoisomérases



**Figure 21 : Schématisation des 2 types de régions MARs.**

Les MARs constitutives et facultatives lient la matrice nucléaire de façon permanente et réversible, respectivement. De telles régions peuvent promouvoir l'activation de gènes.

(enzymes impliquées dans la gestion des contraintes de torsion de l'ADN). Les régions MARs sont préférentiellement localisées à proximité de régions transcrites, dans des portions introniques et au niveau des télomères (Gluch et al., 2008; Kisseljova et al., 2014; Wang et al., 2010). En se basant sur ces propriétés communes, des outils d'identification et de prédiction des MARs ont été développés (tels que MAR-Finder, SMARTest ...). Cependant le pouvoir de prédiction de ces outils n'est pas optimal et la classification en tant que MAR des régions identifiées par ces outils doit être considérée avec prudence (Evans et al., 2007).

La matrice nucléaire correspond à une définition générique (et relativement simpliste) du réseau de protéines localisé à l'intérieur du noyau, principalement constitué de protéines autres que des histones obtenues après préparation saline d'extraits nucléaires (Berezney and Coffey, 1974; Wang et al., 2010). Les protéines de type lamines, matrines et les ribonucléoprotéines sont les protéines majeures composant la matrice nucléaire (Hakes and Berezney, 1991; Nakayasu and Berezney, 1991; Razin et al., 2014). Des études ont notamment décrits des interactions entre les lamines et les régions MARs (Ludérus et al., 1992; Zhao et al., 1996). Par leur ancrage au réseau fibrillaire constituant la matrice nucléaire, les MARs favorisent une organisation dynamique de la chromatine sous forme de boucle (Heng et al., 2001, 2004; Jackson et al., 1990). Si des études associent la présence des MARs à une augmentation de la transcription (Martins et al., 2004; Skowronska-Krawczyk and Rosenfeld, 2015) ; de telles régions peuvent également jouer le rôle de régions isolatrices de la chromatine (Goetze et al., 2005; Yusufzai and Felsenfeld, 2004). La littérature définit 2 types de régions d'attachement à la matrice nucléaire (**Figure 21**) :

- les MARs dites constitutives liées de façon permanente à la matrice nucléaire et présentant une hypersensibilité à la DNase I dans tous les types cellulaires ;
- et les MARs dites facultatives, telles que les régions encadrant  $E\mu$  ( $MAR_{E\mu}$ ), associées de façon réversible à la matrice nucléaire et présentant une hypersensibilité à la DNase I en fonction du type cellulaire et du statut de différenciation (Britanova et al., 2005; Gluch et al., 2008).

Les  $MAR_{E\mu}$  au locus *IgH* ont initialement été décrites comme des éléments répresseurs du locus *IgH* dans des cellules non B (Imler et al., 1987; Kadesch et al., 1986). Les travaux suivants ont montré que la délétion de ces régions est associée à une diminution significative de la transcription d'un transgène *IgH* dans des cellules pré-B, probablement en altérant l'accessibilité des régions à proximité (Jenuwein et al., 1997a). Si le rôle des  $MAR_{E\mu}$  reste difficile à élucider, c'est notamment parce que ces régions constituent de véritables plateformes de recrutement pour divers facteurs protéiques présentant un potentiel activateur ou répresseur. Les  $MAR_{E\mu}$  possèdent des sites de liaisons pour des protéines liant des séquences riches en bases A/T, telles que les protéines de la famille ARID (AT Rich Interacting Domain) ou SATB (Special AT Binding) (Yokota and Kanakura, 2014) (**Figure 20**).





Parmi les protéines de la famille ARID, la protéine ARID3A ou BRIGHT (B Cell Regulator of IgH Transcription) possède 4 sites de liaisons au sein des régions MARs<sub>EM</sub>. BRIGHT est décrit comme étant un facteur de transcription permettant d'augmenter la transcription de la chaîne lourde d'Ig (Webb et al., 1991). La fixation de BRIGHT semble faciliter l'accessibilité à la chromatine (Herrscher et al., 1995; Lin et al., 2007) ; de même la surexpression de ce facteur induit une augmentation de la sensibilité à la DNase I des régions environnantes (Kaplan et al., 2001a, 2001b). La capacité de BRIGHT à activer la transcription dépend de son domaine de liaison à l'ADN (Nixon et al., 2004). La protéine BRIGHT, est exprimée dans les cellules du lignage B à des stades précoces (pré-B) et tardifs (B transitionnel, B activé, B mémoire et plasmocyte) mais également dans les cellules souches et certains précurseurs hématopoïétiques (An et al., 2010; Ratliff et al., 2014a; Webb et al., 1998). L'élaboration d'un modèle de souris KO pour BRIGHT a révélé son rôle indispensable au bon déroulement de l'érythropoïèse, la délétion de BRIGHT entraînant une létalité embryonnaire (Webb et al., 2011a). Seul 1% des embryons survivent et présentent un blocage du développement B précoce au stade pré-B accompagné d'une diminution des populations B transitionnelles et périphériques (Webb et al., 2011a). La létalité associée au KO de BRIGHT a limité l'étude de sa fonction aux stades B tardifs ; des modèles de KO conditionnels (restreints aux cellules B) ont alors été élaborés et montrent que la suppression de BRIGHT, induite par l'expression de la recombinaison CRE au locus *Cd79a* (Hobeika et al., 2006), induit une diminution de la sous population B1 au profit des cellules B2 en périphérie (Habir et al., 2017). Cette fonction de BRIGHT sur le développement des cellules B1 avait été précédemment observée dans un modèle transgénique exprimant un dominant négatif (Nixon et al., 2008). L'implication de BRIGHT sur l'homéostasie B est par ailleurs confirmée par une étude montrant que son expression forcée dans un modèle de souris transgénique induit une amplification de la sous population marginale (MZ B) associée à la production d'auto-anticorps (Oldham et al., 2011; Shankar et al., 2007). Dans ces modèles, l'implication précise de BRIGHT sur les mécanismes de CSR et de SHM n'a jusqu'alors pas été étudiée.

Deux autres membres de la famille ARID partageant des caractéristiques communes avec BRIGHT ont été identifiés : il s'agit de ARID3B et ARID3C (ou BRIGHT like). ARID3B est également une « MARs Binding Protein » (MAR-BP) et peut s'associer à ARID3A pour faciliter son action favorisant la transcription (Kim et al., 2007; Kurkewich et al., 2016; Numata et al., 1999). La protéine ARID3B ne semble pas être impliquée dans la survie des animaux KO pour ARID3A ; le KO de ARID3B conduisant à une létalité embryonnaire plus précoce (Kurkewich et al., 2016; Webb et al., 2011a). La délétion conditionnelle de ARID3B est associée à une diminution des progéniteurs lymphoïdes communs et des cellules du lignage B qui en découlent (Kurkewich et al., 2016). L'expression de ARID3C semble coïncider avec celle de BRIGHT ; cette protéine agirait comme co-activateur de la transcription médiée par BRIGHT (Tidwell et al., 2011).

Des protéines de la famille SATB (Special AT-rich Binding proteins) ont également la capacité de lier des régions MARs<sub>EM</sub> au locus *IgH*, c'est le cas des protéines SATB1 et SATB2.



La protéine SATB1 est une MAR-BP exprimée de façon très majoritaire dans les thymocytes (Dickinson et al., 1992). De nombreuses études se sont intéressées au rôle de SATB1 au cours du développement lymphocytaire T ; le rôle potentiel de cette MAR-BP dans les cellules B n'a jusqu'alors pas été approfondi. Les modèles de souris déficients pour SATB1 sont caractérisés par un thymus de petite taille, des défauts multiples à différents stades du développement T et une mortalité à 3 semaines (Akiba et al., 2018; Alvarez et al., 2000a; Kakugawa et al., 2017). Bien que SATB1 soit exprimée dans les progéniteurs lymphoïdes communs aux cellules du lignage B et T, la suppression de SATB1 ne semble pas affecter le développement B (Alvarez et al., 2000a; Satoh et al., 2013). SATB1 est un facteur régulant la capacité d'auto-renouvellement et de différenciation des cellules souches hématopoïétiques (Savarese et al., 2009; Will et al., 2013). Cette protéine contribue à l'organisation de la chromatine sous forme de boucles ancrées à la matrice nucléaire et SATB1 est capable de recruter des protéines impliquées dans le remodelage de la chromatine (telles que les enzymes à l'origine des modifications des histones) (Cai et al., 2006; Hao et al., 2015; Pavan Kumar et al., 2006; Yasui et al., 2002).

La protéine SATB2 présente 67% d'homologie avec SATB1 et est majoritairement exprimée dans les cellules pré-B. La liaison de SATB2 aux régions MAR<sub>SEμ</sub> semble nécessaire pour son activité favorisant la transcription. Par ailleurs, l'activité et la localisation de SATB2 à la périphérie du noyau est régulée par des événements de sumoylation (Dobrev, 2003). Le modèle KO de SATB2 n'est pas adapté à l'étude de sa fonction dans les lymphocytes B car cette protéine est également exprimée dans plusieurs structures du cerveau et dans les ostéoblastes : les animaux déficients pour SATB2 ne sont pas viables et présentent de profonds défauts du développement cranio-facial (Dobrev et al., 2006). De la même façon que SATB1, SATB2 semble être impliqué dans le recrutement de facteurs influençant l'état de la chromatine (Gyorgy et al., 2008).

Parmi les protéines liant les MAR<sub>SEμ</sub>, la protéine Cux1/CDP (Cut-like/CCAAT displacement protein) semble être un facteur répresseur de la transcription (Scheuermann and Chen, 1989; Zong and Scheuermann, 1995). Cux1/CDP agirait en compétition avec les facteurs *trans*-activateurs pour la liaison aux régions d'attachement à la matrice nucléaire, expliquant ainsi son rôle répresseur (Goebel et al., 2002; Kim et al., 2017; Zong and Scheuermann, 1995). Cux1/CDP est exprimé dans les cellules B à des stades précoces uniquement, mais également dans des cellules non B (cellules T, fibroblastes, cellules rénales...) (Ledford et al., 2002; Wang et al., 1999a). Le KO de cette protéine est associé à une mortalité néonatale partielle, les animaux déficients pour Cux1/CDP présentent notamment une réduction du nombre de cellules B et T associée à une augmentation de l'apoptose. D'autres défauts sont décrits, tels que des retards de croissance, des altérations du développement des poumons, des cas d'infertilité ...conduisant à une mortalité importante des animaux vers 3 semaines (Sinclair et al., 2001).

Les études décrivant la fonction des diverses protéines capables de lier les MAR<sub>SEμ</sub> suggèrent que ces régions peuvent endosser plusieurs rôles en fonction du type MAR-BP recrutée. Cependant, la localisation des régions MAR<sub>SEμ</sub>, de part et d'autre du cEμ, suggère que la fonction des MAR<sub>SEμ</sub> soit liée à celle de l'enhancer intronique.



Les nombreux modèles d'étude associant une délétion partielle ou totale de la région  $E\mu$  (incluant ou non les MARs situées en 5' ou en 3' du core enhancer) n'ont pas permis de révéler un rôle substantiel de ces régions au cours du développement B. En effet, les défauts constatés lors des étapes de recombinaisons VDJ sont uniquement attribués à la délétion du core  $E\mu$  (Afshar et al., 2006; Perlot et al., 2005; Sakai et al., 1999a). Les MARs $E\mu$  ne semblent pas non plus intervenir de façon significative dans le processus de CSR (Bottaro et al., 1998a; Li et al., 2010a). L'implication des MARs $E\mu$  dans la SHM n'a jusqu'alors jamais été établie : les divers modèles établis (hybridomes, transgéniques et porteurs d'un allèle *IgH* réarrangé) apportent chacun leur lot de contradictions (Giusti and Manser, 1993; Li et al., 2010; Lin et al., 1998; Ronai et al., 2005). Plus récemment, une étude menée dans la lignée cellulaire Ramos (lymphome de Burkitt humain) montre que la délétion d'une portion de la région située en aval du dernier segment  $J_H$  (incluant une partie de la MAR $E\mu$  en amont de  $cE\mu$ ) n'affecte en rien la SHM (Castiblanco et al., 2018).

Au locus *Ig $\kappa$* , une région MAR se trouve également présente en 5' de l'enhancer intronique  $iE\kappa$  (situé entre les segments  $J_{\kappa 5}$  et  $C\kappa$ ) (Cockerill et al., 1986) (**Figure 11**). L'implication de cette région MAR $E\kappa$  dans le processus de SHM a été suggéré dans un modèle de souris transgéniques puis confirmée dans un modèle de souris KO (Betz et al., 1994; Goyenechea et al., 1997; Yi et al., 1999a). A l'inverse des multiples modèles KO incluant les MARs $E\mu$ , la délétion la région MAR au locus *Ig $\kappa$*  induit un réarrangement précoce de type  $V\kappa$  vers  $J\kappa$  et une diminution de la SHM (Yi et al., 1999a).

Dans le but de préciser le rôle des régions MARs $E\mu$  et leur possible implication dans le processus de SHM, un modèle murin KO porteur d'une délétion des 2 régions d'attachement à la matrice nucléaire encadrant l'activateur  $cE\mu$  a été élaboré au sein du laboratoire. Nos études montrent que ces régions MARs $E\mu$  sont de véritables éléments *cis*-régulateurs du processus de SHM. Dans les cellules B de CG issues d'animaux MARs $E\mu$  KO, la fréquence de mutation au locus *IgH* est significativement diminuée. De façon inattendue, les MARs $E\mu$  sont également capables d'agir comme éléments « *trans*-régulateurs », influençant les mutations introduites au locus des chaînes légères mais également aux niveaux des cibles illégitimes de AID. Le rôle des MARs $E\mu$  au locus *IgH* a fait l'objet d'un article présenté dans la partie résultat (article 2). Les données obtenues concernant l'action « *trans*-régulatrice » des MARs $E\mu$  seront également présentées dans la partie résultat (article 3).

## **IV.2.2. La région régulatrice en 3' du locus *IgH* (3'RR).**

### **IV.2.2.1. Structure de la région 3'RR.**

L'enhancer intronique  $E\mu$  a longtemps été considéré comme le régulateur majeur au locus *IgH*. Cependant, plusieurs évidences laissaient suggérer l'existence d'une autre région régulatrice située en 3' du locus *IgH*. Des expériences de délétion de  $E\mu$  dans des lignées cellulaires révèlent que celles-ci étaient toujours capables d'exprimer des gènes du locus *IgH* (Zaller and Eckhardt, 1985). D'autres études montreront ensuite que la suppression, dans des lignées cellulaires, d'une partie du locus *IgH* en aval du gène constant  $C\alpha$  entraîne une



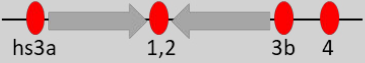

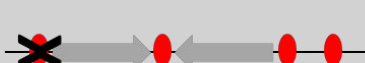

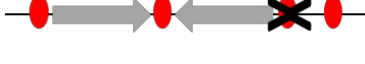

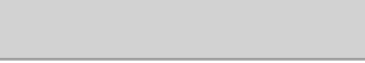
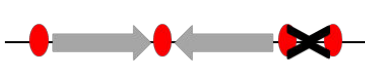


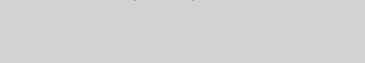
diminution non négligeable de la transcription de la chaîne lourde d'Ig (Gregor and Morrison, 1986). Quatre activateurs transcriptionnels localisés en aval du gène  $C\alpha$ , et dont l'activité est spécifique au lignage B, ont été successivement identifiés (chez le rat, la souris puis chez l'homme) (Dariavach et al., 1991; Pettersson et al., 1990; Pinaud et al., 1997).

Chez la souris, la région régulatrice en 3' du locus *IgH* s'étend sur près de 30 kb en aval du gène constant  $C\alpha$  et est composée de 4 activateurs transcriptionnels identifiés comme des sites d'hypersensibilité à la DNase I. Dans le sens 5' vers 3' les activateurs se répartissent tel que : hs3a, hs1-2, hs3b et hs4 selon une organisation structurale bien définie. En effet, la 3'RR est caractérisée par une structure quasi palindromique, d'environ 25 kb, couvrant les 3 premiers enhancers. Cette structure quasi palindromique est centrée sur l'enhancer hs1-2 qui se trouve flanqué par des régions incluant de nombreuses séquences inversées et répétées dont les limites sont les éléments hs3a en 5' et hs3b en 3'. L'activateur hs4 se trouve en dehors de la structure quasi palindromique et constitue la partie distale de la région 3'RR. La structure particulière de cette région suggère une organisation en 2 modules : un module proximal composé par la région palindromique et ses 3 enhancers hs3a, hs1-2, hs3b et un module distal contenant l'enhancer hs4 (Chauveau and Cogné, 1996; Pinaud et al., 2011) (**Figure 19b**).

Une particularité de la 3'RR est liée au fait qu'elle contient des séquences répétitives dites LS (like switch) comparables aux régions switch. Les séquences LS peuvent être ciblées par AID entraînant une recombinaison de type  $S\mu$  - 3'RR qui entraîne la délétion de l'ensemble des gènes constants du locus *IgH*. Ce type de recombinaison, nommée LSR pour locus suicide recombination, conduit à la mort des cellules B qui sont dans l'incapacité d'exprimer un BCR (Péron et al., 2012).

#### **IV.2.2.2. Fonction de la région 3'RR.**

L'activité des différents éléments composant la 3'RR a initialement été étudiée grâce à des gènes rapporteurs dans des modèles de lignées cellulaires représentatives d'un stade donné du développement B. L'activité individuelle de chacun des enhancers de la 3'RR est plus faible que celle de l'activateur intronique  $E\mu$ . Aux stades précoces, les enhancers hs1-2, hs3a et hs3b ne sont pas actifs. L'activité de ces 3 éléments est associée aux stades B tardifs (B mature et plasmocyte). Seul l'activateur hs4 est décrit comme étant actif dès le stade pro-B et jusqu'au stade plasmocytaire (Madisen and Groudine, 1994; Matthias and Baltimore, 1993; Michaelson et al., 1995). Par ailleurs, des effets synergiques entre les différents éléments composant la 3'RR ont été décrits : l'expression du gène rapporteur atteint son maximum lorsque celui-ci est sous la dépendance de l'enhancer hs1-2 encadré par les éléments hs3 et en configuration orientée inversée (Chauveau et al., 1998). Ce constat indique que la structure palindromique contribue activement à l'activité de la 3'RR. Les études suggèrent qu'aux stades précoces, seuls les activateurs  $E\mu$  et hs4 sont actifs. Aux stades suivants B mature et B activé, l'activité des enhancers de la 3'RR est prépondérante et augmente crescendo jusqu'au

Modèles d'étude	Recb. VDJ	SHM	CSR	Réf.	
<i>wt</i>					
hs1-2 KO		Pas de phénotype			Cogné et al., 1994 ; Manis et al., 1998
hs3a KO		Pas de phénotype			
hs3b KO		Pas de phénotype			
hs4 KO		Pas de phénotype			Vincent-Fabert et al., 2009
hs3b-4 KO		-	✓	∨∨∨ (sauf IgG1)	Pinaud et al., 2001 ; Le Morvan et al., 2003
3'RR KO		✓	∨∨∨	∨∨∨	Rouaud et al., 2012-2013 ; Vincent-Fabert et al., 2010 ; Saintamand et al., 2015a et b
PAL KO		-	∨∨∨	∨∨∨ (sauf IgGA et IgG2b)	Garot et al., 2016
ΔLeftPAL		-	∨∨∨	∨∨ (sauf IgGA)	Saintamand et al., 2016
ΔIRIS		-	∨∨	∨ (sauf IgGA et IgG2b)	Saintamand et al., 2016
C3'RR		-	∨∨∨	∨	Le Noir et al., 2017

**Table 3 : Modèles murins de délétion de la région 3'RR et conséquences sur les événements de recombinaisons VDJ, d'hypermutation somatique (SHM) et de recombinaison de classe (CSR).**



stade plasmocytaire. Cependant les modèles d'études cellulaires, caractéristiques d'un stade donné, ne reflètent pas forcément la complexité des événements qui ont lieu à l'échelle d'un organisme entier. Par la suite, des modèles de délétion (KO ou remplacement) des différents éléments composant la 3'RR (seul ou combiné) ont permis d'apprécier plus finement le rôle de la 3'RR au cours de la maturation B (Pinaud et al., 2011). Dans les paragraphes suivants, je décrirai les principaux modèles de souris élaborés dans le but d'élucider la fonction de la 3'RR (**Table 3**). L'activité de cette région étant maximale aux stades tardifs du développement B, plusieurs études se sont intéressées au rôle de la 3'RR lors des événements de remaniement secondaire (CSR et SHM) et ont démontré son action sur les événements de transcription associés.

#### **IV.2.2.2.1. Modèles de souris invalidant hs1-2 ou hs3a par insertion du gène *neo<sup>r</sup>* ou délétion précise.**

Chez la souris, le remplacement de l'enhancer hs1-2 par une cassette *pGK Neo<sup>r</sup>* entraîne plusieurs défauts : une altération de la CSR vers tous les isotypes (sauf IgG1) corrélée à une diminution de la transcription germinale des régions switch concernées (Cogné et al., 1994). Un phénotype similaire est observé lorsque le remplacement par la cassette *pGK neo<sup>r</sup>* concerne hs3a (Manis et al., 1998). Cependant, un modèle de délétion « propre » incluant l'élimination de la cassette *pGK neo<sup>r</sup>* (via la recombinaison CRE) invalide les conclusions précédentes : aucun effet sur la CSR n'est alors constaté (Manis et al., 1998), remettant en cause l'importance et suggérant la redondance des activateurs de la région 3'.

#### **IV.2.2.2.2. Modèles de souris déficientes pour hs3b.**

Les cellules B issues d'animaux déficients pour l'enhancer hs3b (délétion ciblée « propre », après suppression de la cassette *Neo<sup>r</sup>*) ne présentent pas de défaut de la transcription germinale ou de la CSR après stimulation au LPS *in vitro*. La diminution du taux sérique d'IgG2a et d'IgG1 constatée *in vivo* dans ce modèle pourrait résulter d'une différence de fond génétique entre les souris *wt* et mutantes (Bébin et al., 2010).

#### **IV.2.2.2.3. Modèles de souris déficientes pour hs4.**

Le modèle de souris porteuses d'une délétion de l'élément hs4 montre que cet activateur est dispensable pour le processus de CSR. Cependant la délétion de l'activateur hs4 entraîne une diminution de l'expression de la chaîne lourde Ig $\mu$  uniquement au stade B mature « au repos » ; ce défaut entraîne une diminution de l'expression du BCR à ce stade (Vincent-Fabert et al., 2009).

Le fait que l'élément hs4 soit actif dès le stade pré-B laissait présager un rôle de cet enhancer lors des recombinaisons VDJ. Les études des modèles murins KO incluant la délétion de l'enhancer hs4 ne montrent cependant aucune altération des recombinaisons VDJ, invalidant cette hypothèse (Morvan et al., 2003; Pinaud et al., 2001; Rouaud et al., 2012; Vincent-Fabert et al., 2009).



Les divers modèles de délétion individuelle des activateurs composant la région 3'RR (KO des éléments hs3a, hs1-2 ou hs3b) suggèrent des fonctions redondantes entre les différents éléments, seule la délétion de l'activateur hs4 révèle un rôle spécifique pour cet enhancer (pour l'expression de la chaîne lourde au stade B mature au repos).

Le mode d'action synergique des différents enhanceurs a pu être étudié à partir de modèles de souris porteuses de délétions combinées des éléments composant la région 3'RR.

#### **IV.2.2.2.4. Modèles de souris déficientes pour plusieurs éléments composants la région 3'RR.**

Le modèle de souris combinant la suppression des 2 derniers activateurs hs3b et hs4 (hs3b-4 KO) présente une caractéristique liée à l'absence de hs4 : la diminution de l'expression de l'IgM à la surface des cellules B matures et au repos. Alors que la délétion isolée des éléments hs3b ou hs4 n'affecte pas le mécanisme de CSR, leur suppression combinée perturbe drastiquement la recombinaison de classe. Le défaut de la production sérique de certaines classes d'Ig (IgG2a, IgG2b, IgG3 et IgE) observé dans ce modèle est corrélé à un défaut de CSR après stimulation *in vitro* vers tous les isotypes à l'exception d'IgG1. L'origine de ce défaut de CSR est un défaut de transcription germinale des régions switch acceptrices correspondantes (Pinaud et al., 2001). Une étude ultérieure montre que la suppression des 2 enhanceurs distaux n'affecte pas le processus de SHM (Morvan et al., 2003).

Le défaut de CSR constaté dans le modèle hs3b-4 KO est profondément aggravé dans un modèle de délétion complète de la région 3'RR (Vincent-Fabert et al., 2010). La stimulation *in vitro* de cellules B déficientes pour la totalité de la région 3'RR révèle une quasi abolition du mécanisme de CSR vers tous les isotypes et une diminution de la transcription germinale des régions switch acceptrices concernées et de la région donneuse S $\mu$  (Vincent-Fabert et al., 2010). Ce modèle, incluant également la suppression de hs4, présente la caractéristique commune d'une diminution de l'expression de la chaîne lourde Ig $\mu$  dans les cellules B matures au repos. Plus récemment, une étude précise le rôle de la 3'RR lors de la CSR : cette région régulatrice contrôle l'accessibilité des régions switch à AID (notamment acceptrices) lors de l'étape initiale de la CSR caractérisée par l'introduction des DSB (Saintamand et al., 2015).

Dans ce même modèle (3'RR KO), le processus de maturation d'affinité par SHM est également clairement affecté. La fréquence de mutation au locus *IgH* est quasi abrogée et la proportion de séquences mutées se trouve drastiquement diminuée dans les cellules du CG triées des plaques de Peyer d'animaux 3'RR KO. Dans ces cellules, qui sont normalement la cible des SHM, le recrutement de AID est très fortement diminué et corrélé à une diminution (de moitié) de la transcription primaire au locus *IgH* (initiée au promoteur pV<sub>H</sub> du segment réarrangé). Bien que l'étape de transcription constitue un prérequis pour le processus de SHM, la transcription à minima observée dans les cellules du CG des animaux 3'RR KO n'est à elle seule pas suffisante



pour conférer une accessibilité optimale du locus à AID (Rouaud et al., 2013). De façon attendue, la fréquence de mutation au locus des chaînes légères *Igκ* n'est pas affectée en l'absence de la région 3'RR.

Dans le but de déterminer les spécificités fonctionnelles liées à la structure quasi palindromique de la région 3'RR, de nouveaux modèles murins porteurs de diverses délétions ont été réalisés et récemment étudiés au laboratoire (**Figure 19b**).

Les études menées à partir du modèle PAL KO ont permis de préciser la fonction des 2 modules (proximal et distal) composant la région 3'RR. La présence du module distal (hs4) est suffisante au bon déroulement du développement B lors des étapes indépendantes de l'Ag alors que le module proximal (quasi palindrome incluant les trois premiers enhancers) est requis pour les événements de remaniement tardif au locus *IgH* (CSR et SHM) et pour la production massive d'Ac par les plasmocytes (Garot et al., 2016).

La transcription à distance des régions S à l'origine de l'activation de la CSR par la région 3' RR semble dépendant des éléments « core-enhancers » du module proximal. Un défaut de CSR est en effet une caractéristique commune aux modèles PAL KO, 3'RR KO, ΔleftPAL KO.

Le processus de SHM est quant à lui sévèrement affecté lorsque l'architecture palindromique de la région 3' RR est compromise. La délétion complète ou partielle de la région palindromique (PAL KO, ΔleftPAL et core 3'RR) entraîne une diminution significative de la fréquence de mutation et de la transcription primaire au locus *IgH*. Chez les animaux ΔIRIS et core 3'RR, la présence des core-enhancers (en l'absence de la structure palindromique) ne suffit pas à restaurer un niveau de mutation optimal (Garot et al., 2016; Le Noir et al., 2017; Saintamand et al., 2016).

### **IV.3. Les régions isolatrices au locus *IgH*.**

Les régions isolatrices sont des régions d'ADN servant de barrière entre des régions géniques qui doivent être régulées de façon indépendante (gènes, groupes de gènes ou/et éléments régulateurs de ces derniers). Ces régions délimitent des territoires au sein desquels les éléments régulateurs du groupe de gènes exercent leur fonction tout en prévenant l'action d'éléments régulateurs non appropriés. Les isolateurs peuvent bloquer l'action des enhancers (lorsqu'ils sont notamment localisés entre un enhancer et un promoteur) ou servir de barrière via la création de boucles (Gaszner and Felsenfeld, 2006; Kolovos et al., 2012; Liu et al., 2018).

#### **IV.3.1. La région en 5' du locus *IgH*.**

Les remaniements hautement contrôlés qui s'opèrent au locus *IgH* suggèrent la présence d'autres éléments régulateurs. De telles régions *cis*-régulatrices pourraient être localisées en 5' du locus *IgH* car préservées lors des événements de recombinaison. Une étude s'est intéressée à la région 5' en amont des segments V<sub>H</sub> et a mis en évidence la présence de 4 sites hypersensibles à la DNase I (nommés hs1, hs2, hs3a et hs3b) (**Figure 2**). Seul hs1 présente une hypersensibilité à la DNase I restreinte au stade pro B, le caractère hypersensible des 3 autres sites est observé tout au long du développement B (Pawlitzky et al., 2006). Cependant, l'élaboration de



modèles de souris porteurs d'une délétion du site *hs1* seul ou d'une délétion combinée des 4 sites *hs* ne révèle aucun défaut de développement B (Perlot et al., 2010). Cette région pourrait constituer la limite 5' du locus *IgH* ; elle pourrait participer au positionnement nucléaire du locus *IgH* (Perlot et al., 2010; Yang et al., 2005).

#### **IV.3.2. La région intergénique $V_H$ -D.**

L'étude de la région intergénique localisée entre les segments  $V_H$  et D a révélé la présence de 6 sites hypersensibles à la DNase (nommés *hs1* à *hs6*) (**Figure 2**). Les sites *hs4* et *hs5* contiennent chacun un site CBE (CTCF binding element) (Featherstone et al., 2010). La région intergénique semble jouer le rôle de région isolatrice en empêchant la transcription antisens des régions  $DJ_H$  de se propager vers la région  $V_H$ ; ces isolateurs participent au déroulement ordonné des réarrangements en limitant un accès prématuré des segments  $V_H$  (Featherstone et al., 2010). La délétion de la région intergénique entraîne une augmentation de la transcription germinale antisens  $DJ_H$  et provoque des réarrangements de type  $V_H$  vers  $DJ_H$  dans des thymocytes (Giallourakis et al., 2010). La délétion des 2 sites CBE confirme le rôle de la région intergénique dans la régulation de la transcription germinale et montre son implication dans la formation de boucles chromatiniennes (facilitant les réarrangements des segments  $V_H$  les plus distaux) (Guo et al., 2011a; Lin et al., 2015).

#### **IV.3.3. La région en 3' du locus *IgH*.**

En aval de la région 3'RR (environ 4 kb en aval de *hs4*) se situent 3 sites supplémentaires d'hypersensibilité à la DNase I (nommés *hs5*, *hs6* et *hs7*) (**Figure 2**) (Garrett et al., 2005). Ces régions de chromatine accessible possèdent les caractéristiques des isolateurs par la présence de sites CBE. Dans un modèle de souris, le KO de la portion d'ADN qui comprend les 3 sites *hs5*, *hs6* et *hs7* en aval de la 3'RR du locus *IgH* entraîne une augmentation de la transcription des gènes situés en aval, soulignant la fonction isolatrice de cette région. Cette étude révèle également un biais lors des recombinaisons VDJ associé à un usage préférentiel des segments  $V_H$  proximaux (Volpi et al., 2012). Ainsi la région isolatrice en 3' du locus *IgH* ne constitue pas un élément *cis*-activateur mais un isolateur transcriptionnel et intervient dans les interactions longues distances lors des étapes précoces du développement B (Birshtein, 2012; Degner et al., 2011a).

#### **IV.4. Régulation spatiale du locus *IgH*.**

Les événements de remaniement au locus *IgH* sont hautement ordonnés dans le temps à des stades de différenciation précis ; ils nécessitent également un contrôle dans l'espace du noyau de la cellule B. Au cours du développement B, le locus *IgH* va modifier sa position et s'organiser au sein du noyau en formant des boucles chromatiniennes qui vont guider des interactions longues distances. Ce niveau de régulation, spatial et conformationnel, est nécessaire au bon déroulement des nombreux remaniements qui s'opèrent au locus *IgH* (Hewitt et al., 2010).

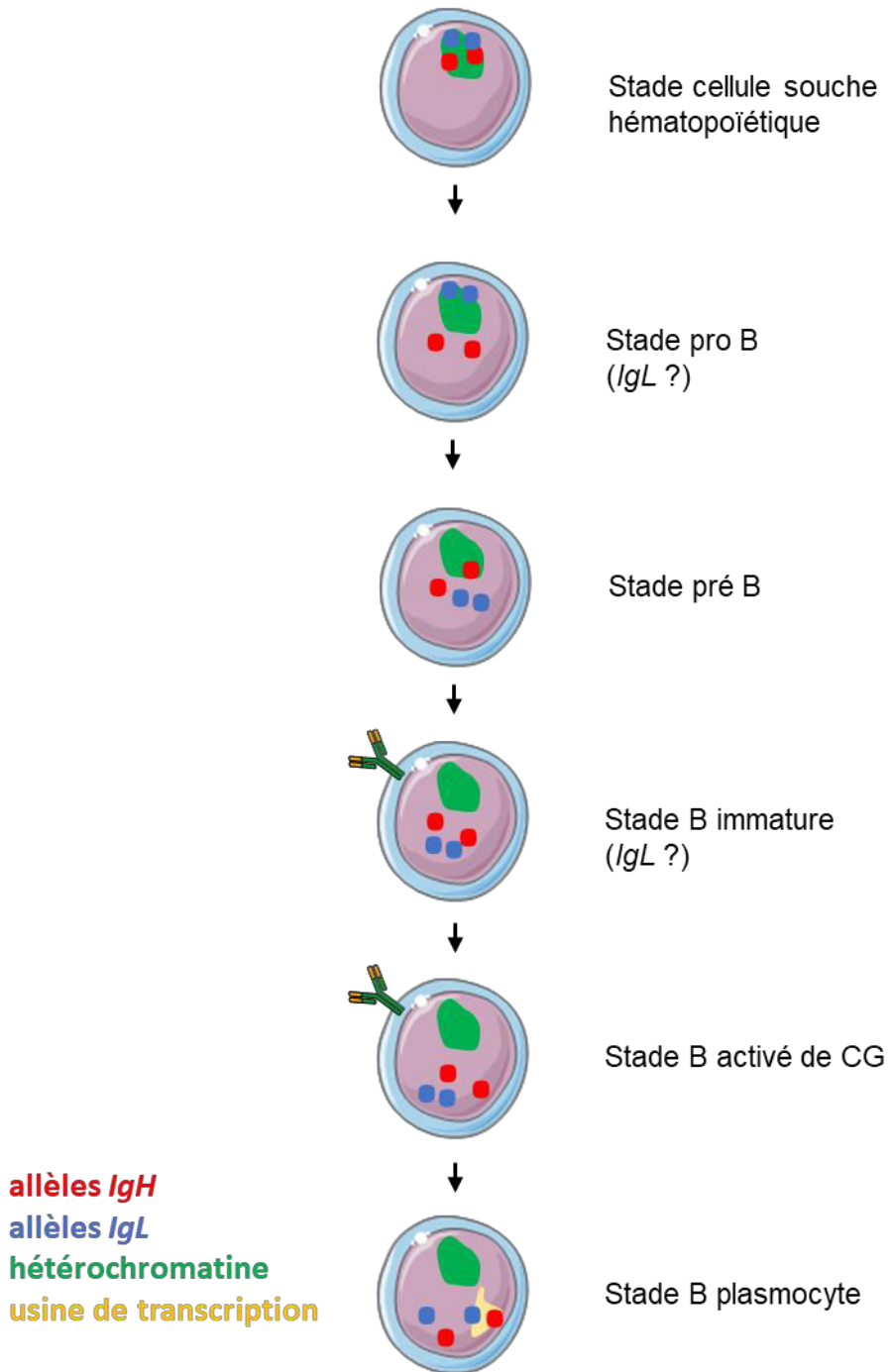


Figure 22 : Représentation schématique du positionnement des loci des gènes d'Ig au cours du développement B.

Pour certains stades de développement (indiqué par « *IgL* ? »), la position des loci *IgL* est à confirmer.



#### IV.4.1. Position des loci des gènes d'Ig.

Dans le noyau des cellules eucaryotes, l'ADN est organisé en association avec les protéines histones pour former la chromatine et selon différents niveaux de compaction qui dépendent des modifications post-traductionnelles des histones (Felsenfeld and Groudine, 2003; Lin and Murre, 2013). Classiquement deux types de chromatine sont décrits : l'euchromatine (EC) et l'hétérochromatine (HC). L'EC correspond à la chromatine sous forme peu condensée et dite transcriptionnellement active. Par opposition, l'HC fait référence à la chromatine sous forme condensée et peut être divisée en 2 types : facultative et constitutive. L'état de compaction de l'HC facultative varie en fonction du type cellulaire et du stade de différenciation. L'HC constitutive conserve son niveau de compaction et est principalement constituée de séquences répétitives (ou régions « satellites ») associées aux régions télomériques et péricentromériques (Ciabrelli and Cavalli, 2015; Kalousi and Soutoglou, 2016; Ritland Politz et al., 2016). Il existe également une portion d'HC enrichie à la périphérie du noyau et associée aux protéines de types lamines, ce type d'HC est nommée « HC null » (Ciabrelli and Cavalli, 2015). Les régions génomiques au contact de la lamine nucléaire sont appelées LAD (pour lamin associated domain) et sont enrichies en gènes peu actifs (van Steensel and Belmont, 2017).

La position des loci des gènes d'Ig n'est pas équivalente entre les cellules du lignage B et les cellules non B (Hewitt et al., 2010). En effet, les gènes d'Ig (*IgH* et *Igκ*) vont se localiser au niveau de la lamine nucléaire en périphérie du noyau dans des cellules non B. Alors que dans le noyau des cellules engagées dans la différenciation lymphocytaire B, la position des gènes d'Ig (*IgH* et *Igκ*) sera plus centrale et va évoluer en fonction du stade de développement (Kosak, 2002; Pichugin et al., 2017; Skok et al., 2001). Différentes études utilisant notamment la technique d'hybridation *in situ* en fluorescence et 3 dimensions (FISH-3D) ont permis de suivre la position des loci des gènes d'Ig dans les noyaux des cellules B à différents stades de développement.

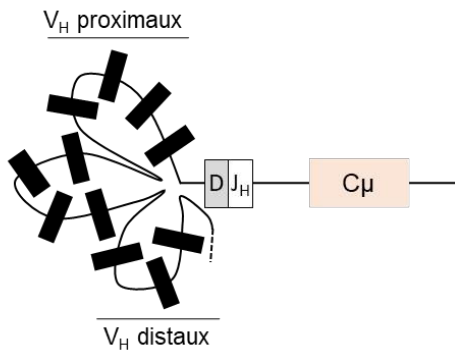
Dans les cellules souches hématopoïétiques, les 2 allèles *IgH* sont localisés dans l'hétérochromatine (compartiment « inactif »). L'engagement dans la voie de différenciation B est associé à des modifications du positionnement allélique (**Figure 22**). Au stade précoce (pré-pro et pro-B), lorsque les réarrangements de type D vers  $J_H$  s'opèrent de façon simultanée sur les 2 allèles *IgH*, ceux-ci ne se trouvent plus localisés à l'HC. Dans environ 20% des noyaux des cellules à ce stade, les 2 allèles *IgH* se trouvent appariés (la distance séparant les 2 allèles est inférieure à  $1\mu\text{m}$ ) ; cet appariement bi-allélique est dépendant de l'expression de l'enzyme RAG1 (Hewitt et al., 2009). Au stade suivant pré-B, au cours duquel a eu lieu les réarrangements de  $V_H$  vers  $DJ_H$ , la position des 2 allèles *IgH* devient asymétrique et inclut un allèle localisé dans l'hétérochromatine. L'hypothèse proposée serait que la relocalisation d'un allèle *IgH* vers un compartiment nucléaire inactif évite un remaniement de  $V_H$  vers  $DJ_H$  simultané des 2 allèles, expliquant la séquence monoallélique de ce type de réarrangement au locus *IgH*. (Roldán et al., 2005). La position non équivalente des allèles *IgH* semble maintenue aux stades suivants lors des réarrangements des chaînes légères d'Ig.



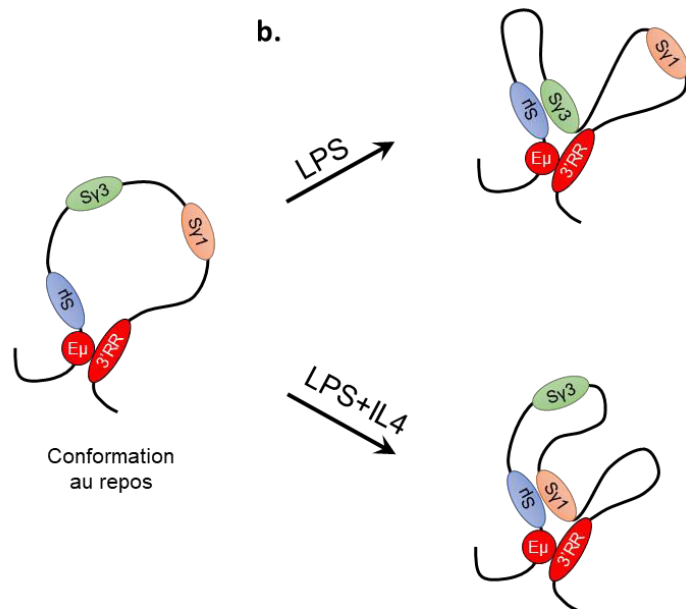
De façon similaire, la position des allèles *Igκ* est modifiée en fonction du stade de développement B. Au stade précoce (pro-B), les 2 allèles *Igκ* semblent se relocaliser de la périphérie vers le centre du noyau, en dehors de l'hétérochromatine (Goldmit et al., 2005; Kosak, 2002). Au stade pré-B, un allèle *Igκ* se localise dans l'hétérochromatine de façon à ne rendre qu'un seul allèle accessible aux réarrangements (Fitzsimmons et al., 2007; Goldmit et al., 2005). D'autres études montrent que la position nucléaire de gènes essentiels à la progression dans le lignage varie également dans le noyau des cellules B en développement. Pour exemple, le gène *Ebf1* est relocalisé en dehors des LAD lors de la transition entre les stades pré-pro-B et pro-B (Lin et al., 2012). Si toutes les études s'accordent sur l'existence d'une régulation spatiale des gènes d'Ig au sein du noyau des cellules B, certaines descriptions divergent selon les études. Pour exemple, la relocalisation des allèles *Igκ* vers le centre du noyau dans les cellules engagées dans le lignage B dès le stade-pro B (Kosak, 2002) ne semble pas être confirmée par une étude plus récente indiquant que les allèles *Igκ* sont relocalisées au centre du noyau uniquement à partir du stade pré-B (Rother et al., 2016).

La majorité des études de positionnement nucléaire ont décrit les positions des gènes d'Ig lors des stades précoces du développement B (en relation avec les recombinaisons V(D)J). Les données sur les stades suivants montrent que, au stade immature tardif, les 2 allèles *IgH* se trouvent à nouveau positionnés en dehors de l'hétérochromatine (Fuxa et al., 2004; Hewitt et al., 2009; Kosak, 2002; Roldán et al., 2005). Dans le noyau des cellules B périphériques, les 2 allèles *IgH* doivent être théoriquement accessibles aux remaniements secondaires (SHM et CSR) (Delpy et al., 2003). Bien que des études aient suggéré une position non équivalente des allèles *IgH* dans les noyaux des cellules B matures après stimulation *in vitro* (Skok et al., 2001), de nouvelles évidences montrent que les 2 allèles se trouvent localisés dans un compartiment nucléaire actif (Holwerda et al., 2013). Ces observations corrént avec d'autres études montrant que les 2 allèles *IgH* (productif et non productif) sont transcrits dans les cellules B ayant terminé leurs réarrangements (Daly et al., 2007). Une autre étude récente montre, dans les noyaux de cellules B humaines en cours de maturation, que les 2 allèles *IgH* ne sont pas être associés à l'hétérochromatine (Pichugin et al., 2017). Cette dernière étude propose également, lors de la différenciation B tardive, que les loci d'Ig se relocalisent dans un compartiment à proximité du nucléole qui pourrait constituer une zone de recrutement de l'enzyme AID. L'équipe de Garrard montre que les gènes des Ig sont préférentiellement localisés en périphérie du noyau au stade plasmocyte. A ce stade, bien que localisés sur différents chromosomes, les loci des chaînes lourdes et légères ainsi que les loci codant la chaîne J sont associés en « cluster » et partagent la même usine de transcription (Park et al., 2014). En accord avec la position des gènes dans le noyau, d'autres études soulignent l'importance de la relocalisation nucléaire après des dommages causés à l'ADN. Des lésions de type cassures double brin (telles que les cassures induites lors de la CSR), localisées à la périphérie du noyau vont être prises en charge préférentiellement par la voie du NHEJ (Lemaître et al., 2014).

a.



b.



**Figure 23 : Exemples de topologie adoptée par le locus *IgH*.**

a. A l'issue des recombinaisons D-J<sub>H</sub>, le locus se contracte permettant le rapprochement des segments V<sub>H</sub> distaux vers le segment DJ<sub>H</sub> réarrangé.

b. Lors de la CSR, après stimulation appropriée, la région switch activée se rapproche de la région S<sub>μ</sub> donneuse, facilitant le processus de recombinaisons (d'après Wuerffel *et al.*, 2007).

#### IV.4.2. Topologie tridimensionnelle du locus *IgH*.

Il est admis que la probabilité de recombinaisons entre 2 régions dépend de la distance qui les sépare. Les évènements de recombinaison prenant place dans un locus aussi étendu que le locus *IgH* suggère qu'il existe un niveau de régulation conformationnelle guidant des interaction longues distances capables de mettre à proximité les régions et éléments qui doivent recombiner. Des études utilisant des techniques de FISH et plus récemment de capture de conformation chromosomique (3C et 4C) ont permis de mettre en évidence des modifications dynamiques de topologie du locus *IgH* au cours de ses différents remaniements.

Les changements de conformation au locus *IgH* sont corrélés à la formation de boucles chromatinienne. Celles-ci contribuent à rapprocher les éléments à recombiner au sein d'une même boucle alors que les éléments non impliqués dans la recombinaison sont isolés dans d'autres boucles (Guo et al., 2011b; Jhunjhunwala et al., 2008). Lors des recombinaisons de type D vers J<sub>H</sub>, les régions D-J<sub>H</sub>-E<sub>μ</sub> sont incluses dans une même boucle et à l'écart des segments V<sub>H</sub> (Degner et al., 2011b; Guo et al., 2011a). Lors des recombinaisons suivantes de type V<sub>H</sub> vers DJ<sub>H</sub>, le locus « se contracte » à nouveau, favorisant la proximité entre des groupes de segments V<sub>H</sub> et le segment DJ<sub>H</sub> réarrangé (Fuxa et al., 2004; Guo et al., 2011b; Medvedovic et al., 2013). La contraction du locus permet aux segments V<sub>H</sub> proximaux (proches des régions D<sub>H</sub> et J<sub>H</sub>) mais aussi distaux (les plus éloignés, situés en 5' du locus) d'être à une distance équivalente du segment DJ<sub>H</sub> permettant un usage relativement homogène des différents segments V<sub>H</sub> (Cobaleda et al., 2007; Jhunjhunwala et al., 2008) (**Figure 23a**). Lorsque la recombinaison VDJ est productive, la signalisation induite par le pré BCR facilite la décontraction du locus *IgH* (Roldán et al., 2005).

La topologie du locus *IgH* va également être modifiée lors des évènements de CSR. Pour exemple, des études rapportent un repliement du locus *IgH* suite à une stimulation cytokinique appropriée de façon à rapprocher les régions S donneuse et acceptrice et les régions régulatrices E<sub>μ</sub> et 3'RR (Feldman et al., 2015; Kenter et al., 2012; Wuerffel et al., 2007) (**Figure 23b**).



## *Résultats*





Le contrôle du processus de SHM est complexe et fait intervenir différents niveaux de régulation impliquant un ensemble de facteurs *cis* (les régions régulatrices aux loci des gènes d'Ig) et *trans* interconnectés (la machinerie de transcription, AID et ses partenaires, les facteurs de la réparation...). L'ensemble des éléments contribuant à la régulation spatio-temporelle de la SHM sont jusqu'à présent mal connus et leurs actions concertées posent de nombreuses d'interrogations.

L'étape de transcription est un prérequis essentiel au processus de SHM en favorisant vraisemblablement l'exposition de l'ADN sous forme simple brin à la machinerie de mutation. Cependant l'ensemble des gènes transcrits au stade cellules B du CG ne sont pas ciblés par AID : bien qu'essentielle la transcription n'est pas suffisante pour initier l'introduction des mutations. Par ailleurs, l'action de AID est soumise à des régulations dynamiques telles que des modifications post traductionnelles et sa translocation au noyau des cellules B activées. La fonction de AID en tant qu'enzyme modifiant l'ADN représente une menace pour l'intégrité du génome, pourtant l'action de AID n'est pas strictement restreinte aux gènes d'Ig puisque des gènes non Ig (ou cibles illégitimes) se trouvent mutés dans les noyaux des cellules B de CG. Les questions qui ont motivé nos études sont les suivantes : Comment le recrutement de AID vers ses cibles légitimes et illégitimes est-il médié ? Pourquoi certains gènes dont des oncogènes sont-ils plus propices à être ciblés par AID ? Quels sont les facteurs responsables de la prise en charge par la réparation fidèle ou infidèle suite à l'action de AID ?

Les régions *cis*-régulatrices au locus *IgH*, en particulier la région 3'RR, sont capables de moduler la SHM en agissant conjointement sur les mécanismes de transcription et de recrutement de la machinerie de mutation (Garot et al., 2016; Le Noir et al., 2017; Rouaud et al., 2013; Saintamand et al., 2016). Concernant la région  $E\mu$ , bien que l'activateur  $cE\mu$  ne soit pas nécessaire au bon déroulement de la SHM (Perlot et al., 2005), l'implication potentielle des régions  $MARs_{E\mu}$  dans ce processus demeurerait une hypothèse censée qui méritait d'être précisément étudiée. Mon travail de thèse s'est intéressé à l'étude du mécanisme de SHM dans les cellules B issues de modèles murins porteurs d'une délétion des régions  $MARs_{E\mu}$  encadrant l'activateur intronique  $cE\mu$  (modèle créé avant mon arrivée au laboratoire). La suppression des  $MARs_{E\mu}$  n'interfère pas avec le développement B, l'étude de la SHM dans ce modèle peut donc être menée à partir de cellules B de CG triées directement issues des animaux  $MARs_{E\mu}$  KO.

Au cours de ma thèse, j'ai confirmé la contribution des régions  $MARs_{E\mu}$  au processus de SHM au locus *IgH*. Nous avons proposé un mécanisme moléculaire pouvant expliquer le mode d'action des  $MARs_{E\mu}$  dans l'optimisation de la SHM : cette étude est présentée dans un article soumis (auquel je contribue comme 1<sup>er</sup> auteur : **article 2**).

Notre étude de la fonction des  $MARs_{E\mu}$  s'est poursuivie au-delà du locus *IgH* (sur le chromosome 12), et de façon surprenante, notre équipe a montré que le taux d'hypermutation au locus des chaînes légères  $Ig\kappa$  (sur le chromosome 6) est également diminué en l'absence de ces régions *cis*-régulatrices situées sur un autre



chromosome. Ces résultats ont soulevé bon nombre d'interrogations parmi lesquelles une possible influence des MARs<sub>Eμ</sub> sur les mutations infligées aux cibles illégitimes de AID. La fréquence de mutation de ces cibles illégitimes pouvant être cent fois plus faible que celle des gènes d'Ig, nous avons tout d'abord développé un nouvel outil bio-informatique spécifiquement dédié à l'analyse des mutations rares à partir de données de séquençages haut débit. Les performances de cet outil, nommé DeMinEr (Detection of Minor variant by Error correction), ont fait l'objet d'une publication dans « The Journal of Immunology » (à laquelle je contribue comme 1<sup>er</sup> auteur : **article 1**). A l'aide de DeMinEr, nous montrons pour la 1<sup>ère</sup> fois que les régions MARs<sub>Eμ</sub> situées sur le chromosome 12 influencent significativement les mutations « en *trans* » dans les cellules B du CG : à la fois au locus des chaînes légères Igκ mais aussi sur 2 des cibles illégitimes de AID situées sur d'autres chromosomes.

Nous proposons également que le processus de SHM soit régulé « spatialement » dans les noyaux des lymphocytes B grâce aux régions MARs<sub>Eμ</sub>. Cette hypothèse a été étudiée par une technique d'hybridation *in situ* en fluorescence et 3 dimensions (FISH 3D) et d'acquisition étagée, nous permettant d'analyser la position relative des gènes cibles de AID (légitimes et illégitimes) dans des noyaux de cellules B dont l'architecture tridimensionnelle a été préservée. Notre étude révèle que les régions MARs<sub>Eμ</sub> influencent la position relative des gènes d'Ig et « off targets », en faveur d'une régulation « spatiale » du processus de SHM. Ces résultats seront décrits dans la partie résultat : **article 3**. Les données préliminaires de mutations des gènes d'Ig précédemment obtenues au laboratoire ont été complétées par séquençage haut débit à mon arrivée. J'ai également contribué à l'étude du positionnement nucléaire des gènes d'Ig dans les modèles animaux. J'ai réalisé l'ensemble des travaux (mutations, position des allèles, analyse des transcrits) concernant les cibles « off targets » dans nos modèles.

Le mode d'action des régions MARs<sub>Eμ</sub>, en *cis* (au locus *IgH*) comme en *trans* (au locus Igκ et aux « off targets »), implique vraisemblablement les protéines interagissant avec ces régions (MAR binding proteins ou MAR-BP). Le potentiel rôle des MAR-BP dans le mécanisme de SHM n'a jusqu'alors jamais été exploré, probablement à cause de la létalité associée au KO de la plupart de ces protéines chez l'animal.

A partir de nos données de RNA-seq, nous avons observé que certaines MAR-BP sont soit surexprimées (ARID3A, ARID3B et CUX1) ou sous-exprimées (SATB1) entre les stades B au repos et du CG. Ces variations d'expression au cours de l'activation des cellules B dans le CG font de ces protéines des candidats régulateurs potentiels du processus de SHM. Nous avons donc mis en place divers modèles destinés à une étude *in vivo* de la fonction de ces protéines : 2 modèles de KO conditionnels, plusieurs modèles de knock down (KD) et envisageons des modèles de surexpression. Ces approches sont décrites plus en détails dans la partie discussion/perspectives.



Durant ma thèse, j'ai également réalisé des expériences de FISH 3D dans le but d'étudier le positionnement du locus *IgH* à l'hétérochromatine dans des noyaux de cellules B à différents stades (B au repos, B activés *in vitro*, B activés de CG *ex vivo* et plasmocytes). Ces résultats ont été intégrés à une étude récemment publiée par le groupe de Laurent Delpy au laboratoire (annexe : voir **article 1**).

Au cours d'une collaboration, j'ai apporté ma contribution pour la mise au point d'un essai de qPCR visant à quantifier la présence de cellules tumorales d'origine humaines dans des modèles de souris greffées mimant une tumeur colorectale (annexe : **voir article 2**).

Par ailleurs, avant-même la publication de notre outil bio-informatique DeMinEr, nous avons mis à profit nos compétences pour la préparation, le séquençage haut débit et l'analyse des mutations rares dans des cibles « off targets » de AID dans une étude collaborative publiée par le groupe de Uttiya Basu s'intéressant au rôle du RNA exosome et de ses partenaires dans la distribution des mutations (annexe : **voir article 3**).



## Article 1

### Detecting rare AID-induced mutations in B-lineage oncogenes from high-throughput sequencing data using DeMinEr.

Ophélie Alyssa Martin, Armand Garot, Sandrine Le Noir, Jean-Claude Aldigier, Michel Cogné, Eric Pinaud and Francois Boyer.

*The Journal of Immunology*, DOI: 10.4049/jimmunol.1800203

Comme mentionné précédemment (chapitre III. : « AID : enzyme clé du développement B tardif. »), l'enzyme AID, à l'origine des mutations introduites aux gènes d'Ig dans le but d'améliorer l'affinité des Ig pour les Ag, peut également agir en marge des gènes d'Ig au niveau de cibles illégitimes subissant des mutations tels que des dommages collatéraux induits par AID.

Parmi les cibles illégitimes, les gènes *Bcl6* et *Cd83* (appartenant aux gènes du groupe 1 selon l'étude menée par le groupe de Schatz) semblent être préférentiellement ciblés par AID et se trouvent les plus mutés dans un contexte *wt*. Cependant, la fréquence de mutation de ces cibles illégitimes reste bien inférieure à celle observée aux gènes d'Ig (au moins 10 fois inférieure). Pour les autres cibles illégitimes identifiées (*Pim1*, *Btg1*, *c-myc*...) la fréquence de mutation est encore moindre.

Ces dernières années les techniques de séquençage ont largement évolué : le séquençage nouvelle génération (next generation sequencing, NGS) à haut débit vient peu à peu remplacer le séquençage Sanger classique. La rapidité d'obtention et la quantité des données générées par NGS, permettant de réaliser plus facilement des tests de significativité, font du NGS une méthode largement utilisée en routine pour la recherche comme pour le diagnostic médical.

Nous avons utilisé la technologie haut débit Ion Proton pour le séquençage de 4 cibles illégitimes de AID (*Bcl6*, *Cd83*, *Pim1* et *Btg1*) à partir d'ADN issu de cellules B de CG triées d'animaux *wt*, déficients pour AID ou pour UNG. La fréquence de mutation obtenue (dans les gènes cibles cités précédemment) est théoriquement inexistante dans des cellules B du CG d'animaux déficients pour AID. Le profil de mutation associé à la délétion de UNG étant bien caractérisé (augmentation des mutations aux bases C et G avec une diminution des mutations de type transversions), nous avons utilisé ce modèle pour valider nos données de séquençage sur le plan qualitatif (nature des substitutions).

Le séquençage haut débit dit « ultra deep » (consistant à lire un grand nombre de fois une région génomique particulière) est cependant associé à un bruit de fond non négligeable qui limite l'étude des mutations rares. En effet la fréquence de mutation « basale » engendrée par le bruit de fond peut couvrir la fréquence de mutation d'évènements peu représentés. Il est communément admis de filtrer les données de séquençage haut débit et de ne considérer que les évènements de substitution dont la fréquence de mutation est





supérieure à 1%. Le taux de mutation des cibles illégitimes de AID se situe autour de 0,1‰, les filtres usuels ne sont donc pas adaptés pour l'identification de mutations rares, telles que celle initiées par AID aux niveaux des gènes « off targets ».

Nous avons alors développé un nouvel outil (DeMinEr) dédié à l'identification des mutations rares et basé sur l'utilisation d'un échantillon contrôle pour lequel les mutations observées seront associées au bruit de fond. Les échantillons déficients pour AID ont été utilisés comme contrôle ; une alternative consiste à utiliser des échantillons provenant de tissus non lymphoïdes B dans lesquels la SHM est inexistante. Les fréquences de mutation obtenues pour chaque base et à chaque position de la région séquencée dans les échantillons « tests » sont comparées à celles obtenues dans un contexte déficient pour AID. Le filtre appliqué permet, pour chaque position, de déterminer si une mutation est réellement induite par AID lorsque la fréquence d'insertion de la base mutée est significativement supérieure à celle obtenue pour la même position dans l'échantillon contrôle AID KO. Le filtre que nous appliquons est affiné par plusieurs paramètres notamment le nombre de cellules utilisées initialement dans l'échantillon et la profondeur de lecture. Nos résultats montrent que le filtrage par DeMinEr révèle des taux de mutation dans les cibles *Bcl6*, *Cd83*, *Pim1* et *Btg1* en accord avec les considérations biologiques précédemment publiées. Sur le plan qualitatif, les profils de mutation de ces gènes cibles obtenus dans un contexte UNG KO, sont tout à fait caractéristiques d'une altération de la voie de réparation BER et valident la fiabilité de notre outil.



## Article 2

### **Antisense transcripts driven by $E\mu$ -matrix attachment regions optimize somatic hypermutation upstream from the enhancer.**

Ophélie Martin, Marie Marquet, Armand Garot, François Boyer, Mylène Brousse, Sébastien Bender, Claire Carrion, Jean Claude Aldigier, Patricia J. Gearhart, Robert Maul, Laurent Delpy, Michel Cogné, Sandrine Le Noir and Eric Pinaud.

Dans le but de préciser le rôle des régions  $MAR_{E\mu}$  et leur possible implication dans le processus de SHM, un modèle murin porteur d'une délétion des 2 régions d'attachement à la matrice nucléaire encadrant l'activateur  $cE\mu$  a été élaboré dans le laboratoire. Si des études antérieures n'avaient pas permis d'élucider clairement l'implication des  $MAR_{E\mu}$  dans le mécanisme de SHM, nous montrons que ces régions sont des éléments *cis*-régulateurs nécessaires au bon déroulement du processus de maturation d'affinité.

La délétion des régions  $MAR_{E\mu}$  n'interfère pas avec le développement B, les différentes sous populations médullaires et périphériques sont normales chez les animaux mutants (seule une légère augmentation du nombre de cellules B du CG est observée). Le mécanisme de CSR et la production d'Ac sécrétés ne sont pas affectés chez les souris  $MAR_{E\mu}$  KO. En revanche, en l'absence des régions  $MAR_{E\mu}$ , la fréquence des mutations au locus *IgH* se trouve perturbée : diminuée en aval des segments réarrangés (introns  $J_{H3}$  et  $J_{H4}$ ) et augmentée dans la région située juste en aval du core enhancer  $cE\mu$ . La perte des régions  $MAR_{E\mu}$  peut être assimilée à la perte d'une « barrière » limitant et optimisant la machinerie de SHM à la région située en amont de l'enhancer  $cE\mu$ . La transcription étant un prérequis au mécanisme de mutation, nous avons quantifié avec précision les transcrits en amont de la région  $E\mu$  au locus *IgH*. Bien que la suppression des  $MAR_{E\mu}$  n'affecte pas la transcription globale, les transcrits antisens sont moins abondants dans les cellules B mutantes. Une diminution de la transcription antisens est compatible avec une diminution de la SHM selon l'hypothèse proposée par l'équipe de Alt indiquant que les zones de transcription convergente recrutent AID. D'après cette hypothèse, nous avons étudié le recrutement de AID en amont de la région  $E\mu$  au locus *IgH* (par immunoprécipitation de la chromatine, ChIP) et montré que l'absence des  $MAR_{E\mu}$  conduit à une diminution de l'abondance de cette déaminase. L'ensemble des expérimentations de quantification des transcrits et de ChIP ont également été menées dans des cellules B stimulées *in vitro*, ne subissant pas de SHM, pour lesquelles ni la transcription antisens ni le recrutement de AID n'est affecté par la délétion des  $MAR_{E\mu}$ . Ces résultats confirment l'hypothèse selon laquelle le défaut de transcription antisens est bien en lien avec le défaut de SHM.



## **Antisense transcripts driven by $E\mu$ -matrix attachment regions optimize somatic hypermutation upstream from the enhancer**

Ophélie Martin<sup>1,5</sup>, Marie Marquet<sup>1,5</sup>, Armand Garot<sup>1</sup>, François Boyer<sup>1</sup>, Mylène Brousse<sup>1</sup> Sébastien Bender<sup>1,2</sup>, Claire Carrion<sup>1</sup>, Jean Claude Aldigier<sup>1,2</sup>, Patricia J. Gearhart<sup>3</sup>, Robert Maul<sup>3</sup>, Laurent Delpy<sup>1</sup>, Michel Cogné<sup>1,2,4</sup>, Sandrine Le Noir<sup>1,5\*</sup> and Eric Pinaud<sup>1\*</sup>

<sup>1</sup>*UMR 7276, Centre National de la Recherche Scientifique and Université de Limoges : Contrôle de la Réponse Immune B et des Lymphoproliférations, 2, rue du Dr. Marcland, 87025 Limoges, France*

<sup>2</sup>*Centre Hospitalier Universitaire Dupuytren, 2 Avenue Martin Luther King, F-87042 Limoges, France*

<sup>3</sup>*Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224*

<sup>4</sup>*Institut Universitaire de France, Université de Limoges, 87000 Limoges, France*

<sup>5</sup>O.M., M.M. and S.L.N. contributed equally to this work

\*Corresponding authors: [eric.pinaud@unilim.fr](mailto:eric.pinaud@unilim.fr) and [sandrine.le-noir@unilim.fr](mailto:sandrine.le-noir@unilim.fr)

Correspondence and requests should be addressed to E.P. ([eric.pinaud@unilim.fr](mailto:eric.pinaud@unilim.fr)) and S.L.N. ([sandrine.le-noir@unilim.fr](mailto:sandrine.le-noir@unilim.fr))



## Abstract

Two scaffold/matrix attachment regions (5'- and 3'- $MAR_{E\mu}$ ) flank the intronic core enhancer ( $cE\mu$ ) within the immunoglobulin heavy chain locus ( $IgH$ ). Besides their conservation in mice and humans, the physiological role of  $MAR_{E\mu}$  is still unclear. We performed endogenous deletion of  $MAR_{E\mu}$  regions in mice and specified its significance in optimizing somatic hypermutation (SHM) targeting. In germinal centre B cells,  $MAR_{E\mu}$  deletion decreased SHM frequency in the intronic region downstream from the variable exon rearranged to  $J_{H3}$  and  $J_{H4}$ . This defect was correlated with a decrease of AID recruitment and RNA pol II elongating form, a process that we assigned to a defect of antisense transcription towards the variable exon. We concluded that, in addition to the required sense transcription,  $MAR_{E\mu}$  favour the occurrence of antisense RNAs upstream from the intronic enhancer in order to get the optimal convergent transcription balance required for efficient SHM targeting.





The IgH locus, encoding the immunoglobulin heavy chain, is among the most complex in mammals, with multiple *cis*-regulatory elements controlling stepwise DNA accessibility to recombination and mutation through mechanisms that mainly rely on transcription (Perlot and Alt, 2008). Current studies of the dynamic processes that regulate chromatin conformation changes and subnuclear location have renewed interest in *cis*-regulatory regions that delimit differentially regulated chromosomal domains. Among such DNA regulatory regions, nuclear Scaffold/Matrix Attachment Regions (MARs) have been implicated in the structural and functional organization of these domains. The juxtaposition of MARs to intronic enhancer elements in both *IgH* and *IgL* loci and their conservation in humans, mice and rabbits (Scheuermann and Garrard, 1999) suggest that such regions serve physiological functions. They participate in the regulation of gene expression notably by increasing enhancer function and facilitating their action over large distances. Several proteins found to bind MARs are expressed ubiquitously or in a tissue-specific manner, respectively defining constitutive or facultative MARs (Gluch et al., 2008). Once attached to the nuclear matrix in a tissue specific fashion, facultative MARs could form topological barriers that could isolate or fasten chromatin regions (Gluch et al., 2008). Such barriers could induce DNA torsional strain with positive and negative DNA supercoiling, respectively, upstream and downstream from the RNA pol II-induced transcription bubble (Teves and Henikoff, 2014). The supercoils are then released by the action of dedicated topoisomerases (Pommier et al., 2016).

The *IgH E $\mu$*  enhancer region is a combination of both the core *E $\mu$*  (*cE $\mu$* ) enhancer element (220 bp) and two 310–350-bp flanking MARs (*MARs<sub>E $\mu$</sub>* ) that were first defined by *in vitro* matrix-binding assays (Cockerill et al., 1987). This region, especially *cE $\mu$* , controls early VDJ recombination events (Afshar et al., 2006; Perlot et al., 2005) and is also involved in Ig  $\mu$  chain expression in pre-B cells (Marquet et al., 2014). However, its role in SHM remains unclear. An elegant model of deletion in the endogenous *E $\mu$*  region of hybridoma cells, enforced for human AID expression, suggested the requirement of *cE $\mu$*  and a substantial function of *MARs<sub>E $\mu$</sub>*  for SHM (Ronai et al., 2005). Similarly, when added to transgenes, *cE $\mu$*  and its flanking MARs contribute to Ig  $\mu$  chain expression and high levels of SHM (Azuma et al., 1993; Bachl and Wabl, 1996; Giusti and Manser, 1993; Lin et al., 1998; Motoyama et al., 1994; Ronai et al., 1999). In contrast, knock out (KO) models underlined the complexity of its physiological regulation. In a mouse model carrying the pre-rearranged VB1-8A region, *E $\mu$*  deletion still resulted in a high rate of SHM in Peyer's patch GC B cells, arguing for a non-essential role of the enhancer (Li et al., 2010). More clearly, deletion of *cE $\mu$*  in the mouse germline did not reduce SHM frequency but only slightly increased the proportion of unmutated alleles (likely due to the reduced inflow of peripheral and, consequently, GC B cells in this model) (Perlot et al., 2005). Strikingly, the role of *MARs<sub>E $\mu$</sub>*  was also elusive and somewhat controversial. Whereas their endogenous deletion, analysed in mouse chimeras by the RAG-2 complementation assay, demonstrated that *MARs<sub>E $\mu$</sub>*  are dispensable for VDJ recombination and IgH expression (Sakai et al., 1999a), the ambiguous function of *MARs<sub>E $\mu$</sub>*  was sustained by



the discrepancy between their ability to either bind negative regulatory factors (Kohwi-Shigematsu et al., 1997; Wang et al., 1999), improve  $cE\mu$  enhancer efficiency (Kaplan et al., 2001), or substitute for  $cE\mu$  to maintain IgH expression (Wiersma et al., 1999). At the  $\kappa$  light chain locus, the intronic enhancer  $Ei\kappa$  region also contains an upstream MAR. The implication of  $MAR_{Ei\kappa}$  as an enhancer of SHM was first suggested in transgenic studies (Goyenechea et al., 1997) and then tested in the mouse KO model that accumulated premature light chain rearrangements with a mild SHM defect (Yi et al., 1999), an effect comparable to one observed at  $IgH$  in hybridoma cells devoid of  $MAR_{SE\mu}$  (Ronai et al., 2005). The potential function for  $MAR_{Ei\kappa}$  in modulating SHM was later contradicted by a study comparing  $3'E\kappa$ - and  $MAR_{Ei\kappa}$  KO models (Inlay et al., 2006).

To address the controversy over the role of the scaffold in SHM, we generated a mouse model carrying a germline deletion of  $MAR_{SE\mu}$ , and their function was assessed by analyzing SHM-related events upstream and downstream from the intronic enhancer in GC B cells. We note that  $MAR_{SE\mu}$  are able to focus and optimize SHM events upstream from the  $E\mu$  region by promoting antisense transcription towards the rearranged segment; suggesting that convergent transcription can recruit AID for SHM.



## RESULTS

### Normal B cell development and Ig production in the absence of *MARs<sub>E $\mu$</sub>*

We generated a mouse mutant line carrying an endogenous deletion of both the 5' and 3' *IgH* matrix attachment regions that flank the *J<sub>H</sub>-C<sub>H</sub>* intronic *cE $\mu$*  enhancer. Although generated with slightly different targeting vector backbone and homology arms, the resulting *IgH* allele, so-called *MARs<sub>E $\mu$</sub> <sup>Δ</sup>* (Fig. 1a), is similar to that generated by Sakai et al. (Sakai et al., 1999a). Bone marrow subsets of B cell precursors were analysed in *wt* and homozygous *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* deficient mice. When compared to age-matched *wt* animals, *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* mice exhibited normal proportions and numbers of pre-proB, pro-B and pre-B cell precursors (Supplementary Table 1). Unlike endogenous deletion of the entire *E $\mu$*  region (Marquet et al., 2014), *MARs<sub>E $\mu$</sub>*  deletion did not modify Ig  $\mu$  heavy chain expression in early B lineage cells since proportions of IgM-expressing bone marrow B cell populations (immature, transitional and mature recirculating B cell subsets) were comparable to those of *wt* (Supplementary Table 1). Mature B cell subsets were also similar to *wt* in the spleen and peritoneal cavity of homozygous *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* mutants (Supplementary Table 1). In agreement with the normal inflow of mature B cells in *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* animals, Peyer's patches were efficiently colonized by centroblasts and GC B cells. Numbers and proportions of GC B cells were even significantly increased in homozygous mutants (Fig. 1b, bottom panel and Supplementary Table 1). The similar proportion of proliferating KI67<sup>+</sup> GC B cells in *wt* and *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* mice implied that this increase was not due to over-proliferation of Peyer's patch B cells (Fig. 1b). Finally, levels of serum Ig isotypes were unaffected in *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* animals (Fig. 1c). This normal B cell homeostasis in homozygous mutants confirmed that *MARs<sub>E $\mu$</sub>*  are dispensable for B cell ontogeny, CSR and antibody production.

### *MARs<sub>E $\mu$</sub>* deletion inverts SHM distribution on both sides of the *E $\mu$* enhancer region

To assess whether *MARs<sub>E $\mu$</sub>*  deletion could affect *IgH* somatic hypermutation, we first quantified mutations within the 500-bp regions lying respectively downstream from the *J<sub>H</sub>3* and *J<sub>H</sub>4* exons in Peyer's patch GC B cells sorted from *wt* and *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* mice (overall data were reported in Fig. 2a and b). Interestingly, *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* GC B cells displayed significant differences in the distribution of mutations: an increased proportion of unmutated sequences (30.6 to 45.5%, data collected from several experiments obtained in pool or individual mice, see details in Fig. 2a) of independent clones in mutants compared to less than 10% in *wt*. Another effect of *MARs<sub>E $\mu$</sub>*  deletion on *IgH* SHM targeting was the strong decrease in highly mutated sequences (>10bp per sequence). In *wt*, the proportion of highly mutated alleles reached ~20% (18.9 to 28.3%) while in mutants they were barely present (~2%) in cells devoid of *MARs<sub>E $\mu$</sub>*  (Fig. 2a, pie charts). When comparing only the mutated sequences, mutation frequency was decreased at least by two fold in *MARs<sub>E $\mu$</sub> <sup>Δ/Δ</sup>* mutants, with 6.73 mutations per Kbp (in average) compared to 14.9 in *wt* (in average) (Fig. 2a). By next generation sequencing (NGS) performed in the same cell subset and data processing using our DeMinEr tool



(Martin et al., 2018) a decrease in mutation rate was also observed (Fig. 2b). To confirm the mutation defect, we analyzed SHM by NGS in a large number of GC B cells sorted from spleen of SRCB-immunized  $MARs_{E\mu}^{\Delta/\Delta}$  and *wt* mice. In the intronic region downstream from the  $J_H4$  segment, SHM frequency dropped from 5.3 or 4.6 mutations per Kbp in *wt* cells (average, GS Junior or Ion Proton respectively) to 4.8 or 1.7 in  $MARs_{E\mu}^{\Delta/\Delta}$  cells (Fig. 2c and d). This showed that mutations accumulated by GC B cells only eight days after stimulation were consistently decreased in the absence of  $MARs_{E\mu}$ . Similarly to what was observed in Peyer's patch GC B cells,  $MARs_{E\mu}$ -deficient splenic B cells also displayed an increased proportion of unmutated or poorly mutated sequences (Fig. 2c, pie charts). This confirmed that the intronic region was less efficiently targeted by SHM in mutants.

Since our  $MARs_{E\mu}$  deletion includes the 3' *Hinfl-XbaI* genomic region that contains transcription start sites and part of the  $I\mu$  exon (Cockerill et al., 1987; Lennon and Perry, 1985), we also quantified SHM immediately downstream from this exon in a 600bp region described as mutated in GC B cells (Nagaoka et al., 2002; Petersen et al., 2001). Unlike the intronic regions downstream from the rearranged *VDJ* exon, the overall mutation frequency immediately downstream from  $I\mu$  was significantly increased in GC B cells devoid of  $MARs_{E\mu}$  region and reached 6.9 mutations per Kbp (by the use of either sanger or NGS method) compared to 2.6 and 3.9 mutations per Kbp in *wt* cells (Sanger or GS Junior methods, respectively) (Fig. 2e). This suggests that the region downstream from the  $cE\mu$  was more efficiently targeted in the absence of its  $MARs$ ; a hypothesis supported by the increased proportion of highly mutated sequences (3.4 to 8.%) in  $MARs_{E\mu}$ -deficient GC B cells when compared to *wt* (Fig. 2e, pie charts).

We also compared mutation patterns in NGS Ion Proton and DeMinEr processed large dataset obtained from *wt* and  $MARs_{E\mu}$ -deficient B cells, in order to look for typical hallmarks of DNA repair pathways acting downstream from AID-deamination. Beyond minor differences between individual mice samples, the only remarkable hallmark was an increased in A to G substitutions observed in  $MARs_{E\mu}$ -deficient samples (Supplementary Table 2). The origin of such a difference is questionable although we could suppose some repair pathways downstream from AID deamination are involved.

### **$MARs_{E\mu}$ deletion decreases antisense transcription downstream from the *VDJ* unit in GC B cells**

To decipher the molecular mechanism responsible for targeting SHM around the  $E\mu$  enhancer, we compared GC B cells from Peyer's patches to splenic B cells activated *in vitro*, from *wt* and  $MARs_{E\mu}$ -deficient animals.

It is well established that SHM in Ig V segments is coupled to transcription initiated at V promoters (Fukita et al., 1998). To investigate transcription-related events in SHM-targeted regions upstream and downstream from  $E\mu$ , we precisely quantified the amounts of primary transcripts by using multiple q-PCR probes located respectively downstream from  $J_H3$  (probe C), downstream from  $J_H4$  (probe A, already described in Tinguely et al. (Tinguely et al., 2012), complemented by probe A') (Fig. 3a, details in Supplementary Fig. 1). The use of





cDNA templates conducted with random hexamers showed that the amount of total *IgH* primary transcripts running upstream from  $E\mu$  did not display significant variations between *wt* and  $MARs_{E\mu}$ -deficient cells in both GC and *in vitro*-stimulated samples (probes C, A and A', Fig. 3b). The same results were obtained for primary transcripts running downstream from  $E\mu$  (probe B, Supplementary Fig. 2).

The intriguing discrepancy between the mutation phenotype observed in  $MARs_{E\mu}$ -deficient CG B cells and the silent effect on global transcription motivated a more complete study of transcription events occurring upstream from  $E\mu$ , particularly sense and antisense transcription since the latter has been found in cells undergoing SHM (Perlot et al., 2008). To proceed, we generated cDNA templates with sense transcripts, initiated at the promoter of the rearranged *VDJ* segment, with three primers located downstream from the  $J_H4$  segment (S1 and S2) and within the *cEμ* enhancer (S3) (Fig. 3a and Supplementary Fig. 1). Reciprocally, we generated cDNA templates with antisense transcripts, initiated in the intronic regions upstream from  $E\mu$  as described by Perlot et al. (Perlot et al., 2008), with four primers respectively located downstream from  $J_H2$  (AS0),  $J_H3$  (AS1) and  $J_H4$  (AS2 and AS3) (Fig. 3a and Supplementary Fig. 1). For both sense and antisense, quantification of transcripts was possible with the same probes (C, A and A'). For strand-specific quantification assays with a given probe, the baseline level was either provided by a control reaction measuring endogenous priming since devoid of primer (P-) or by one strand-specific template that cannot be detected with the current probe (§) as reported by Zhao et al. (Bolland et al., 2004; Zhao et al., 2009).

Side by side comparison of *wt* and  $MARs_{E\mu}$ -deficient activated B cells samples revealed several interesting differences. Primary sense transcripts running through introns were more readily detected with probes close to the  $J_H$  segments (probes C and A). Globally, in both GC and *in vitro*-stimulated B cells, sense transcription was never consistently decreased in the absence of  $MARs_{E\mu}$  (Fig. 3c, whole panel). While sense transcript variations in GC cells were not significant, an upward trend could be noticed when shorter transcripts (S1) were detected with C and A probes (Fig. 3c, top and middle filled histograms). In *in vitro*-activated cells, in which sense transcripts were often more abundant, the increase upon  $MARs_{E\mu}$ -deletion becomes significant for long transcripts that could reach the *cEμ* (S3 templates detected with C and A probes,  $p=0.004$ ) (Fig. 3c, top and middle emptied histograms). This indicates that the absence of  $MARs_{E\mu}$  certainly does not hamper RNA pol II to progress 3' to the *VDJ* unit and might even facilitate this process in *in vitro*-stimulated cells. Consequently, the SHM decrease observed in  $MARs_{E\mu}$ -deficient GC B cells could not be attributed to a sense transcription defect.

In contrast, the antisense transcription patterns differed between cell types and genotypes. In *in vitro*-activated cells, antisense transcription downstream from  $J_H$  segments was barely detectable (Fig. 3d, emptied histograms) and, when detected,  $MARs_{E\mu}$ -deletion never hampered transcription (Fig. 3d, emptied histograms, C and A' probes). The opposite was observed in  $MARs_{E\mu}$ -deficient GC B cells in which antisense transcripts running through the intron were always less abundant than in *wt* (Fig. 3d, filled histograms). The



decrease in antisense transcripts was significant when optimal detection was possible by combining closely located primers and probes (Fig. 3d, middle and bottom filled histograms, AS2/probe A,  $p=0.027$  and AS3/probe A',  $p=0.043$ ). Our data highlighted a significant effect of  $MARs_{E\mu}$  deletion on antisense transcription in the region upstream from  $E\mu$ . Seemingly transient, specific to cell subsets and occurring upstream from an enhancer, such antisense transcripts could be substrates for RNA exosome like upstream antisense RNAs (uaRNA/PROMPT) or enhancer RNAs (eRNA) (Flynn et al., 2011; Li et al., 2016; Rothschild and Basu, 2017). To note, this could explain why our longer antisense templates were less readily detected (Fig. 3d, top and bottom filled histograms, AS0/probe C and AS2/probe A'), supposedly being processed through their 3' end by the RNA exosome complex. The noticeable transcription changes in Peyer's patch GC B cells (the subset poised for SHM), suggests that antisense transcription is the  $MARs_{E\mu}$ -driven molecular mechanism involved in the SHM defect.

### **$MARs_{E\mu}$ deletion modifies RNA pol II progression and impairs AID recruitment upstream from $E\mu$ in GC B cells**

To investigate chromatin accessibility of the SHM-targeted regions upstream and downstream from  $E\mu$ , we estimated and compared the enrichment of RNA pol II initiation/stalled and elongation forms (phosphorylated respectively at the Ser5 and Ser2 in its C-terminal domain) by chromatin immunoprecipitation (ChIP). In the same regions, we also estimated the enrichment of AID protein by ChIP. All assays were precisely quantified by using the same q-PCR probes than for transcription assays (Fig. 4a). In both GC and *in vitro*-stimulated B cells, we also verified that  $MARs_{E\mu}$  deletion did not impact *Aicda* gene expression by RT-qPCR: as expected, similar and consistent levels of AID transcripts were found in *wt* and mutant cells in all samples (Supplementary Fig. 3).

RNA pol II ChIP-qPCR experiments pointed out an effect of  $MARs_{E\mu}$  deletion on the density of the various RNA pol II complexes, especially in the intronic region upstream from  $E\mu$  and specifically in GC B cells. By contrast,  $MARs_{E\mu}$ -deletion did not impact RNA pol II recruitment/progression in *in vitro*-activated B cells, since ChIP data display non-significant variations (Fig. 4b and c, emptied histograms).

First, when density of the initiating and/or paused form of RNA pol II (Ser5P) was assayed in the region upstream from  $E\mu$ ,  $MARs_{E\mu}$ -deficient B cells displayed non-significant variations. However, a discrete downward trend was observed with all probes in Peyer's patch GC samples devoid of  $MARs_{E\mu}$  (Fig. 4b, probes C, A & A', filled histograms). Second, RNA pol II Ser5P form was more abundant downstream from  $E\mu$  and both genotypes displayed comparable polymerase enrichment (Fig. 4b, probe B).

Unexpectedly, in Peyer's patches GC B cells, density for the elongation form of RNA pol II (Ser2P) was significantly decreased within introns downstream from  $J_H3$  and  $J_H4$  in the absence of  $MARs_{E\mu}$  (Fig. 4c, filled histograms,  $p=0.004$  with probe C,  $p=0.030$  with probe A;  $p=0.048$  with probe A'). A slight but not significant



decrease was also observed downstream from  $E\mu$  region in  $MARS_{E\mu}$ -deficient GC cells (Fig. 4c, filled histograms, probe B).

The decrease in RNA Pol II Ser2P form observed in  $MARS_{E\mu}$ -deficient GC B cells could result from a global decrease in enhancer activity since it has been shown that active enhancers recruit Pol II Ser2P form on both sides of its core element (Zentner et al., 2011).

More readily than global primary transcription and RNA pol II progression, ChIP experiments using antibody specific to AID revealed a consistent and significant decrease in AID recruitment downstream from  $J_H3$  ( $p=0.032$  with probe C) and  $J_H4$  ( $p=0.016$  with probe A,  $p=0.008$  with probe A') in  $MARS_{E\mu}$ -deficient GC B cells (Fig. 4d, filled histograms). At the opposite, in *in vitro*-activated cells that do not undergo significant SHM (Maul et al., 2014),  $MARS_{E\mu}$  deletion did not influence AID recruitment (Fig. 4d, emptied histograms). In the region downstream from  $E\mu$ , AID recruitment showed non-significant variations between genotypes and cell types (Fig. 4d, probe B). We conclude that  $MARS_{E\mu}$  deletion modified chromatin accessibility to AID in the region upstream from  $E\mu$ , specifically in Peyer's patch GC B cells (the only subset that contains cells undergoing high levels of SHM).



## DISCUSSION

One simplistic model would argue that the most important regulatory regions for *IgH* locus expression are conserved upon any reshaping event occurring in developing B lineage cells (VDJ recombination, CSR and SHM). Mouse KO models targeting either *IgH cE $\mu$*  or the 3'RR support this hypothesis since they led to major phenotype changes (for reviews (Perlot and Alt, 2008; Pinaud et al., 2011)). Accordingly, our current study identifies *MARs<sub>E $\mu$</sub>* , also conserved upon any rearrangement, as part of these most important *IgH* regulatory regions. By linking convergent transcription to SHM targeting efficiency, our study identifies *MARs<sub>E $\mu$</sub>*  as a complementary *cis* regulatory partner of the 3'RR that optimize AID targeting to the *VDJ* unit.

Harbouring normal B-lineage cell subsets and serum Ig levels, our *MARs<sub>E $\mu$</sub>* -deficient mice confirmed, in agreement with previous studies of an analogous MAR region in the *Ig $\kappa$*  locus, (Sakai et al., 1999a; Yi et al., 1999) that intronic MARs are dispensable for B cell ontogeny and antibody production.

Importantly, our mouse model clearly assigns a specific function for endogenous *MARs<sub>E $\mu$</sub>*  on SHM at the *IgH* locus, in accord with the requirement of similar regions for efficient SHM previously pointed out in the endogenous *Ig $\kappa$*  Kappa light chain locus. This pioneer study, describing the specific deletion of a 420pb MAR region upstream from the intronic Kappa enhancer (*Ei $\kappa$* ), highlighted a modest decrease in SHM by quantifying mutations downstream from the *J $\kappa$ 5* segment in GC B cells from Peyer's patches (Yi et al., 1999). However, at that time, the molecular mechanism leading to this effect was enigmatic since no transcription defect was observed in the absence of the *MAR<sub>Ei $\kappa$</sub>* . This model and the numerous transgenic mice that disclosed a possible role of MARs for SHM instigated a variety of hypotheses (Franklin and Blanden, 2005). Taking advantage of the deletion of *MARs<sub>E $\mu$</sub>*  in the mouse, our current study rationalizes some of the molecular mechanisms that physiologically enhance and limit SHM to the genomic region upstream from the *E $\mu$*  intronic enhancer.

First, our study first suggests that *MARs<sub>E $\mu$</sub>*  act as the 3' boundary of the SHM machinery. This hypothesis is mostly supported by the increased SHM frequency downstream from the *cE $\mu$*  enhancer in the absence of *MARs<sub>E $\mu$</sub>* , a finding consistent with previous works describing increased *S $\mu$*  internal deletions in hybridomas devoid of MARs regions (Sakai et al., 1999b). We could speculate that one physiological function of *MARs<sub>E $\mu$</sub>*  regions in GC B cells is to tightly isolate the *VDJ* transcription unit by, at least temporarily, attaching the *E $\mu$*  region to the nuclear matrix and thus provide to this "locked" target, an optimal environment for somatic mutations by trapping the transcription machinery and its co-factors including AID. This topological barrier could, at the same time, partially protect downstream constant regions from SHM; although this configuration should be brief since regions downstream from *E $\mu$*  are also efficiently targeted by AID in GC B cells (Xue et al., 2006).





Several studies indeed proposed that *J-C* intronic MARs help generate negative supercoiling and consequently increased ssDNA and potential other secondary structures that could promote accessibility to AID (Lebecque and Gearhart, 1990; Shen and Storb, 2004; Wright et al., 2008). The hypothesis that  $MARs_{E\mu}$  add again more DNA strain to the sense-transcribed *VDJ* transcription unit is relevant to the positive effect of topoisomerase depletion on AID targeting and SHM (Kobayashi et al., 2011; Maul et al., 2015; Shen and Storb, 2004). While the DNA strain induced by the sense transcription bubble alone might not be sufficient, it is tempting to speculate that, by creating a topological barrier,  $MARs_{E\mu}$  augment DNA strain within the intron and upstream regions to a sufficient level to enable accessibility to AID for SHM, for example, by providing single-stranded patches of DNA as proposed by several groups (Parsa et al., 2012; Ronai et al., 2007; Wang et al., 2014).

Second, our study supports the hypothesis that, complementary to the required sense transcription, antisense transcription in Ig V regions is a relevant regulatory mechanism for SHM. This is strongly supported by recent pioneer works demonstrating that convergent transcription efficiently targets AID (Meng et al., 2014). It then becomes obvious that SHM-related sense and antisense regulatory transcripts should be precisely regulated. Our mouse model brings at least a partial answer by highlighting a function for  $MARs_{E\mu}$  in enhancing SHM-related antisense transcription at the *IgH* locus. Beyond germline sense transcripts, it was shown, in developing B cell precursors undergoing VDJ recombination, that multiple antisense regulatory transcripts, take place in the region upstream from  $E\mu$  (Afshar et al., 2006; Bolland et al., 2004, 2007; Giallourakis et al., 2010) and involve the enhancer and chromatin insulators (Bolland et al., 2007; Featherstone et al., 2010; Guo et al., 2011). At post-medullar stages, antisense regulatory transcripts have been found in switch regions of murine activated B cells prior CSR (Haddad et al., 2011; Perlot et al., 2008) and also in the intronic region upstream from  $E\mu$  in GC B cells as well as in the Ramos cell line (Perlot et al., 2008; Ronai et al., 2007). In all cases, antisense transcripts were detected at the steady-state level, but their relative abundance has been difficult to assess (Zhao et al., 2009). In Ramos and GC B cells, authors showed that levels of SHM-related antisense transcripts were very low when compared to the spliced form of the *VDJ* transcription unit (Perlot et al., 2008; Ronai et al., 2007). The obvious unbalanced sense/antisense transcription ratio could result from either weak transcription rates or instability of antisense products. Nevertheless, Perlot et al. identified by RACE assays, in GC B cells, multiple antisense transcript initiation start sites downstream from every  $J_H$  region and raised the question of specific enhancers. When using  $cE\mu$ -deficient GC B cells or reporter assays and, the authors could not document any antisense-specific promoter associated to the  $E\mu$  region (Perlot et al., 2008), which corresponds to the normal SHM frequency in  $cE\mu$ -deficient B cells (Perlot et al., 2005). While Perlot et al. initially proposed that open chromatin downstream from the *VDJ* transcription unit could be sufficient for initiation of low level antisense transcription in order to prepare SHM (Perlot et al., 2008), our current study refines this previous hypothesis by identifying  $MARs_{E\mu}$



as regulators of SHM-related antisense transcripts that could be eRNA or PROMPT/uaRNA since found upstream from the enhancer (Li et al., 2016).

The ability of  $MARS_{E\mu}$  to modulate antisense eRNA transcription could either result from a direct boost through specific factors or from a passive mechanism providing chromatin accessibility when sufficient DNA torsional strain is reached. One attractive scenario would be that  $MARS_{E\mu}$  actively bind, specifically in GC B cells poised for SHM, some transcription factors able to promote initiation of antisense transcripts moving forward the rearranged  $VDJ$  segment. This is supported by numerous studies, based on gene reporters or transgenes (Forrester et al., 1994; Jenuwein et al., 1997), designing  $MARS_{E\mu}$  as transcriptional regulators through their ability to recruit AT-rich binding factors such as proteins of the ARID or SATB families (Ratliff et al., 2014). Our study also describes  $MARS_{E\mu}$  as “facultative MARS” (*i.e.* acting in a tissue specific manner) since their ability to enhance  $V_H$  antisense transcription is restricted to GC B cells. This scenario does not refute the ability of  $MARS_{E\mu}$  to induce DNA torsional strain but assigns a transcriptional enhancer function that is complementary to its topological barrier role.

The distribution of SHM within and downstream from the  $VDJ$  transcription unit remains enigmatic since, in contrast to  $S$  regions, such regions do not form R-loops and can't pile up RNA pol II complexes (Hackney et al., 2009). Another possible function for  $MARS_{E\mu}$  is their potential to interfere during the elongation or pausing process of the transcription machinery. Although, the fact that  $MARS_{E\mu}$ -deletion modifies the sense/antisense transcripts ratio in the  $J_H-E\mu$  introns solely complicate interpretation of RNA pol II ChIP data: in this context, our experiments should be interpreted with caution. While several studies highlighted accumulation of the stalled form of polymerase with its transcription cofactor Spt5 in the same regions of cells poised for SHM (Maul et al., 2014; Wang et al., 2014), our  $MARS_{E\mu}$ -deletion induced quite unexpected RNA pol II landscapes in the  $J_H-E\mu$  introns. The faint decrease of the Ser5P form in GC B cells devoid of  $MARS_{E\mu}$  could be correlated to the incomplete defect of SHM in GC B cells of mutant mice. The marked decrease of the elongation form (Ser2P) could be linked to the decrease of antisense transcription that broadly initiates at multiple sites downstream from the  $VDJ$  unit (Perlot et al., 2008); this would be in agreement with a study reporting that RNA pol II complexes associated to antisense PROMPT/uaRNAs displayed hallmarks of transcription elongation (Flynn et al., 2011).

In agreement with both SHM and antisense transcription defects, our ChIP data clearly demonstrate that  $MARS_{E\mu}$  promotes accessibility to AID downstream from the  $VDJ$  unit specifically in Peyer's patch GC B cells. Surprisingly,  $MARS_{E\mu}$ -deletion did not increase AID enrichment downstream from  $E\mu$  where mutations were more abundant; we could speculate that such cells accumulate mutations in this region by spending more time in the GC, as suggested by the increased numbers in Peyer's patch of this mutant model. A modified base substitution bias was observed in  $MARS_{E\mu}$  deficient cells, displaying a discrete but specific increase in A to G substitutions. The relationship between this bias and transcription is unclear but one could suppose its



origin takes place in an altered MSH2/6 function in this region in the absence of  $MARs_{E\mu}$  (Saribasak and Gearhart, 2012).

In DT40 and Ramos cells, it was shown that blocking pol II progression increased SHM (Kodgire et al., 2013; Wang et al., 2014). We could speculate that optimal SHM in GC B cells depends on how efficiently  $MARs_{E\mu}$  can slow down RNA pol II in order to give this region access to AID.

An attractive hypothesis proposed by Alt and colleagues (Meng et al., 2014) would be that the optimal chromatin environment for AID targeting would be provided by convergent transcription as the result of fine balance between sense and antisense events. As previously described in yeast (Hobson et al., 2012), pol II complexes proceeding in opposite directions cannot bypass each other and their collision lead to stalling or stopping transcription. Beyond AID recruitment, pol II collision could recruit RNA exosome or other RNase activities able to provide ssDNA targets (Basu et al., 2011; Pefanis et al., 2014). Part of the underlying mechanism has been recently described by the Basu group by showing the involvement of the RNA exosome and its associated helicase in the strand-dependent targeting of somatic mutation in the B cell genome (Lim et al., 2017). Our study suggests that, at Ig gene loci, scaffold regions are actors of this process.



## Material and Methods

**Generating  $MARs_{E\mu}$  KO mice.** Gene targeting for matrix attachment regions flanking the *IgH E $\mu$*  enhancer element was performed by homologous recombination, in the murine E14 ES cell line, with a vector kindly provided by Dr. Frederick Alt that permitted replacement of the 995 pb region (including *cE $\mu$*  and its flanking *MARs*) by a 220 pb *Hinfl* genomic fragment that reintroduced the *cE $\mu$*  enhancer followed by a “*loxP-pGK-Neo<sup>R</sup>-loxP*” cassette (Sakai et al., 1999a). Once introduced in the mouse germline, the selection cassette was deleted *in vivo* by *cre-loxP* recombination as previously described (Marquet et al., 2014) to obtain the  $MARs_{E\mu}^{\Delta}$  *IgH* allele devoid of both 5' and 3'  $MARs_{E\mu}$  (respectively 344 pb *XbaI-Hinfl* and 426bp *Hinfl-XbaI* genomic fragments) (Fig. 1a). Animal procedures were performed on 8 weeks old male and female mice.

**Immunisation.** Mice were immunised by intraperitoneal injection of 200 $\mu$ L of sheep red blood cells and sacrificed 8 days after.

**Southern blots and PCR analysis of cre-mediated  $MARs_{E\mu}$  deletion.** Genomic Southern blots were performed as follows: 20  $\mu$ g genomic DNA were digested by *SacI* or *BamHI* and submitted to electrophoresis on a 0.7% agarose gel. DNA was transferred to nylon membranes (MP Biomedicals) by capillarity. Blots were hybridized with [<sup>32</sup>P]-labeled probes generated by random priming. Hybridization with “X” probe (0.803 kb *SacI-SphI* fragment) and “Y” probe (0.8 kb *XbaI-BamHI* fragment) located outside the homology arms were used to identify ES cell clones in which  $MARs_{E\mu}$  were replaced by the *loxP-pGK-Neo<sup>R</sup>-loxP* cassette (Fig. 1a).

**Flow cytometry and cell sorting.** Flow cytometry analysis was performed on LSR-Fortessa cell analyzer (BD Biosciences) on single-cell suspensions from fresh organs. Once washed with 2% fetal calf serum-PBS, lymphoid cells from bone marrow, spleen, peritoneal cavity and Peyer's patches were labeled with various conjugated Abs:  $\alpha$ B220-V450,  $\alpha$ IgM<sup>a</sup>-FITC,  $\alpha$ CD117-PE,  $\alpha$ IgM<sup>b</sup>-PE and  $\alpha$ CD43-PE for bone marrow cells.  $\alpha$ B220-V450,  $\alpha$ CD21-PE,  $\alpha$ CD23-FITC,  $\alpha$ IgM total-PE,  $\alpha$ IgD-FITC and  $\alpha$ CD3e-FITC,  $\alpha$ IgMa-FITC,  $\alpha$ IgMb-PE for splenocytes.  $\alpha$ B220-V450,  $\alpha$ IgM total-PE,  $\alpha$ CD5-FITC for peritoneal cavity.  $\alpha$ B220-V450,  $\alpha$ B220-APC,  $\alpha$ IgA-FITC,  $\alpha$ IgM total-PE,  $\alpha$ PNA-FITC,  $\alpha$ FAS-PE,  $\alpha$ Ki67-FITC for Peyer's Patches. (Southern Biotechnology Associates; eBioscience; Sigma and BD Biosciences). Flow cytometry cell sorting was performed on an ARIA 3 (BD Biosciences) apparatus on single-cell suspensions from spleens or Peyer's patches. Once washed with 2% fetal calf serum-PBS, cells were labeled with  $\alpha$ B220 and  $\alpha$ GL7 or  $\alpha$ PNA and  $\alpha$ FAS antibodies and sorted based on distinct gates defined as germinal centre B cells (B220<sup>+</sup>/GL7<sup>+</sup> or B220<sup>+</sup>/PNA<sup>High</sup>/Fas<sup>+</sup>).

**Cell culture.** Splenocytes were collected, after red blood cells lysis, CD43<sup>+</sup> cells were depleted using anti-CD43 MicroBeads (Miltenyi Biotec). CD43<sup>-</sup> splenic B cells were cultured for 3 days at a density of 1x10<sup>6</sup> cells per ml in RPMI 1640 supplemented in 10% serum calf fetal, sodium pyruvate (Lonza), amino acid (NEAA 100x Lonza) and Penicillin-Streptomycin (Gibco) with 1 $\mu$ g/ml LPS (Invivogen) alone (for transcription assays) or plus 20ng/ml IL4 (PeproTech) (for ChIP experiments).





**SHM assays.** SHM analysis was either performed by cloning followed by classical Sanger method as described (Rouaud et al., 2013) or performed directly on PCR products by next generation sequencing using GSJunior (Roche) or Ion Proton system (Applied Biosystem). For GS Junior, sequencing libraries were prepared according to the manufacturer's instructions, adaptor sequences were added to the previous amplification primer sequences in order to be compatible with the GS-Junior sequencing technology. Amplifications were performed with Phusion® High-Fidelity DNA Polymerase (New England Biolabs) according to the following program: DNA was denatured 40 s at 98°C and then submitted to 38 cycles consisting of 98°C for 10 s, 68°C for 30 s and 72°C for 30 s, and 1 cycle at 72°C for 10 min. PCR products were first purified using NucleoSpin kit (Macherey-Nagel) followed by Ampure bead purification (Beckman Coulter). PCR products were subjected to "PCR emulsion step" (GS-Junior+ emPCR Kit (Lib-A), Roche) and sequenced using GS junior sequencing kit XL+ (Roche) according to the manufacturer's instructions.

Raw sequences were aligned against reference sequences of IgHJ4-downstream intron or Smu and only full length sequences were kept for mutation analysis. For IgHJ4, clonally related sequences were removed based on the sequence of VDJ junction similarity. No further filtering steps were implemented in our analysis workflow. Mutations were called on each sequence using pairwise alignment algorithm (from biopython package) and only base substitutions were reported. Mutation rates were computed as the ratio between the sum of mutated bases in all complete sequences over the total number of aligned bases.

For Ion Proton, sequencing libraries were prepared according to the user guide Ion Xpress™ Plus gDNA Fragment Library Preparation (Cat. no. 4471269, Life Technologies). Briefly, PCR products (100ng) were fragmented by enzymatic digestion (Ion Shear™ Plus Reagents Kit, Cat. no. 4471248) and ligated to Barcodes and Adapters (Ion Plus Fragment Library Kit, Cat. no. 4471252). After 200 bp size selection step on E-Gel precast agarose electrophoresis system, final amplification was performed. Raw data were processed using DeMinEr tool as described in (Martin et al., 2018).

**RT-PCR and q-PCR.** Total RNA was prepared by using TRIzol reagent (Ambion) procedures. RNA samples were first treated with DNase I (Invitrogen) for 15 min at 25°C. RT was performed on 200 ng of total RNA with random hexamers or with specific primer (10µM) (listed in Supplementary Fig. 2.) using superscript III enzyme (Invitrogen). As control, we performed a reverse transcription without primer to determine the threshold. Each real-time qPCR reaction was performed in duplicate on 10 ng of RNA equivalent, using TaqMan Universal or SYBR green Mastermix (TAKARA) on StepOnePlus system (Applied Biosystems). Primary transcription at *IgH* locus was quantified as previously described (Tinguely et al., 2012) and completed with both a set of primers and q-PCR probes close to *J<sub>H</sub>* segments (listed in Supplementary Fig.2.) and a set of primers located 5' to *S<sub>μ</sub>* : Smu-Fw (5'-ACCCAGGCTAAGAAGGCAAT-3'), Smu-Rev (5'-CCTTCCTTCTGCGTATCCAT-3'). AID transcription was analyzed using TaqMan probe (Mm01184115\_m1



*Aicda*, Applied Biosystem). Relative mRNA levels were normalized to *Gapdh* transcripts with the appropriate TaqMan probe (Mm99999915\_g1, Applied Biosystem). Data were analyzed by comparing threshold cycle (CT) values according to the  $2^{-(\Delta\Delta CT)}$  method. The *wt* mice templates used as calibrators were S2 for sense transcripts, AS0 or AS2 for antisense transcripts.

**Total serum Ig quantification by ELISA.** Sera from *wt* and *MARS<sub>E $\mu$</sub>  <sup>$\Delta/\Delta$</sup>*  mice were analyzed for the presence of different Ig classes and subclasses by ELISA as previously described (Pinaud et al., 2001).

**Chromatin immunoprecipitation.** ChIP-IT High Sensitivity kit (Active Motif) was used according to the manufacturer's instructions. ChIP experiments were performed on isolated Peyer's patches cells and on *in vitro* activated B cells. Briefly,  $5 \times 10^6$  cells were cross-linked with 1% formaldehyde at room temperature for 15 min. The reaction was stopped by adding glycine to a final concentration of 0.125 M. After the cross-linking reaction, cell lysis and sonication were performed on Vibracell sonicator (Sonics and Materials, Inc). For immunoprecipitation, 4  $\mu$ g of each antibody was added to the lysate at 4 °C overnight. Polyclonal rabbit antibodies directed against AID were previously described (Kohli et al., 2010). Anti-RNA pol II Ser2P, anti-RNA pol II Ser5P and normal rabbit IgG were purchased from Abcam and Millipore respectively. Immunoprecipitated DNA was used for real-time qPCR. The quantitative PCR signal was normalized using the % input method after subtraction of IgG background. Primers used for ChIP-qPCR analysis were the same as described above (see q-PCR section and Supplementary Fig. 2).

**Statistical analysis** Statistical Mann Whitney tests were performed using GraphPad Prism (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ).



## **Acknowledgments**

We thank Sylvie Desforges, Bernadette Remeyrand and Nadine Cogné for mouse work. We are grateful to Dr. Zéliha Oruc-Ratinaud regarding the animal health status. We are grateful to Dr. Fred Alt for providing  $MAR_{SE\mu}$  targeting construct backbone. We thank Drs. Dominik Schenten, Sébastien Storck, Christophe Sirac, Yves Denizot, Sophie Péron, Brice Laffleur, Alexis Saintamand, Jérôme Saulière for discussions and helpful comments, we are also grateful to Drs. Robert Maul and Jeanne Moreau for critical review of the manuscript. This work was supported by a grant from CHU de Limoges [CORC : Comité d'Orientation Recherche en Cancer Limousin « 2014 »], sponsored by Conseil Régional and ARS Limousin (to E.P.), grants from ARC (PGA120150202338), Agence Nationale de la Recherche (ANR grant 16-CE15-0019-01), Institut National du Cancer (INCa grant #9363 and PLBIO #15-256). This work also benefited from the support INCa-Cancéropôle GSO Emergence (to EP). M.M. was supported by a PhD scholarship from Centre National de la Recherche Scientifique and Région Limousin; S.L.N. was supported by a post-doc fellowship Grant ANR-11-BSV302701; O.M. was supported by the M.E.N.S.E.R. PhD fellowship of the french government; F.B. was supported by Fondation Partenariale de l'Université de Limoges and ALURAD. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

## **Author contribution**

OM, MM, AG, MB, SB, CC, EP and SLN performed experiments. FB performed bioinformatics analysis. JCA, MC, LD, EP and SLN conceived and supervised the study. MM developed the experimental model. OM, PG, LD, EP and SLN wrote the manuscript.

## **Additional information**

**Competing financial interest:** The authors declare no competing financial interests.



## References

- Afshar, R., Pierce, S., Bolland, D.J., Corcoran, A., and Oltz, E.M. (2006). Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J. Immunol.* *176*, 2439–2447.
- Azuma, T., Motoyama, N., Fields, L.E., and Loh, D.Y. (1993). Mutations of the chloramphenicol acetyl transferase transgene driven by the immunoglobulin promoter and intron enhancer. *Int. Immunol.* *5*, 121–130.
- Bachl, J., and Wabl, M. (1996). Enhancers of hypermutation. *Immunogenetics* *45*, 59–64.
- Basu, U., Meng, F.-L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* *144*, 353–363.
- Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004). Antisense intergenic transcription in V(D)J recombination. *Nat. Immunol* *5*, 630–637.
- Bolland, D.J., Wood, A.L., Afshar, R., Featherstone, K., Oltz, E.M., and Corcoran, A.E. (2007). Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer *Emu*. *Mol. Cell. Biol.* *27*, 5523–5533.
- Cockerill, P.N., Yuen, M.H., and Garrard, W.T. (1987). The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *J. Biol. Chem.* *262*, 5394–5397.
- Featherstone, K., Wood, A.L., Bowen, A.J., and Corcoran, A.E. (2010). The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. *J. Biol. Chem* *285*, 9327–9338.
- Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. U.S.A.* *108*, 10460–10465.
- Forrester, W.C., van Genderen, C., Jenuwein, T., and Grosschedl, R. (1994). Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. *Science* *265*, 1221–1225.
- Franklin, A., and Blanden, R.V. (2005). Hypothesis: biological role for J-C intronic matrix attachment regions in the molecular mechanism of antigen-driven somatic hypermutation. *Immunol. Cell Biol.* *83*, 383–391.
- Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* *9*, 105–114.
- Giallourakis, C.C., Franklin, A., Guo, C., Cheng, H.-L., Yoon, H.S., Gallagher, M., Perlot, T., Andzelm, M., Murphy, A.J., Macdonald, L.E., et al. (2010). Elements between the IgH variable (V) and diversity (D) clusters influence antisense transcription and lineage-specific V(D)J recombination. *Proc. Natl. Acad. Sci. U.S.A* *107*, 22207–22212.
- Giusti, A.M., and Manser, T. (1993). Hypermutation is observed only in antibody H chain V region transgenes that have recombined with endogenous immunoglobulin H DNA: implications for the location of cis-acting elements required for somatic mutation. *J. Exp. Med.* *177*, 797–809.





- Gluch, A., Vidakovic, M., and Bode, J. (2008). Scaffold/matrix attachment regions (S/MARs): relevance for disease and therapy. *Handb Exp Pharmacol* 67–103.
- Goyenechea, B., Klix, N., Yélamos, J., Williams, G.T., Riddell, A., Neuberger, M.S., and Milstein, C. (1997). Cells strongly expressing Ig(kappa) transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. *EMBO J.* 16, 3987–3994.
- Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., et al. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424–430.
- Hackney, J.A., Misaghi, S., Senger, K., Garris, C., Sun, Y., Lorenzo, M.N., and Zarrin, A.A. (2009). DNA targets of AID evolutionary link between antibody somatic hypermutation and class switch recombination. *Adv. Immunol* 101, 163–189.
- Haddad, D., Puget, N., Laviolette-Malirat, N., Conte, C., and Khamlichi, A.A. (2011). Seeking sense of antisense switch transcripts. *Transcription* 2, 183–188.
- Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. *Mol. Cell* 48, 365–374.
- Inlay, M.A., Gao, H.H., Odegard, V.H., Lin, T., Schatz, D.G., and Xu, Y. (2006). Roles of the Ig kappa light chain intronic and 3' enhancers in Igk somatic hypermutation. *J. Immunol.* 177, 1146–1151.
- Jenuwein, T., Forrester, W.C., Fernández-Herrero, L.A., Laible, G., Dull, M., and Grosschedl, R. (1997). Extension of chromatin accessibility by nuclear matrix attachment regions. *Nature* 385, 269–272.
- Kaplan, M.H., Zong, R.T., Herrscher, R.F., Scheuermann, R.H., and Tucker, P.W. (2001). Transcriptional activation by a matrix associating region-binding protein. contextual requirements for the function of bright. *J. Biol. Chem.* 276, 21325–21330.
- Kobayashi, M., Sabouri, Z., Sabouri, S., Kitawaki, Y., Pommier, Y., Abe, T., Kiyonari, H., and Honjo, T. (2011). Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19305–19310.
- Kodgire, P., Mukkavar, P., Ratnam, S., Martin, T.E., and Storb, U. (2013). Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID. *J. Exp. Med.* 210, 1481–1492.
- Kohli, R.M., Maul, R.W., Guminski, A.F., McClure, R.L., Gajula, K.S., Saribasak, H., McMahon, M.A., Siliciano, R.F., Gearhart, P.J., and Stivers, J.T. (2010). Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J. Biol. Chem.* 285, 40956–40964.
- Kohwi-Shigematsu, T., Maass, K., and Bode, J. (1997). A thymocyte factor SATB1 suppresses transcription of stably integrated matrix-attachment region-linked reporter genes. *Biochemistry* 36, 12005–12010.
- Lebecque, S.G., and Gearhart, P.J. (1990). Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *J. Exp. Med.* 172, 1717–1727.
- Lennon, G.G., and Perry, R.P. (1985). C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. *Nature* 318, 475–478.



- Li, F., Yan, Y., Pieretti, J., Feldman, D.A., and Eckhardt, L.A. (2010). Comparison of identical and functional Igh alleles reveals a nonessential role for E $\mu$  in somatic hypermutation and class-switch recombination. *J. Immunol.* *185*, 6049–6057.
- Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* *17*, 207–223.
- Lim, J., Giri, P.K., Kazadi, D., Laffleur, B., Zhang, W., Grinstein, V., Pefanis, E., Brown, L.M., Ladewig, E., Martin, O., et al. (2017). Nuclear Proximity of Mtr4 to RNA Exosome Restricts DNA Mutational Asymmetry. *Cell* *169*, 523–537.e15.
- Lin, M.M., Green, N.S., Zhang, W., and Scharff, M.D. (1998). The effects of E  $\mu$ , 3'alpha (hs 1,2) and 3'kappa enhancers on mutation of an Ig-VDJ-Cgamma2a Ig heavy gene in cultured B cells. *Int. Immunol.* *10*, 1121–1129.
- Marquet, M., Garot, A., Bender, S., Carrion, C., Rouaud, P., Lecardeur, S., Denizot, Y., Cogné, M., and Pinaud, E. (2014). The E $\mu$  enhancer region influences H chain expression and B cell fate without impacting IgVH repertoire and immune response in vivo. *J. Immunol.* *193*, 1171–1183.
- Martin, O.A., Garot, A., Le Noir, S., Aldigier, J.-C., Cogné, M., Pinaud, E., and Boyer, F. (2018). Detecting Rare AID-Induced Mutations in B-Lineage Oncogenes from High-Throughput Sequencing Data Using the Detection of Minor Variants by Error Correction Method. *J. Immunol.* *201*, 950–956.
- Maul, R.W., Cao, Z., Venkataraman, L., Giorgetti, C.A., Press, J.L., Denizot, Y., Du, H., Sen, R., and Gearhart, P.J. (2014). Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo-activated cells. *J. Exp. Med.* *211*, 2297–2306.
- Maul, R.W., Saribasak, H., Cao, Z., and Gearhart, P.J. (2015). Topoisomerase I deficiency causes RNA polymerase II accumulation and increases AID abundance in immunoglobulin variable genes. *DNA Repair (Amst.)* *30*, 46–52.
- Meng, F.-L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.-R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberger, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* *159*, 1538–1548.
- Motoyama, N., Miwa, T., Suzuki, Y., Okada, H., and Azuma, T. (1994). Comparison of somatic mutation frequency among immunoglobulin genes. *J. Exp. Med.* *179*, 395–403.
- Nagaoka, H., Muramatsu, M., Yamamura, N., Kinoshita, K., and Honjo, T. (2002). Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region: implication of AID involvement in a common step of class switch recombination and somatic hypermutation. *J. Exp. Med.* *195*, 529–534.
- Parsa, J.-Y., Ramachandran, S., Zaheen, A., Nepal, R.M., Kapelnikov, A., Belcheva, A., Berru, M., Ronai, D., and Martin, A. (2012). Negative supercoiling creates single-stranded patches of DNA that are substrates for AID-mediated mutagenesis. *PLoS Genet.* *8*, e1002518.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* *514*, 389–393.
- Perlot, T., and Alt, F.W. (2008). Cis-regulatory elements and epigenetic changes control genomic rearrangements of the IgH locus. *Adv. Immunol.* *99*, 1–32.



- Perlot, T., Alt, F.W., Bassing, C.H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc. Natl. Acad. Sci. U.S.A* *102*, 14362–14367.
- Perlot, T., Li, G., and Alt, F.W. (2008). Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proc. Natl. Acad. Sci. U.S.A* *105*, 3843–3848.
- Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R., et al. (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* *414*, 660–665.
- Pinaud, E., Khamlichi, A.A., Le Morvan, C., Drouet, M., Nalesso, V., Le Bert, M., and Cogné, M. (2001). Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* *15*, 187–199.
- Pinaud, E., Marquet, M., Fiancette, R., Péron, S., Vincent-Fabert, C., Denizot, Y., and Cogné, M. (2011). The IgH locus 3' regulatory region: pulling the strings from behind. *Adv. Immunol.* *110*, 27–70.
- Pommier, Y., Sun, Y., Huang, S.-Y.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* *17*, 703–721.
- Ratliff, M.L., Templeton, T.D., Ward, J.M., and Webb, C.F. (2014). The Bright Side of Hematopoiesis: Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. *Front Immunol* *5*, 113.
- Ronai, D., Berru, M., and Shulman, M.J. (1999). Variegated expression of the endogenous immunoglobulin heavy-chain gene in the absence of the intronic locus control region. *Mol. Cell. Biol.* *19*, 7031–7040.
- Ronai, D., Iglesias-Ussel, M.D., Fan, M., Shulman, M.J., and Scharff, M.D. (2005). Complex regulation of somatic hypermutation by cis-acting sequences in the endogenous IgH gene in hybridoma cells. *Proc. Natl. Acad. Sci. U.S.A.* *102*, 11829–11834.
- Ronai, D., Iglesias-Ussel, M.D., Fan, M., Li, Z., Martin, A., and Scharff, M.D. (2007). Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *J. Exp. Med.* *204*, 181–190.
- Rothschild, G., and Basu, U. (2017). Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends Genet.* *33*, 143–154.
- Rouaud, P., Vincent-Fabert, C., Saintamand, A., Fiancette, R., Marquet, M., Robert, I., Reina-San-Martin, B., Pinaud, E., Cogné, M., and Denizot, Y. (2013). The IgH 3' regulatory region controls somatic hypermutation in germinal center B cells. *J. Exp. Med.* *210*, 1501–1507.
- Sakai, E., Bottaro, A., Davidson, L., Sleckman, B.P., and Alt, F.W. (1999a). Recombination and transcription of the endogenous Ig heavy chain locus is effected by the Ig heavy chain intronic enhancer core region in the absence of the matrix attachment regions. *Proc. Natl. Acad. Sci. U.S.A.* *96*, 1526–1531.
- Sakai, E., Bottaro, A., and Alt, F.W. (1999b). The Ig heavy chain intronic enhancer core region is necessary and sufficient to promote efficient class switch recombination. *Int. Immunol.* *11*, 1709–1713.
- Saribasak, H., and Gearhart, P.J. (2012). Does DNA repair occur during somatic hypermutation? *Semin. Immunol.* *24*, 287–292.
- Scheuermann, R.H., and Garrard, W.T. (1999). MARs of antigen receptor and co-receptor genes. *Crit. Rev. Eukaryot. Gene Expr.* *9*, 295–310.



- Shen, H.M., and Storb, U. (2004). Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc. Natl. Acad. Sci. U.S.A.* *101*, 12997–13002.
- Teves, S.S., and Henikoff, S. (2014). DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus* *5*, 211–218.
- Tinguely, A., Chemin, G., Péron, S., Sirac, C., Reynaud, S., Cogné, M., and Delpy, L. (2012). Cross talk between immunoglobulin heavy-chain transcription and RNA surveillance during B cell development. *Mol. Cell. Biol.* *32*, 107–117.
- Wang, X., Fan, M., Kalis, S., Wei, L., and Scharff, M.D. (2014). A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. *Nat Commun* *5*, 4137.
- Wang, Z., Goldstein, A., Zong, R.T., Lin, D., Neufeld, E.J., Scheuermann, R.H., and Tucker, P.W. (1999). Cux/CDP homeoprotein is a component of NF- $\mu$ NR and represses the immunoglobulin heavy chain intronic enhancer by antagonizing the bright transcription activator. *Mol. Cell. Biol.* *19*, 284–295.
- Wiersma, E.J., Ronai, D., Berru, M., Tsui, F.W., and Shulman, M.J. (1999). Role of the intronic elements in the endogenous immunoglobulin heavy chain locus. Either the matrix attachment regions or the core enhancer is sufficient to maintain expression. *J. Biol. Chem.* *274*, 4858–4862.
- Wright, B.E., Schmidt, K.H., Davis, N., Hunt, A.T., and Minnick, M.F. (2008). II. Correlations between secondary structure stability and mutation frequency during somatic hypermutation. *Mol. Immunol.* *45*, 3600–3608.
- Xue, K., Rada, C., and Neuberger, M.S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*<sup>-/-</sup> *ung*<sup>-/-</sup> mice. *J. Exp. Med.* *203*, 2085–2094.
- Yi, M., Wu, P., Trevorrow, K.W., Claflin, L., and Garrard, W.T. (1999). Evidence that the I $\kappa$  gene MAR regulates the probability of premature V-J joining and somatic hypermutation. *J. Immunol.* *162*, 6029–6039.
- Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* *21*, 1273–1283.
- Zhao, Y., Dunn-Walters, D.K., Barone, F., and Spencer, J. (2009). Antisense transcripts of V(D)J rearrangements; artifacts caused by false priming? *Mol. Immunol.* *46*, 2357–2362.



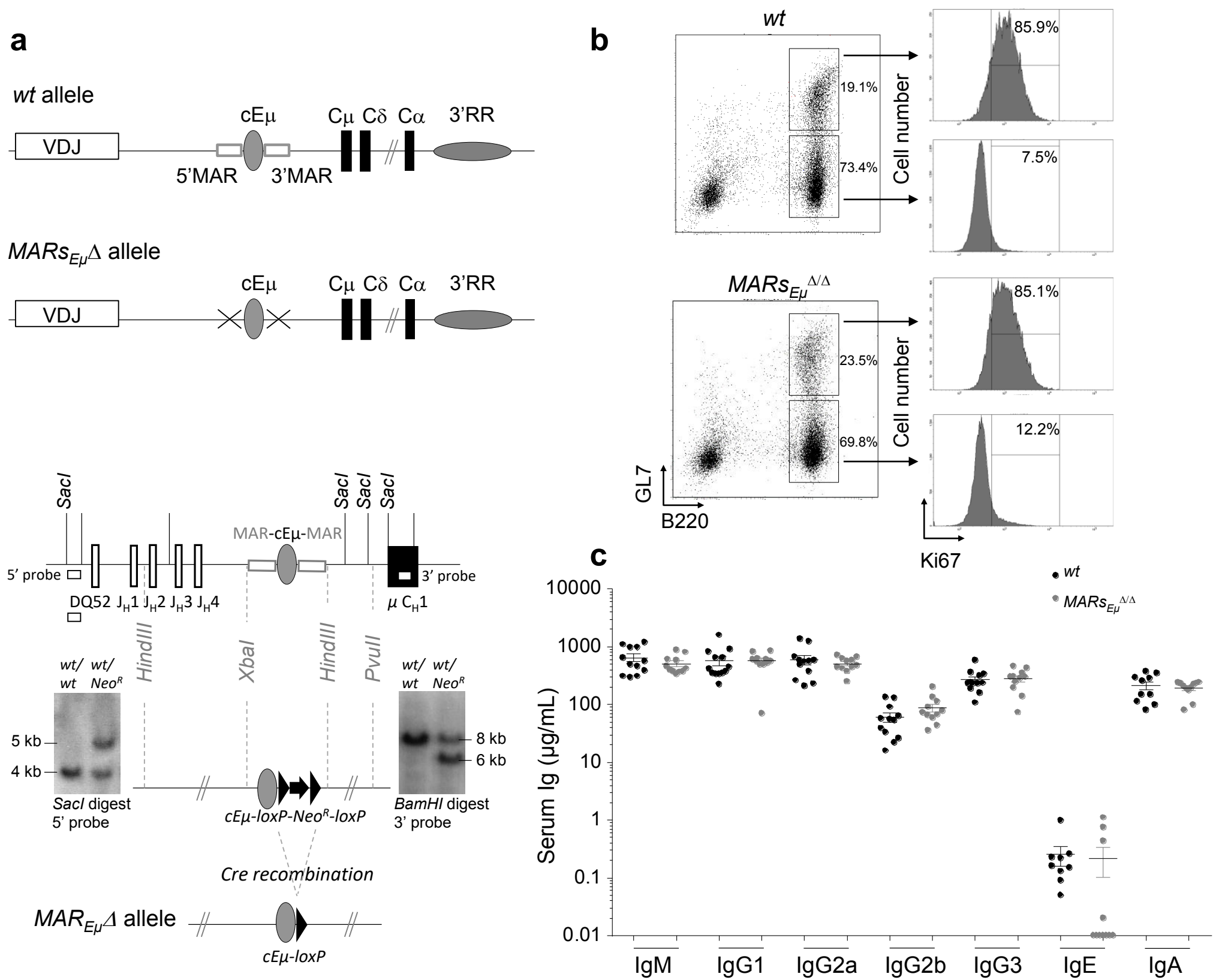


**Figure 1.  $MAR_{S_{E\mu}}$  deletion supports efficient *in vivo* Ig isotype production and GC B cell development**

**(a)** Schematic representation of *wt* and  $MAR_{S_{E\mu}}\Delta$  alleles (top). Targeting construct and Southern blot performed on recombinant ES cells with *Neo<sup>R</sup>* insertion. Hybridization with the 5' probe detected 4 kb and 5 kb *SacI* genomic fragments respectively for *wt* and recombined alleles. Hybridization with the 3' probe detected 8 kb and 6 kb *BamHI* genomic fragments respectively for *wt* and recombined alleles.  $MAR_{E\mu}\Delta$  allele preserved the *cE $\mu$*  enhancer after Cre-recombination. **(b)** Comparison of Peyer's patch B cells subsets from *wt* and  $MAR_{S_{E\mu}}^{\Delta/\Delta}$  animals by flow cytometry: dot plots showed percentage of naïve (B220<sup>+</sup>/GL7<sup>-</sup>) and GC (B220<sup>+</sup>/GL7<sup>+</sup>) B cells (left panel) and, for each subset, the percentage of dividing cells (Ki67<sup>+</sup>) was shown as histogram plots (right panel). Experiments were performed twice with a minimum of 3 mice per group. **(c)** Immunoglobulin isotype secretion in sera from *wt* and  $MAR_{S_{E\mu}}^{\Delta/\Delta}$  mice determined by ELISA (n=9 to 12 mice, mean±SEM).



Figure 1





**Figure 2.  $MARs_{E\mu}$  deletion impairs the overall SHM frequency and distribution within the *IgH J-C* intronic region**

(a-b) Somatic hypermutation downstream from  $J_H3$  and  $J_H4$  segments in Peyer's patch GC B cells sorted from *wt* and  $MARs_{E\mu}^{\Delta/\Delta}$  mice; quantified after cloning and sequencing by classical Sanger method (a) or by next generation sequencing (b). (c-d) Somatic hypermutation downstream from  $J_H4$  segments in spleen GC B cells sorted from SRBC-immunized *wt* and  $MARs_{E\mu}^{\Delta/\Delta}$  mice; quantified by next generation sequencing GS Junior (c) or Ion Proton system (d). (e) Somatic hypermutation downstream from  $cE\mu$  region from Peyer's patch GC B cells sorted from *wt* and  $MARs_{E\mu}^{\Delta/\Delta}$  mice; quantified after cloning and sequencing by classical Sanger method or by next generation sequencing.

For each genotype, pie charts (obtained from Sanger and GS Junior sequencing method) showed the percentage of mutation per sequence (proportional to the area in each slice) and total number of individual clones is indicated in the center (after removal of clonally related sequences based on VDJ junction). Since no indication in sequence distribution are available using Ion Proton sequencing method, data were represented as scattered plots, each point refer to a mutation frequency from one individual mice.

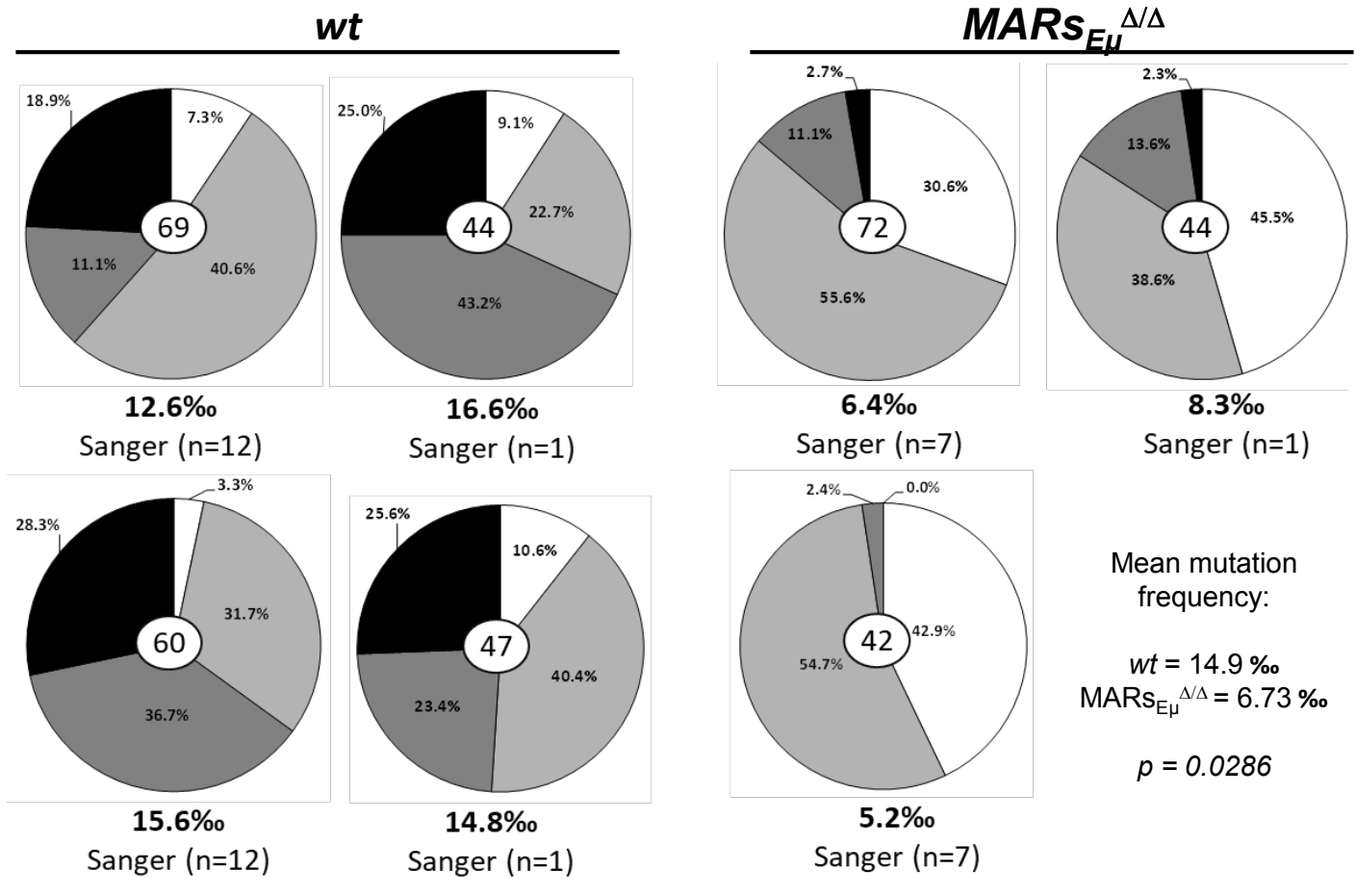
For each figures (pie charts and scattered plots) the overall mutation frequency, the sequencing method and the number of mice are indicated. \* $P < 0.05$ , (Mann–Whitney U-test, one tailed).



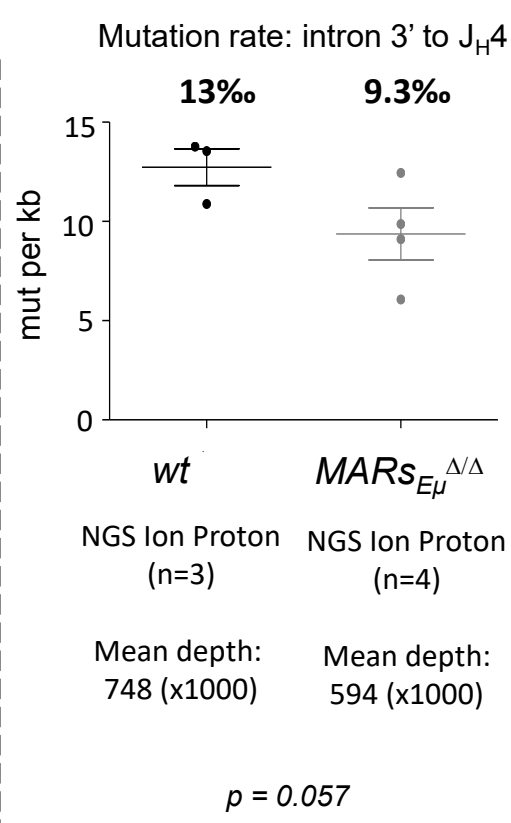
# Figure 2

**a**

Intron 3' to  $J_H3$  or  $J_H4$  Peyer's patches GC

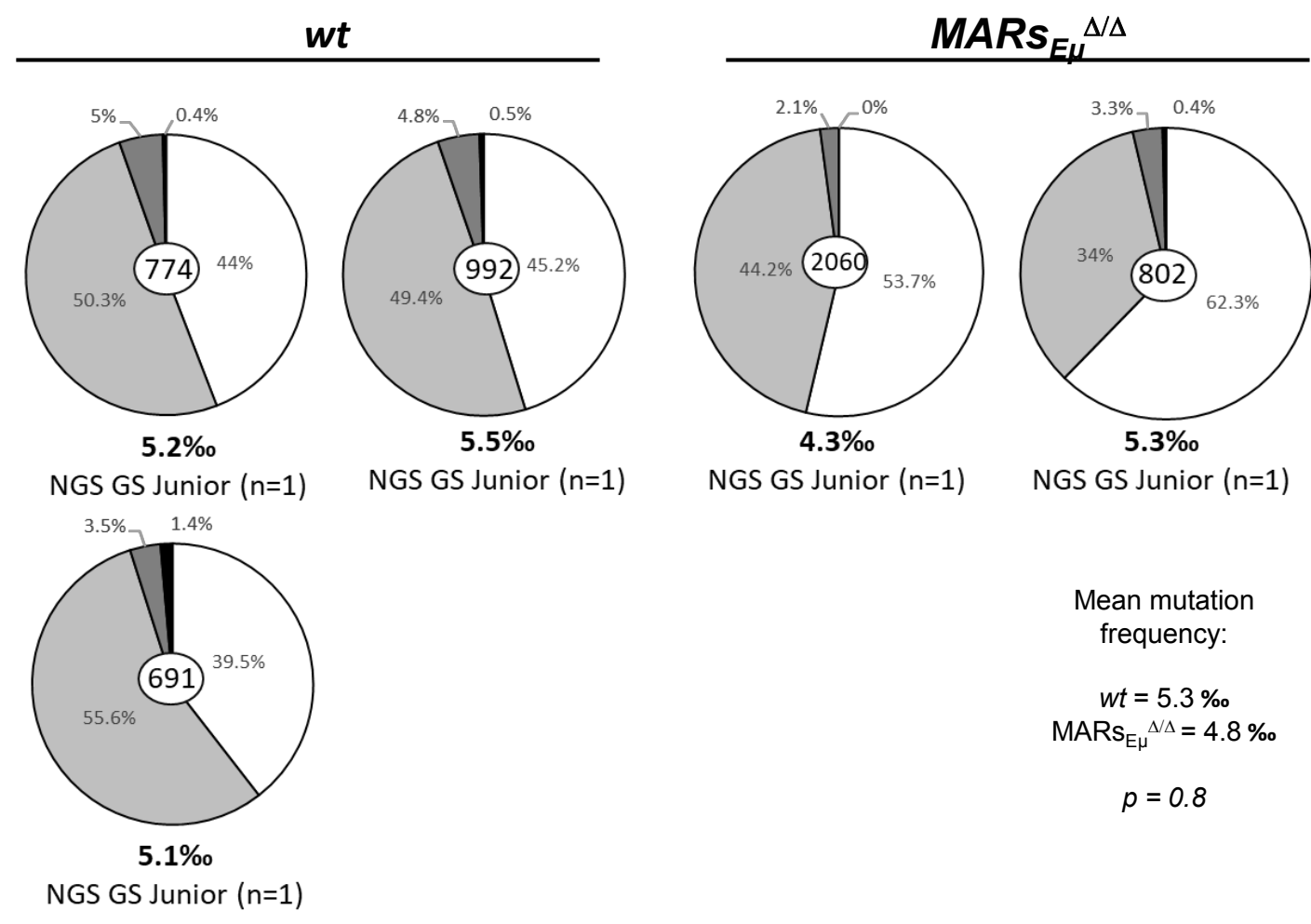


**b**

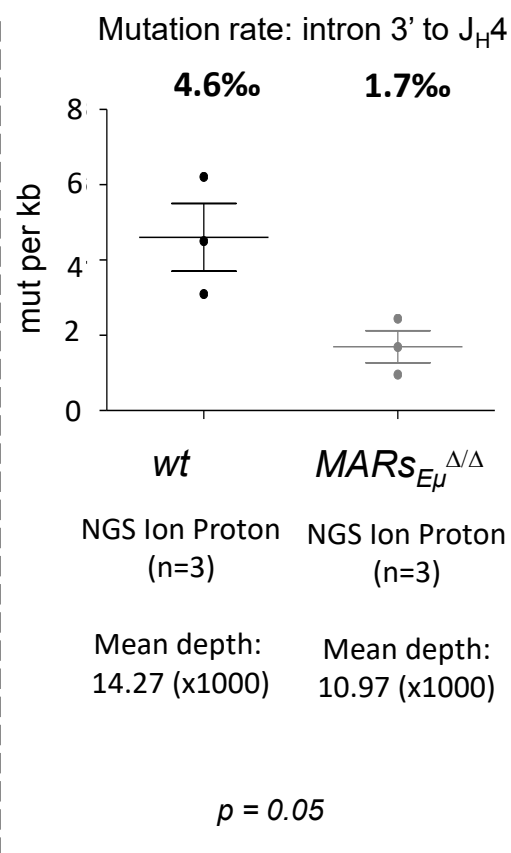


**c**

Intron 3' to  $J_H4$  spleen GC

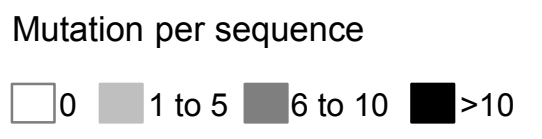
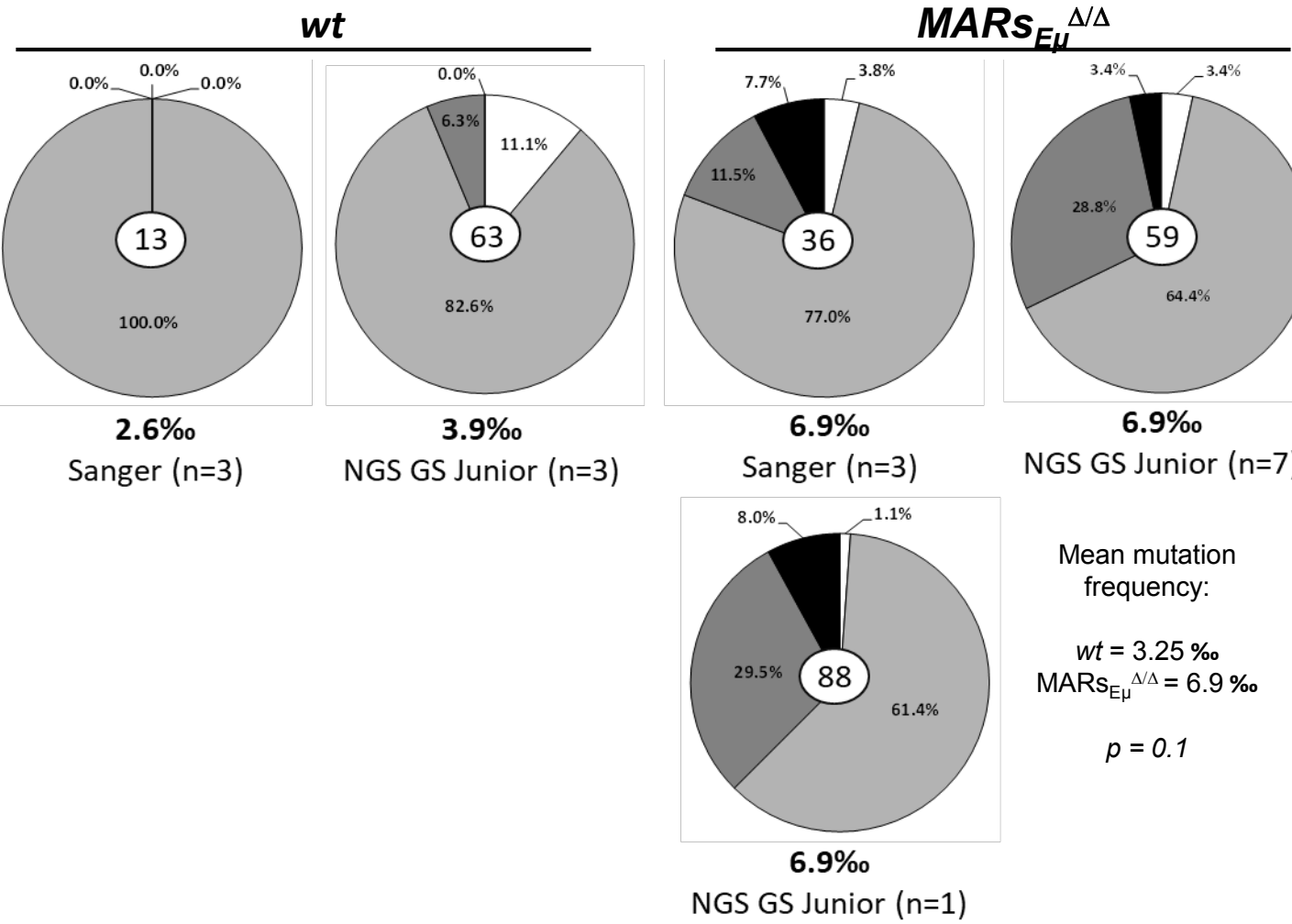


**d**



**e**

Intron 5' to  $Smu$  Peyer's patch GC





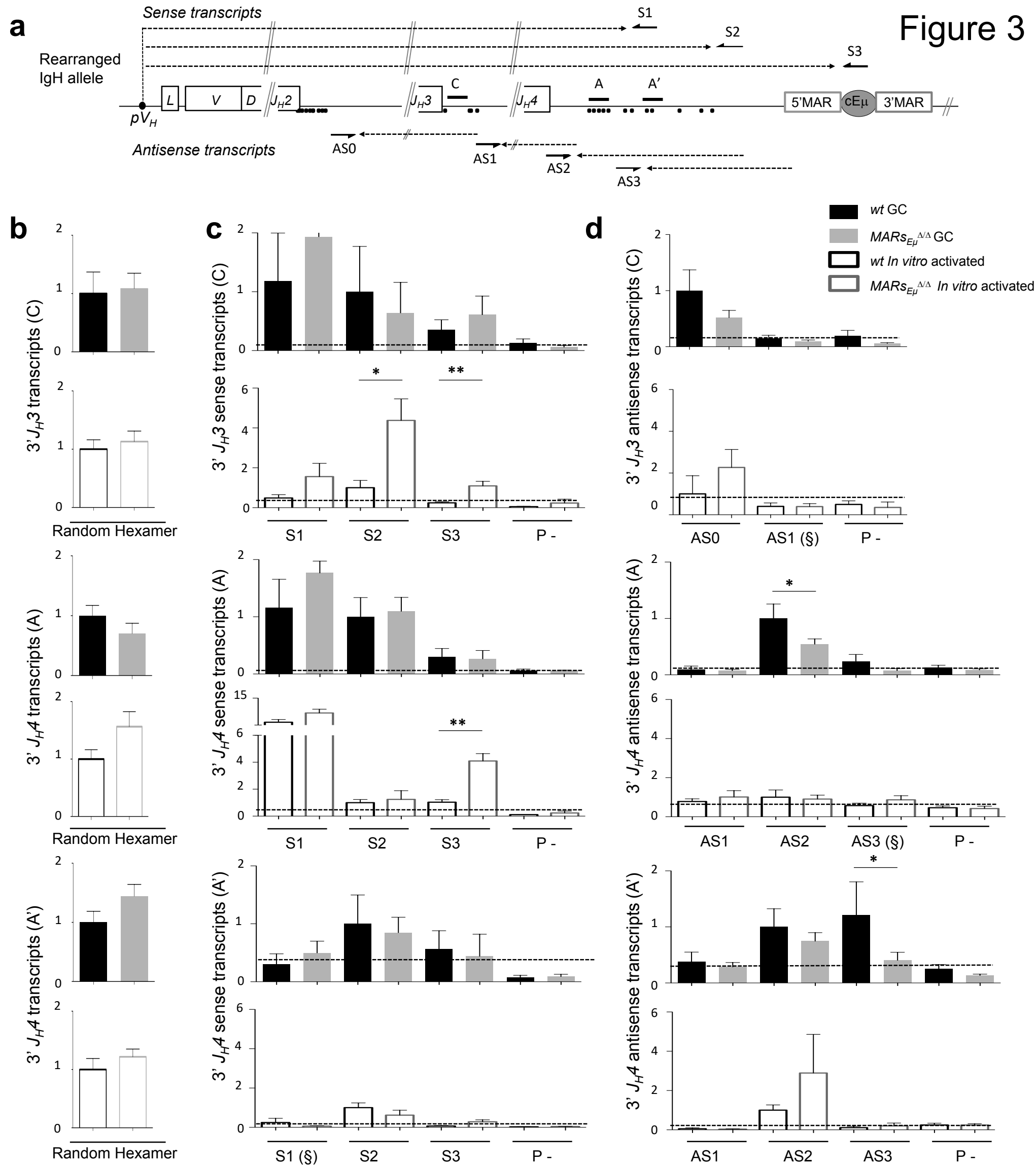


**Figure 3.  $MARs_{E\mu}$  deletion impairs strand-specific transcription of the VDJ unit**

(a) Murine *IgH* locus (not scale) indicating location of primers (S1, S2, S3, AS0, AS1, AS2 and AS3) within introns downstream from  $J_H2$ ,  $J_H3$  and  $J_H4$  used for strand-specific reverse transcription to detect sense/antisense transcripts (dotted arrows). S1, S2 and S3 templates detect sense transcripts; AS0, AS1, AS2 and AS3 templates detect antisense transcripts. Black bars (C, A and A') indicate location of q-PCR probes. Dots indicated antisense transcripts start sites according to Perlot et al. (Perlot et al., 2008). (b) Total primary transcripts quantified with C, A and A' probes in Peyer's patch GC B cells (filled histograms, n=5-19) and *in vitro*-activated B cells (emptied histograms, n=5-6) from *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  mice. (c) Primary sense transcripts from *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  in Peyer's patch GC B cells (filled histograms, n=4-13) and *in vitro*-activated B cells (emptied histograms, n=4-6). (d) Primary antisense transcripts from *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  in Peyer's patch GC B cells (filled histograms, n=7-15) and *in vitro*-activated B cells (emptied histograms, n=4-6).

Baseline level was either provided by a control reaction measuring endogenous priming since devoid of primer (P-) or by one strand-specific template that cannot be detected with the current probe (§). Histograms show mean $\pm$ SEM of two to three independent experiments. \*P<0.05, \*\*P<0.01 (Mann-Whitney U-test, two tailed).





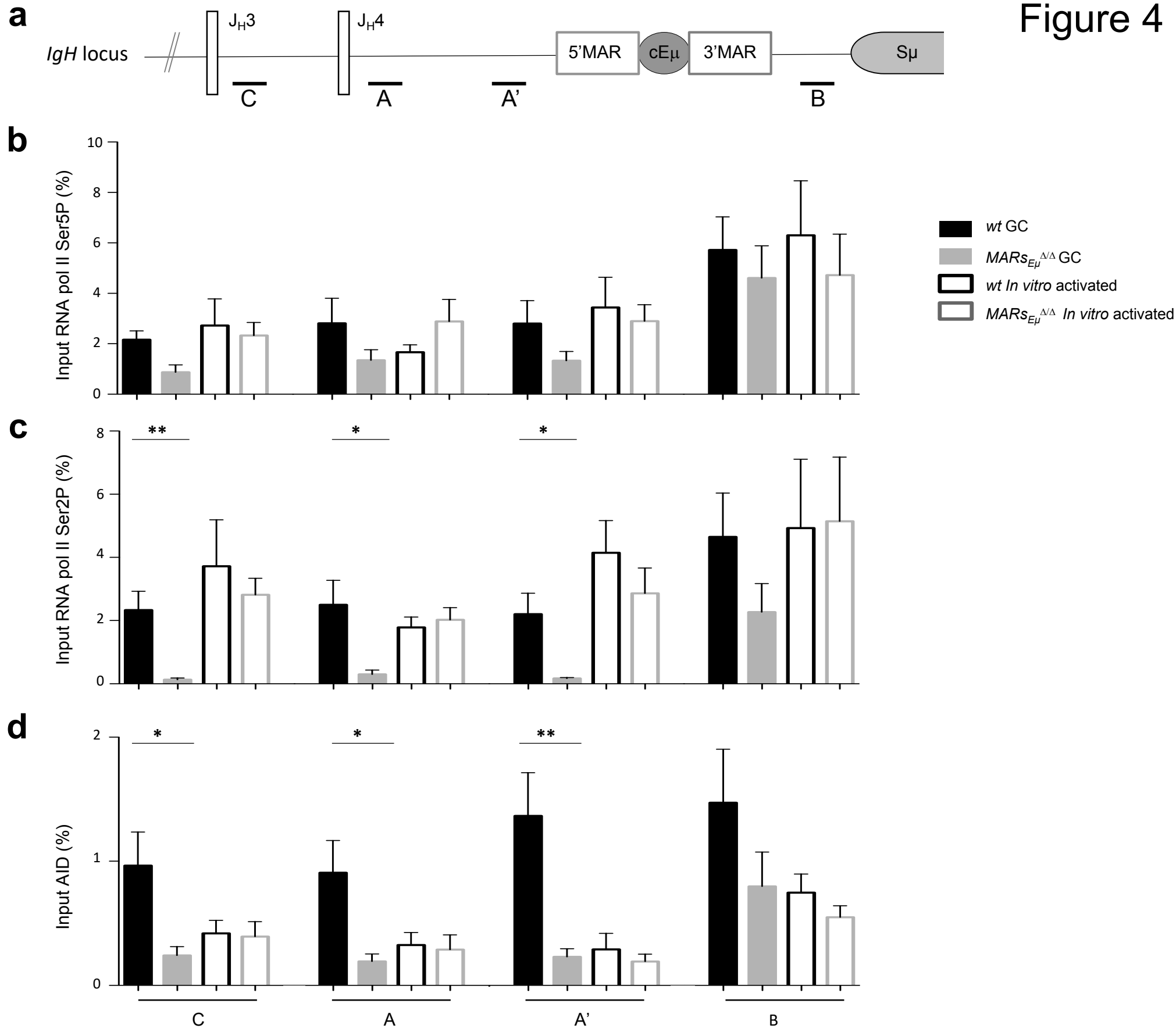


**Figure 4.  $MARS_{E\mu}$  deletion modifies RNA pol II progression upstream from  $E\mu$**

(a) Schematic representation of mouse *IgH* locus (not scale), black bars (C, A, A' and B) indicate q-PCR probes used to quantify DNA template for ChIP experiments. (b) ChIP analysis of RNA pol II Ser5P antibody performed on Peyer's patch cells (Filled histograms, n=5 animal samples) and on *in vitro* activated splenic B cells with LPS + IL4 from *wt* and  $MARS_{E\mu}^{\Delta/\Delta}$  animals (emptied histograms, n=3 to 5). (c) ChIP analysis of RNA pol II Ser2P performed on the same cell subsets (n=4 to 7 animal samples). (d) ChIP analysis of AID performed on the same cell subsets (n=4 to 5 animal samples). ChIP-qPCR values are normalized to the total input DNA. Histograms show mean $\pm$ SEM. \*P<0.05, \*\*P<0.01 (Mann–Whitney U-test, two tailed).



Figure 4







## Supplementary Figure 1

### Annotated nucleotide map of the *IgH-J<sub>H</sub>1* to *E $\mu$* germline region from of 129 wt mice.

Both *J<sub>H</sub>* exons and *coreE $\mu$*  element are indicated by bold characters. Start sites for antisense transcripts are reported as (\*) according to Perlot et *al.* 2008. Location of primers used for strand-specific reverse transcription (S1, S2, S3, AS0, AS1, AS2, AS3) are indicated by double underlines. TaqMan qPCR amplicons (C, A, A') are highlighted in grey.



# Supplementary Figure 1

## J<sub>H</sub>1

**ctactgggtacttcgatgtctggggcgcagggaccacgggtcacctctcctcag**gtaagctggcttttttctttctgcac  
\*  
attccattctgaaatgggaaaagatattctcagatctccccatgtcaggccatctgccacactctgcatgctgcagaag  
\*  
cttttctgtaaggatagggctcttcaactcccaggaaaagagggcagtcagaggctagctgcctgtggaacagtgacaatca  
\*  
tggaatataggcatttacattgtaggctacatgggtagatgggtttttgtacaccactaaaggggtctatgatagtg  
tg

## J<sub>H</sub>2

**actactttgactactggggccaaggcaccactctcacagttcctcag**gtgagtccttacaacctctctcttctattca  
\* \* \* \* \*  
gcttaaatagatcttactgcatttgttgggggggaaatgtgtgtatctgaatttcaggtcatgaaggactagggacacc  
ttgggagtcagaaagggctattgggagccctggctgatgcagacagacatcctcagctcccagacttcatggccagaga  
AS0  
tttatagggatcctggccagcattgccgctaggtccctctcttctatgctttctttgtccctcactggcctccatctga  
gataatcctggagccctagccaaggatcatttattgtcaggggtctaatacattggtgtcacaatgtg

## J<sub>H</sub>3

**cctggtttgcttactggggccaagggactctgggtcactgtctctgcag**gtgagtcctaacttctcccattctaaatgca  
\* C  
tgttggggggattctgagccttcaggaccaagattctctgcaaacgggaatcaagattcaaccctttgtcccaaagtt  
\* \* \*  
gagacatgggtctgggtcagggactctctgcctgctgggtctgtggtgacattagaactgaagtatgatgaaggatctgc  
cagaactgaagcttgaagctctgaggcagaatcttgtccagggcttatcggactcttgtgagaattaggggctgacagtt  
gatgggtgacaatttcagggctcagtgactgtctggtttctctgaggtgaggctggaatataggtcaccttgaagacttaa  
AS1

gaggggtccagggggcttctgcacaggcagggaaacagaatgtggaacaatgacttgaatgggtgattcttgtgtgacac  
caggaattggcataatgtctgagttgccaggggtgattctagtcagactctgggggtttttgtcgggtatagaggaaa  
atccactattgtg

## J<sub>H</sub>4

**attactatgctatggactactgggggtcaaggaaacctcagtcaccgtctcctcag**gtaagaatggcctctccagggtcttt  
\* \* \* \* \* AS2  
atttttaacctttgttatggagttttctgagcattgcagactaatcttggatatcttgcctgagggagccggctgaga  
\* \* A AS3  
gaagttgggaaataaactgtctagggatctcagagcctttaggacagattatctccacatctttgaaaaactaagaatc  
\* S1 \* \* \* \* \*  
tgtgtgatgggtggtgggagtccttggatgatgggatagggactttggaggctcatttgaggagatgctaaaacaat  
A'  
cctatggctggagggatagttggggctgtagttggagatcttccagtttttagaataaaagtattagctgcggaatatac  
\*  
ttcaggaccacctctgtgacagcatttatacagtatccgatgcatagggacaaagagtggagtggggcactttcttttag  
\*  
atctgtgaggaatgttccacactagattgtttaaaacttcatttgttggaggagagctgtcttagtgattgagtcag  
\*  
ggagaaaggcatctagcctcggctctcaaaagggtagttgctg  
S2

## 5' MAR

tctagagaggtctgggtggagcctgcaaaagtccagctttcaaaggaaacacagaagtatgtgtatggaatattagaagat  
gttgcttttactcttaagttgggtcctagggaaaaatagttaaatactgtgactttaaaatgtgagaggggttttcaagta  
ctcatttttttaaagtccaaaattttgtcaatcaatttgaggcttgtttgtgtagaactgacattacttaaagttt  
aaccgaggaatgggagtgaggctctctcataccctattcagaactgacttttaacaataataaattaagtttaaaat  
tttaaatgaattgagcaatggtgagttggagtc

## Core Eμ

**aagatggccgatcagaaccagaacacctgcagcagctggcaggaagcaggtcatgtggcaaggctat**ttgggggaaggga  
S3  
**aaataaaaccactaggtaaacttgtagctgtggtttgaagaagtggttttgaaacactctgtccagccccaccaaaccg**  
**aaagtcaggctgagcaaaacaccacctgggtaatttgcaatttctaaaataagttgag**

## 3' MAR

gattcagccgaaactggagaggtcctcttttaacttattgagttcaaccttttaatttttagcttgagtagttctagttt  
cccaaacttaagtttatcgacttcaaaatgtatttagaattcattttcaaaattaggttatgtaagaaattgaaggac  
tttagtgcttttaatttctaataatatttagaaaacttcttaaaattactctattattcttccctctgattattgggtctc  
cattcaattcttttccaatacccgagcatttacagtgactttgttcatgatcttttttagttggtttgtttgccttac  
tattaagactttgacattctgggtcaaaacggcttcacaaatctttttcaagaccactttctgagtagttcatttttaggag  
aaataacttttttttaaatgaatgcaattatctaga

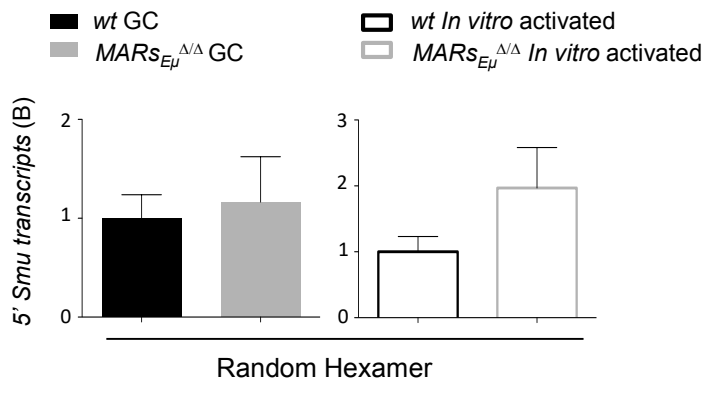
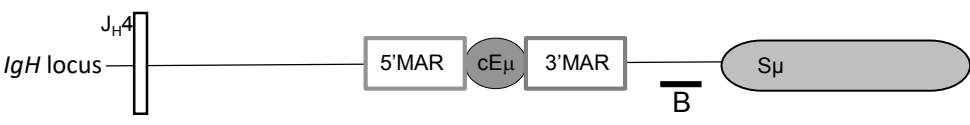
## Supplementary Figure 2

### ***MAR<sub>E $\mu$</sub>* deletion does not modify global primary transcription downstream from the *E $\mu$* enhancer**

Top: Murine *IgH* locus (not scale) indicating TaqMan probe location used for quantified primary transcripts in intron 5' to *S $\mu$* .

Bottom: Relative primary transcripts quantified in intron 5' to *S $\mu$*  in Peyer's patch GC B cells (filled histograms) and *in vitro* activated B cells (emptied histograms) from *wt* and *MAR<sub>E $\mu$</sub>  <sup>$\Delta/\Delta$</sup>  mice. Histograms show mean $\pm$ SEM of two independent experiments with at least 5 mice.*







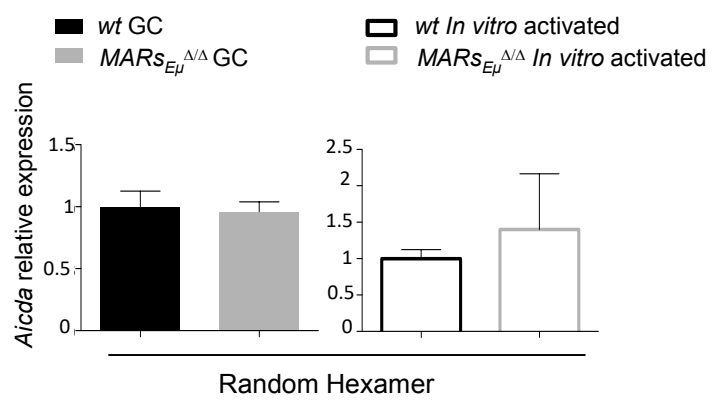
### Supplementary Figure 3

#### ***MARSE $\mu$* deletion does not impair *Aicda* gene expression**

AID relative transcripts quantified in Peyer's patch GC B cells (filled histograms) and *in vitro* activated B cells (emptied histograms) from *wt* and *MARSE $\mu^{\Delta/\Delta}$*  mice. Histograms show mean $\pm$ SEM of two independent experiments with at least 5 mice.









### Supplementary Table 1

#### ***MARs<sub>E $\mu$</sub>* deletion does not impair B cell development in mice**

Bone Marrow and peripheral B cell subsets counts in *wt* and *MARs<sub>E $\mu$</sub>  <sup>$\Delta/\Delta$</sup>*  mice. Absolute numbers are reported as mean $\pm$ SEM. Significance was assessed with Student T test.



## Supplementary Table 1

	<i>wt</i> Cell numbers (x10 <sup>6</sup> )	<i>MARs</i> <sub>E<math>\mu</math></sub> <sup><math>\Delta/\Delta</math></sup> Cell numbers (x10 <sup>6</sup> )	Significance
<b>Bone marrow</b>			
B-lineage cells (B220 <sup>+</sup> )	84.96 $\pm$ 3.326 N=5	75.27 $\pm$ 4.241 N=6	NS
Pre-pro B cells (B220 <sup>+</sup> /CD117 <sup>+</sup> )	2.183 $\pm$ 0.1974 N=6	2.567 $\pm$ 0.2951 N=6	NS
Pro-B cells (IgM <sup>-</sup> /B220 <sup>+</sup> /CD43 <sup>high</sup> )	7.380 $\pm$ 1.131 N=6	7.050 $\pm$ 0.9949 N=6	NS
Pre-B cells (IgM <sup>-</sup> /B220 <sup>+</sup> /CD43 <sup>low</sup> )	12.50 $\pm$ 1.965 N=6	13.54 $\pm$ 2.203 N=6	NS
<b>Spleen</b>			
B-lineage cells (B220 <sup>+</sup> )	58.16 $\pm$ 4.593 N=6	70.68 $\pm$ 3.524 N=6	NS
Naïve mature B cells (B220 <sup>+</sup> /IgM <sup>+</sup> /IgD <sup>+</sup> )	45.09 $\pm$ 3.145 N=6	52.76 $\pm$ 1.874 N=5	NS
Marginal zone B cells (B220 <sup>+</sup> /CD21 <sup>high</sup> /CD23 <sup>low</sup> )	7.235 $\pm$ 1.504 N=6	7.740 $\pm$ 0.5829 N=6	NS
Follicular B cells (B220 <sup>+</sup> /CD21 <sup>low</sup> /CD23 <sup>high</sup> )	42.01 $\pm$ 3.506 N=6	45.75 $\pm$ 2.992 N=6	NS
<b>Peritoneal cavity</b>			
B-lineage cells (B220 <sup>+</sup> )	1.864 $\pm$ 0.3514 N=5	2.517 $\pm$ 0.2998 N=5	NS
B1a cells (CD5 <sup>+</sup> /IgM <sup>+</sup> )	0.5924 $\pm$ 0.1002 N=5	0.8280 $\pm$ 0.1308 N=5	NS
B1b cells (CD5 <sup>-</sup> /IgM <sup>+</sup> )	1.076 $\pm$ 0.2745 N=5	1.474 $\pm$ 0.2010 N=5	NS
<b>Peyer's patches</b>			
B-lineage cells (B220 <sup>+</sup> )	4.765 $\pm$ 1.106 N=6	5.841 $\pm$ 1.172 N=6	NS
B220 <sup>+</sup> /IgA <sup>+</sup> cells	0.5625 $\pm$ 0.1273 N=6	0.8837 $\pm$ 0.2218 N=6	NS
B220 <sup>+</sup> /IgM <sup>+</sup> cells	2.828 $\pm$ 0.5269 N=6	3.782 $\pm$ 0.8263 N=6	NS
Naive B cells (B220 <sup>+</sup> /PNA <sup>low</sup> /Fas <sup>low</sup> )	3.684 $\pm$ 0.6369 N=5	3.556 $\pm$ 0.5026 N=5	NS
Germinal centre B cells (B220 <sup>+</sup> /PNA <sup>high</sup> /Fas <sup>high</sup> )	1.044 $\pm$ 0.8822 N=8	3.258 $\pm$ 1.2 N=11	P=0.0007



## Supplementary Table 2

### ***MAR<sub>S<sub>E $\mu$</sub></sub>*** deletion impaires nucleotide substitution pattern

Spectra of mutations in the intron downstream from *J<sub>H</sub>4* obtained from individual *wt* (n=3, left) and *MAR<sub>S<sub>E $\mu$</sub></sub>*  <sup>$\Delta/\Delta$</sup>  (n=4) mice. Data are presented as percentage of mutations, corrected for the base composition of the sequence.





**wt****MARs<sub>Eμ</sub><sup>Δ/Δ</sup>**Corrected mutation spectra (% events):  
intron 3' to J<sub>H</sub>4 Peyer's patch GC B cells

from \ to	A	T	C	G	Total
A		5	12	10	27
T	8		8	2	18
C	3	20		3	26
G	15	6	8		29

from \ to	A	T	C	G	Total
A		11	9	17	37
T	5		8	2	15
C	4	14		5	23
G	14	4	8		26

from \ to	A	T	C	G	Total
A		12	6	10	28
T	8		8	3	19
C	6	18		4	28
G	14	3	9		26

from \ to	A	T	C	G	Total
A		7	7	17	31
T	4		5	3	12
C	2	20		5	27
G	19	3	7		29

from \ to	A	T	C	G	Total
A		9	7	9	25
T	5		7	4	16
C	6	18		7	31
G	14	6	9		29

from \ to	A	T	C	G	Total
A		8	7	15	30
T	7		9	2	18
C	4	15		6	25
G	16	3	9		28

from \ to	A	T	C	G	Total
A		8	5	14	27
T	4		7	2	13
C	3	20		7	30
G	19	4	5		28



### Article 3

#### **Evidence for *trans*-dominant impact of *IgH* E $\mu$ Matrix Attachment Regions on somatic hypermutation and subnuclear positioning of both *Ig* genes and AID off targets.**

Ophélie Martin, Marie Marquet, Armand Garot, François Boyer, Mylène Brousse, Sébastien Bender, Claire Carrion, Patricia J. Gearhart, Laurent Delpy, Michel Cogné, Sandrine Le Noir and Eric Pinaud.

En parallèle de l'étude des mutations au locus *IgH* dans le modèle homozygote MAR<sub>E $\mu$</sub>  KO, nous avons également quantifié, chez des animaux homozygotes et hétérozygotes porteurs d'allèles *IgH* d'haplotypes différents (*IgH* « a » dépourvu des régions MAR<sub>E $\mu$</sub>  et *IgH* « b » wt), les mutations aux 2 loci *IgH* (mutant et wt) ainsi qu'aux loci des chaînes légères *Ig $\kappa$* . De façon surprenante, la délétion des MAR<sub>E $\mu$</sub>  (au chromosome 12) affecte les mutations introduites au locus *Ig $\kappa$*  (chromosome 6), les modèles hétérozygotes MAR<sub>E $\mu$</sub> <sup>*Δa/wtb*</sup> ont confirmé l'effet *trans*-régulateur lié à la suppression des régions MAR<sub>E $\mu$</sub> . Au-delà de leur effet *cis*-régulateur sur la région située juste en amont (article 2) ; notre étude indique que les régions MAR<sub>E $\mu$</sub>  sont capables d'agir en *trans* (au locus des chaînes légères *Ig $\kappa$* ) mais également, dans le cas de leur suppression sur un seul allèle, en exerçant un effet *trans*-dominant sur le second allèle *IgH*.

Ces résultats nous ont conduit à rechercher un effet *trans*-régulateur des MAR<sub>E $\mu$</sub>  sur les mutations infligées aux cibles illégitimes de AID. Nous avons alors utilisé notre outil DeMinEr pour déterminer la fréquence de mutation de 2 cibles illégitimes de AID : *Bcl6* et *Cd83*, dans un contexte wt et MAR<sub>E $\mu$</sub>  KO. Nos résultats ont confirmé l'action en *trans* des régions MAR<sub>E $\mu$</sub>  sur les 2 cibles étudiées *Bcl6* et *Cd83* : ces dernières se trouvent significativement moins mutées dans les cellules B de CG issues d'animaux mutants.

Compte tenu du potentiel des régions MARs à participer à l'organisation nucléaire et de la capacité des MAR<sub>E $\mu$</sub>  à lier la matrice nucléaire de façon réversible (MARs dites facultatives) ; nous avons émis l'hypothèse que de telles régions pourraient être impliquées dans une régulation spatiale du processus de SHM dans l'espace du noyau des cellules B. Notre étude s'est donc poursuivie par une analyse par FISH 3D du positionnement nucléaire des cibles de AID dans les noyaux de cellules B naïves ou de CG. Nous montrons alors pour la 1<sup>ère</sup> fois, que la position des gènes cibles de AID est influencée par les régions MAR<sub>E $\mu$</sub> . Et bien que certaines données de positionnement sont à compléter, il semblerait que la présence des MAR<sub>E $\mu$</sub>  soit associée à une localisation préférentielle des cibles de AID plus en périphérie du noyau.

Nos observations sont en faveur d'une régulation spatiale du processus de SHM médiée au moins en partie par les régions MAR<sub>E $\mu$</sub>  du locus *IgH*. Nos travaux illustrent que la position relative des gènes d'Ig et des cibles « off targets » influence le processus de SHM initié par l'enzyme AID.



**Evidence for trans-dominant impact of *IgH* E $\mu$  Matrix Attachment Regions on somatic hypermutation and subnuclear positioning of both Ig genes and AID off targets**

**Condensed title: MARs<sub>E $\mu$</sub>  enhance *trans* chromosomal somatic mutations**

Ophélie Martin<sup>1,5</sup>, Marie Marquet<sup>1,5</sup>, Armand Garot<sup>1</sup>, François Boyer<sup>1</sup>, Mylène Brousse<sup>1</sup> Sébastien Bender<sup>1,2</sup>, Claire Carrion<sup>1</sup>, Patricia J. Gearhart<sup>3</sup>, Laurent Delpy<sup>1</sup>, Michel Cogné<sup>1,2,4</sup>, Sandrine Le Noir<sup>1,5\*</sup> and Eric Pinaud<sup>1\*</sup>

<sup>1</sup>*UMR 7276, Centre National de la Recherche Scientifique and Université de Limoges : Contrôle de la Réponse Immune B et des Lymphoproliférations, 2, rue du Dr. Marcland, 87025 Limoges, France*

<sup>2</sup>*Centre Hospitalier Universitaire Dupuytren, 2 Avenue Martin Luther King, F-87042 Limoges, France*

<sup>3</sup>*Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224*

<sup>4</sup>*Institut Universitaire de France, Université de Limoges, 87000 Limoges, France*

<sup>5</sup>O.M., M.M. and S.L.N. contributed equally to this work

\*Corresponding authors: [eric.pinaud@unilim.fr](mailto:eric.pinaud@unilim.fr) and [sandrine.le-noir@unilim.fr](mailto:sandrine.le-noir@unilim.fr)

Correspondence and requests should be addressed to E.P. ([eric.pinaud@unilim.fr](mailto:eric.pinaud@unilim.fr)) and S.L.N. ([sandrine.le-noir@unilim.fr](mailto:sandrine.le-noir@unilim.fr))



## Abstract

Immunoglobulin (Ig) genes undergo genomic rearrangements and mutations in B-lineage cells. In the germinal center structure, Ig heavy (*IgH*) and light chain (*IgL*) loci of activated B cells are targeted by somatic hypermutation (SHM) with the aim to change Ig affinity. SHM are initiated by the cytidine deaminase AID, although Ig genes are AID “legitimates” targets, this enzyme can also inflict “collateral damages” to non Ig genes called “off target” genes, some of them involved in lymphomagenesis. Somatic mutations at *IgH* locus are controlled by several *cis*-regulatory regions including the  $E\mu$  region and the 3' Regulatory Region. Over the past 30 years, multiple models (transgenes or knock out) involving the  $E\mu$  region, that includes a core enhancer element flanked by two regions of anchorage to the nuclear matrix ( $MARs_{E\mu}$ ), underlined its importance in B-lineage cells. However,  $MARs_{E\mu}$  function remains elusive: either transcription activators or silencers. We recently found that  $MARs_{E\mu}$  play an important role in optimizing *IgH* somatic hypermutation, probably through promoting convergent regulatory transcripts efficiently targeted by AID (Martin et al., article 2). Studies of the dynamic processes that regulate chromatin conformation changes and nuclear location arouse a new interest for *cis*-regulatory regions that delimit differentially regulated chromosome domains. Among such DNA regulatory regions, nuclear matrix attachment regions are thought to either constitutively border chromatin domain (constitutive *MARs*) or bind tissue-specific nuclear factors that participate to gene regulation expression (facultative *MARs*). In this study, we propose a specific function for the  $MARs_{E\mu}$  regions acting as a *trans*-regulatory element for the SHM process by promoting nuclear organization.





Antibody maturation affinity is a SHM driven mechanism, leading to point mutation into the variable regions of Ig genes, with the aim to increase Ig affinity. SHM is initiated by AID in germinal center B cells following antigen activation. AID is a cytidine deaminase enzyme responsible for the deamination of single strand DNA cytidine to uridine leading to a DNA U-G mismatch. Single strand DNA which constitute AID substrate is provided during transcription (Peled et al., 2008), however all transcribed regions in GC B cells are not deaminated by AID meaning that transcription alone is not sufficient to recruit AID to its targets (Inlay et al., 2006). The U-G mismatch can be handled by either the replication machinery or by the DNA repair machinery, through the base excision or mismatch repair pathways. In B cells, these DNA repair pathways are error-prone allowing the introduction of frequent point mutation in Ig heavy (*IgH*) and light (*IgL*) chain gene loci (Weill and Reynaud, 2008). As the initiator of SHM, AID enzyme has a strong preference for Ig genes and especially a quite restricted region (about 1 kilobase long) encompassing the rearranged variable exon and its downstream intron of heavy and light chain loci. B cells from the GC harbour a quite high mutation rate within Ig-targeted region that reaches 15 to 20 mutations within this 1000pb region. Although acting primarily on Ig genes, AID can also induced mutations as “collateral damages” to non Ig genes (including oncogenes) so called “AID off-targets” (Pasqualucci et al., 2001). For example, mutations in *Bcl6* are frequent in diffuse large B cell- and follicular lymphomas (Jardin and Sahota, 2005; Jardin et al., 2007). The high risk of AID-induced collateral damages in the B cell genome implies that remodelling events at Ig gene loci (recombination and SHM) are precisely surveyed (Liu et al., 2008). Among *cis*-regulatory elements involved in this survey (transcriptional enhancers, chromatin isolators and anchors...), one best candidate is the enhancer and nuclear scaffolding regions located within J-C intronic region of *IgH* and *IgL* loci first because they are conserved in mice and humans and second because they remain functional upon any locus rearrangement. In the murine *IgH* (Chr. 12) and *IgL* kappa loci (Chr. 6), J-C intronic enhancer region is a combination of both a core enhancer element (cE $\mu$  or cEik) and one or two flanking nuclear Matrix Attachment (*MAR*<sub>S $\mu$</sub>  5' and 3' or *MAR*<sub>Eik</sub> 5') (Cockerill and Garrard, 1986; Cockerill et al., 1987). Studies of the dynamic processes that regulate chromatin conformation changes and nuclear location arouse a new interest for *cis*-regulatory regions that delimit differentially regulated chromosome domains. Among such DNA regulatory regions, nuclear matrix attachment regions are thought to either constitutively border chromatin domain (constitutive MARs) or bind tissue-specific nuclear factors that participate to gene regulation expression (facultative MARs) (Gluch et al., 2008).

So far, the physiological role of MARs flanking Ig genes intronic enhancers was still elusive, although it was proposed a kappa light chain SHM-enhancing function for the *MAR*<sub>Eik</sub> (Yi et al., 1999). Our recent finding using a *MAR*<sub>S $\mu$</sub>  KO mouse model carrying the deletion of the two endogenous *IgH* *MAR*<sub>S $\mu$</sub>  indicate that these regions are able to facilitate SHM to the *IgH* variable exon by enhancing antisense convergent transcription: the *MAR*<sub>S $\mu$</sub>  induced antisense non-coding RNA increased AID recruitment upstream from the core intronic E $\mu$  enhancer (Martin et al., article 2). Nevertheless, the precise SHM targeting mechanism recruiting AID on Ig



and especially to non-Ig genes is still poorly understood. Using our *MARs<sub>Eμ</sub>* KO mouse model, we assessed in this study an unexpected dominant effect of *MARs<sub>Eμ</sub>* deletion on SHM within both Ig heavy and light chain genes (*IgH* and *Igκ* loci) but also within AID off target genes (*Bcl6* and *Cd83* loci). The global mutation decrease in multiple AID-targets is coupled to a change in relative allelic subnuclear positioning of the corresponding genes and to a defect in AID recruitment. Our study proposes that dynamic repositioning of AID-targeted genes in B cell nuclei is a mechanism involved in optimizing SHM.



## RESULTS

### Heterozygous $MARS_{E\mu}$ deletion decreased SHM at $IgH$ of the homologous chromosome

We previously generated a mouse mutant line carrying endogenous deletion of  $IgH$  matrix attachment regions that flank the core  $cE\mu$  enhancer element (Martin et al., article 2). The resulting  $IgH$  allele, so-called  $MARS_{E\mu}^{\Delta}$  is similar to that generated by Sakai et al. 1999. For the purposes of this study, we compared  $wt$  animals to either F1 heterozygous animals carrying  $wt$   $IgH$  allele of “b” haplotype and  $MARS_{E\mu}$ -deficient allele of “a” haplotype ( $MARS_{E\mu}^{\Delta a}/wt^b$ ) or homozygous mutants  $MARS_{E\mu}^{\Delta\Delta}$ . In heterozygous models, IgM-allotype expression in developing B-lineage cells reflects competition between  $IgH$   $wt$  (“b” allotype) and mutant (devoid of  $MARS_{E\mu}$ , “a” allotype) alleles. We first verified, by flow cytometry, that heterozygous mutant displayed similar proportions of  $IgH^a$  and  $IgH^b$ -expressing B cells in Peyer’s patches (Fig. 1 a). In the heterozygous context,  $MARS_{E\mu}$  deletion still allowed normal  $IgH$  transcription in Peyer’s patches naive and GC B cells since intracellular IgM<sup>a</sup> mean fluorescence was similar in  $MARS_{E\mu}^{\Delta a}/wt^b$  and  $wt^a/wt^b$  F1 cells (Fig. 1 b). When SHM was analyzed in splenic GC B cells sorted from SRBC-immunized  $wt^a/wt^b$  mice, mutation frequency quantified downstream from the rearranged  $J_H3$  and  $J_H4$  segments were comparable in both  $IgH^a$  and  $IgH^b$  alleles: respectively 10.6 and 11.9‰ (Fig. 1 c). In  $MARS_{E\mu}^{\Delta a}/wt^b$  mice, mutation frequency was decreased within the  $IgH^a$  mutant locus (7.8 ‰) but against all odds, SHM in the  $wt$   $IgH^b$  allele was also lowered (8.6‰) (Fig. 1 c), revealing a dominant effect of the  $MARS_{E\mu}$  deletion for  $IgH$  somatic mutation. To confirm this, we quantified mutations in Peyer’s patches B cells by next generation sequencing and found the same dominant effect of a single  $MARS_{E\mu}$  deletion on SHM of both  $IgH$  alleles since mutation frequency raised 7.6 and 8.5‰ in  $IgH$ - $MARS_{E\mu}^{\Delta a}$  and  $IgH$ - $wt^b$  alleles respectively (Fig. 1. d). We also carefully quantified SHM in functional (in frame) and non-functional (out of frame)  $IgH$  alleles. In the case of  $MARS_{E\mu}^{\Delta a}/wt^b$  cells, both the functional and non-functional alleles were equally affected (Fig. 1 d) indicating that an IgM-BCR driven by a  $MARS_{E\mu}$ -deficient allele did not suffer of any bias at the cell selection level. In the past, *trans*-complementation between mouse  $IgH$  alleles was described in the context of CSR (Dougier et al., 2006): in this case, a CSR rescue was carried out by the functional 3’  $IgH$  enhancers of the second allele by providing an acceptor site for recombination in *trans* when acceptor sites were not available in *cis*. Although devoid of enhancer activity, our data assigned to  $MARS_{E\mu}$  a quite different *trans*-acting function for SHM, which occurred independently of any inter-allelic recombination.

### $IgH$ $MARS_{E\mu}$ deletion decreased SHM at $IgL \kappa$ loci

It was then questionable whether the SHM *trans*-dominant effect of  $MARS_{E\mu}$  could apply to loci other than  $IgH$ , and notably to Ig light chain  $\kappa$  loci that is expressed in 95% of mouse peripheral B cells. We then quantified SHM in the 554 bp region located just downstream from the  $J\kappa5$  segment, in immunized splenic GC B cells from heterozygous mice, and found a clear two-fold decrease of SHM frequency in  $MARS_{E\mu}^{\Delta a}/wt^b$  mutants (6.2‰) compared to  $wt^a/wt^b$  (12.4‰) (Fig. 1 c, bottom). Rationally, the dominant *trans*-acting effect of  $MARS_{E\mu}$



deletion was also observed in Peyer's Patches GC B cells from homozygous animals carrying the deletion of either the  $MARs_{E\mu}$  region only ( $MARs_{E\mu}^{\Delta\Delta}$ ) or a larger deletion including both  $MARs_{E\mu}$  and  $cE\mu$  enhancer ( $E\mu^{\Delta\Delta}$ ) (Marquet et al., 2014). When measuring  $Ig\kappa$  SHM frequency in the same region, both  $MARs_{E\mu}^{\Delta\Delta}$  and  $E\mu^{\Delta\Delta}$  mutants displayed consistent reduction: 8.1 and 7.8‰ compared to 12.1‰ in *wt* cells, respectively (Fig. 1 e). GC B cells from homozygous mutants devoid of  $MARs_{E\mu}$  showed an increased proportion of non-mutated sequences (respectively 19% and 47% for mutants compared to only 8% in *wt*, Fig. 1 e), this feature being also observed in splenic GC B cells from heterozygous  $MARs_{E\mu}^{\Delta a}/wt^b$  (57% of unmutated sequences in mutants heterozygous mutants compared to only 20% in *wt*, Fig. 2 c, bottom). Using next generation sequencing in Peyer's patch GC B cell population sorted from homozygous  $MARs_{E\mu}^{\Delta\Delta}$  animals, we found a less marked but consistent decrease of SHM that we could explain by the pathogen-free housing conditions of this batch of animals (7‰ in mutants compared to 9‰ in *wt*) (Fig. 1 f).

Since it was established that SHM in Ig V segments is coupled to transcription initiated at V promoters (Fukita et al., 1998), we quantified  $Ig\kappa$  locus primary transcription and found that it was unaffected by  $MARs_{E\mu}^{\Delta}$  mutation (Fig. 1 g, left panel). This was consistent with normal levels of intracellular  $Ig\kappa$  protein levels in  $MARs_{E\mu}^{\Delta\Delta}$  naive and GC B cells (Fig. 1 g right panel). The decreased mutation rate observed at  $Ig\kappa$  locus in our mutant animals suggest that  $MARs_{E\mu}$  also act as global *trans*-chromosomal regulators of Ig genes SHM in GC B cells. Altogether with previous studies showing that normal SHM and transcription occurred at  $Ig\kappa$  loci of  $IgH$  3'RR-deficient B cells (Rouaud et al., 2013), we can conclude that the crosstalk between heavy and light chain loci for SHM is imposed by *MAR* regions but not by transcriptional enhancers.

### ***IgH* $MARs_{E\mu}$ deletion decreased SHM within *Bcl6* and *Cd83* off target genes**

Since  $MARs_{E\mu}$  deletion in GC B cells induces a *trans*-acting effect on SHM targeted at both *IgH* and *IgL* genes, we asked whether these regions could be involved in targeting SHM inflicted to the AID-off targets that have been identified in the mouse (Liu et al., 2008). We quantified SHM on the two main off target *Bcl6* and *Cd83* that have been described as poorly repaired in a *wt* context (Liu et al., 2008; Álvarez-Prado et al., 2018). We took advantage of our DeMinEr tool, dedicated to identify rare nucleotide substitutions, to precisely quantify mutation within regions of both *Bcl6* and *Cd83* genes that have been shown to be targeted by AID (Liu et al., 2008; Martin et al., 2018). In our assay, mutation frequencies in *wt* GC B cells are in agreement with values described in the literature by both Schatz and Ramiro's groups (Liu et al., 2008; Álvarez-Prado et al., 2018): 93.1 and 50.0 mutations per  $10^5$  bases analyzed respectively for *Bcl6* and *Cd83*. In both off target genes, deletion of  $MARs_{E\mu}$  lead to a significant decrease of AID-induced mutations: reaching respectively 57.1 ( $p=0.0025$ ) and 32.5 ( $p=0.0206$ ) mutations per  $10^5$  bases in *Bcl6* and *Cd83*. We also quantified primary transcription within the corresponding regions of *Bcl6* and *Cd83* genes in naive and GC B cells of both





genotypes (Fig. 2 b). Similarly to what we observed in the *Igk* locus, primary transcription did not display significant differences between *wt* and mutant mice.

The decrease in AID-induced mutations observed in both *Igk* and off target genes upon *MARs<sub>Eμ</sub>* deletion suggests that such regions of the *IgH* locus participate to a global mechanism involved in the recruitment of AID to its targets. According to Skok group hypothesis, proposing that AID-targeted genes often interact with *IgH* (Rocha et al., 2012), it was important to determine if *MARs<sub>Eμ</sub>* deletion could influence subnuclear positioning of AID-target genes.

### ***IgH MARs<sub>Eμ</sub>* deletion modifies subnuclear positioning of AID targets (Ig and non-Ig genes)**

At the DNA fibre level, spatial organization can affect gene regulation: beside the fact that chromatin composition directly shapes gene expression, three-dimensional chromatin organisation (that delimits differentially regulated chromosomal territories “CT”) is emerging as a critical gene regulatory mechanism. Indeed, large chromatin domains are in close contact with the nuclear periphery in lamina-associated domains (LADs) that help organize chromosomes inside the nucleus and have been associated with gene repression (van Steensel and Belmont, 2017). It is admitted that distal regulatory elements such as enhancer clusters (super-enhancers) together with chromatin anchors could regulate target genes on the same (*cis*) (Downen et al., 2014) or on different chromosomes (*trans*) (Spilianakis et al., 2005; Williams et al., 2010). All these emerging 3D nuclear organisation studies and the abundant literature describing Ig gene positioning in B-lineage cells (Proudhon et al., 2015) allow to draw a model for Ig gene preferential positioning in B cells, where heavy and light chain loci dynamically cohabit for specific functions (Park et al., 2014).

To look for any potential *trans*-chromosomal crosstalk between AID-targeted genes in cells undergoing SHM, we performed Fluorescence *in situ* Hybridization in 3D-preserved nuclei (3D-FISH). We then measured distances, corrected for any change in cell volume, to estimate the relative positioning of *IgH*, *Igk*, *Bcl6* and *Cd83* loci (Fig. 3 a) in interphase nuclei of naive and GC B cells sorted from Peyer’s patches of *wt* and *MARs<sub>Eμ</sub><sup>ΔΔ</sup>* animals. Since MARs have the potency to attach genomic regions to particular nuclear structures, especially lamin-enriched compartments (Gluch et al., 2008), we first estimated the relative position of the gene carrying the MAR-deletion (*IgH*) in the nuclear space by measuring its distances to the nuclear periphery (Fig. 3 b, left); the same was done for *Igk* (Fig. 3 b, right), *Bcl6* and *Cd83* (Fig. 3 c). In the *wt* context and in both naive and activated B cells from the GC, *IgH* alleles are located in a region close to the nuclear periphery, their relative distances to this compartment did not significantly changes between these two stages. In the absence *MARs<sub>Eμ</sub>*, B cell activation induced a marked repositioning of *IgH* to the periphery (p=0.001), although in this case, *MARs<sub>Eμ</sub>*-deficient alleles are still more distant to the periphery than in *wt* GC B (Fig. 3 b, left). The most remarkable effect observed in the absence of *MARs<sub>Eμ</sub>* is that *IgH* loci are systematically more distant to the nuclear periphery (Fig. 3 b, left): *IgH*-periphery corrected distances increased in GC cells when compared to



*wt* ( $p=0.0329$ ), the difference is even more marked in naive cells ( $p=0.004$ ). Such significant changes regarding distances to the nuclear periphery was also true for *Igκ* loci (Fig. 3 b, left). In both naive and GC B cells, the absence of *MARs<sub>Eμ</sub>* significantly increased *Igκ*-corrected distance to the periphery when compared to *wt* counterparts ( $p=0.001$  in naive;  $p=0.0013$  in GC). Concerning *Bcl6* and *Cd83* off targets; the pattern of gene relative positioning to the nuclear periphery seems to mimic the *Igκ* gene changes pattern (Fig. 3 c). *These data have to be completed by measuring off-targets to periphery distances on a consistent number of nuclei.*

Our data suggests that *MARs<sub>Eμ</sub>* address *IgH* locus positioning close to the nuclear periphery in naïve cells and maintain it at this optimal location in GC B cells.

We also evaluated the crosstalk between *IgH* and the other AID-targeted genes by measuring relative intergenic distances in the same cell populations (Fig. 4); in this case all measurements evaluate only strengths that maintain *IgH* locus and one partner at distance but does not reflect positioning in the nuclear volume. In all cases, distances between *IgH* and its AID-targeted partner display similar pattern: in the *wt* context, intergenic distances remain equivalent between naive and GC B cells while *MARs<sub>Eμ</sub>* deletion induces a global disruption in *IgH-Igκ*, *-Bcl6* or *-Cd83* gene distances (Fig. 4). In naive B cells, *IgH* and its partners are in closer proximity in the absence of *MARs<sub>Eμ</sub>* and this could be explained by the preferred location of *IgH* and *Igκ* loci closer to the nuclear center (Fig. 3 b); in GC B cells devoid of *MARs<sub>Eμ</sub>*, the increased distance between *IgH* and its partners is made by the relocation of *IgH* closer to the periphery (Fig. 3 b left). Altogether, these data are in favour of a critical role for *IgH MARs<sub>Eμ</sub>* in subnuclear positioning of AID-targeted genes.

### ***IgH MARs<sub>Eμ</sub>* deletion impairs AID recruitment within *Igκ* and AID off targets**

Since mutations are introduced by AID enzyme, we estimated AID enrichment in legitimate (*Igκ*) and illegitimate (*Cd83*) gene targets by ChIP using an antibody specific to AID. The ChIP-qPCR experiments were performed on *wt* and *MARs<sub>Eμ</sub><sup>ΔΔ</sup>* samples prepared from either *in vitro* activated B cells (a B population that does not undergo SHM) or Peyer's patch GC-enriched cell population (including SHM-targeted B cells). *In vitro* activated B cells from both genotypes displayed comparable levels of AID enrichment in *Igκ* and *Cd83* genes. In agreement with the decreased SHM observed in such genes, Peyer's patch cell samples from *MARs<sub>Eμ</sub><sup>ΔΔ</sup>* animals failed to properly accumulate AID in regions known to be targeted by SHM of *Igκ* and *Cd83* genes: the intronic region downstream from *Jκ5* segment ( $p=0.0079$ ) (Fig. 5a) and the intronic region downstream from *Cd83* exon 1 ( $p=0.0159$ ) (Fig. 5b). Our results suggest that *MARs<sub>Eμ</sub>* are able to modulate SHM by affecting AID recruitment not only in *IgH* locus as previously shown (Martin et al., article 2) but also in other of its targets in GC B cells: *Igκ* and *Cd83*.



## Concluding remarks

Our study shows that while previously considered as facultative *MARs* (Gluch et al., 2008), *MARs<sub>E $\mu$</sub>*  harboured the potential not only to locally recruit SHM but also to cause dynamic changes of nuclear structures. The surprising *trans*-dominant effect of *MARs<sub>E $\mu$</sub>*  deletion, impacting simultaneously nuclear positioning and SHM of both Ig loci (*IgH* and *IgL*) and off targets genes (*Bcl6* and *Cd83*), revealed an additional level of regulation for targeting mutations in activated B cells. We could propose that *MARs<sub>E $\mu$</sub>*  enforced optimal nuclear positioning of *IgH* and *Ig $\kappa$*  loci for SHM. Such function could be complementary to the one of *IgH* and *Ig $\kappa$*  3' regulatory regions, known to strongly activate SHM through transcription and AID specific targeting of their respective loci (Inlay et al., 2006; Rouaud et al., 2013).

In one hand, GC B cells benefits from SHM-optimized positioning of Ig genes permitted by *MARs<sub>E $\mu$</sub>*  but in another hand, the *trans*-acting function of *MARs<sub>E $\mu$</sub>*  also drive illegitimate targets to a nuclear space prone to AID targeting.

The next step in order to elucidate the function of intronic *MARs* for Ig genes SHM is to identify the effector proteins involved in such a mechanism. Some candidates, mostly ubiquitous nuclear proteins are the cut-like factor CUX1 (Wang et al., 1999) and AT-rich binding factors such as proteins of the ARID or SATB families (Ratliff et al., 2014; Yokota and Kanakura, 2014).

Overall, the *IgH* matrix attachment regions seems to be involved in SHM spatial regulation, such regions could address AID targets in specific B cells nucleus area where the mutation machinery is active.



## Material and Methods

***MARs<sub>E $\mu$</sub>  <sup>$\Delta/\Delta$</sup>*  mice.** Mice used for this work were obtained as previously described (article 2). Mice models used for this study were at least 2-month old male and female wild type (*wt*) and *MARs<sub>E $\mu$</sub>  <sup>$\Delta/\Delta$</sup>*  mice, all inbred in mixes B6;129S genetic background. For the needs of this study, additional breeding to *wt* C57Bl/6 strain (carrying an *IgH* allele of “b” haplotype) was done in order to generate F1 heterozygous animals “*MARs<sub>E $\mu$</sub>  <sup>$\Delta/\Delta$</sup> /wt<sup>b</sup>””. *Wt* and *MARs<sub>E $\mu$</sub>  <sup>$\Delta/\Delta$</sup>*  mice were immunised by intraperitoneal injection of 200 $\mu$ L of sheep red blood cells and sacrificed 8 days after.*

**Cell sorting.** Flow cell sorting was performed using ARIA 3 instrument (BD Biosciences) on single-cell suspensions from fresh collected organs. Once washed with 2% fetal calf serum-PBS 1X, cells were labeled with  $\alpha$ B220-APC (1/200) and  $\alpha$ GL7-FITC (1/100) antibodies (BD Biosciences) and sorted based on distinct gates defined as germinal center (GC) B cells (B220<sup>+</sup>/GL7<sup>+</sup>) and naive B cells (B220<sup>+</sup>/GL7<sup>-</sup>).

**Cell culture.** Splenocytes were collected, after red blood cells lysis, CD43<sup>+</sup> cells were depleted using anti-CD43 MicroBeads (Miltenyi Biotec). CD43<sup>-</sup> splenic B cells were cultured for 3 days at a density of 1x10<sup>6</sup> cells per ml in RPMI 1640 supplemented in 10% serum calf fetal, sodium pyruvate (Lonza), amino acid (NEAA 100x Lonza) and Penicillin-Streptomycin (Gibco) with 1 $\mu$ g/ml LPS (Invivogen) and 20ng/ml IL4 (PeProTech) (for ChIP experiments).

**SHM assays.** For Ig genes, SHM analysis was either performed by cloning followed by classical Sanger method as described (Rouaud et al., 2013) or performed directly on PCR products by next generation sequencing using the GS-Junior apparatus (Roche). In this case and according to the manufacturer’s instructions, adaptor sequences were added to the previous amplification primers sequence in order to be compatible with the GS-Junior sequencing technology. Amplifications were performed with Phusion<sup>®</sup> High-Fidelity DNA Polymerase (New England Biolabs) according to the following program: DNA was denatured 40 s at 98°C and then submitted to 38 cycles consisting of 98°C for 10 s, 68°C for 30 s and 72°C for 30 s, and 1 cycle at 72°C for 10 min. PCR products were first purified using NucleoSpin kit (Macherey-Nagel) followed by Ampure beads purification (Beckman Coulter). PCR products were subjected to “PCR emulsion step” (GS-Junior+ emPCR Kit (Lib-A), Roche) and sequenced using GS junior sequencing kit XL+ (Roche) according to the manufacturer’s instructions. For off target genes, SHM analysis was performed by next generation sequencing using the Ion Proton system as previously described in Martin et al., 2018.

**RT-PCR and Q-PCR.** Total RNA was prepared using TRIzol reagent (Ambion) procedures. RNA samples were first treated with DNase I (Invitrogen) for 15 min at 25°C. RT was performed on 200 ng of total RNA with random hexamers priming using the highcapacity cDNA reverse transcription kit (Applied Biosystems). Each real-time qPCR reaction was performed in duplicate on 10 ng of RNA equivalent, using TaqMan Universal or SYBR green Mastermix (TAKARA) on StepOnePlus system (Applied Biosystems). Each real-time qPCR reaction was performed in duplicate on 10 ng of RNA equivalent, using TaqMan Universal Mastermix (TAKARA) on





StepOnePlus system (Applied Biosystems). Primary transcription at *Igκ* locus was performed as previously described (Rouaud *et al.*, 2013). *Bcl6* and *Cd83* transcription was analyzed using TaqMan probe (Mm00477633\_m1 *Bcl6* and Mm00486868\_m1 *Cd83*, Applied Biosystem). Relative mRNA levels were normalized to *Gapdh* transcripts with the appropriate TaqMan probe (Mm99999915\_g1, Applied Biosystem). Data were analyzed by comparing threshold cycle (*CT*) values according to the  $2^{-(\Delta\Delta CT)}$  method.

**Chromatin immunoprecipitation.** ChIP-IT HighSensitivity kit (Active Motif) was used according to the manufacturer's instructions. ChIP experiments were performed on isolated Peyer's patches cells and on *in vitro* activated B cells. Briefly,  $5 \times 10^6$  cells were cross-linked with 1% formaldehyde at room temperature for 15 min. The reaction was stopped by adding glycine to a final concentration of 0.125 M. After the cross-linking reaction, cell lysis and sonication were performed on Vibracell sonicator (Sonics and Materials, Inc). For immunoprecipitation, 4 μg of each antibody was added to the lysate at 4 °C overnight. Polyclonal rabbit antibodies directed against AID were previously described (Kohli *et al.*, 2010). Immunoprecipitated DNA was used for real-time qPCR. The quantitative PCR signal was normalized using the % input method after subtraction of IgG background (normal rabbit IgG were purchased from Millipore). Primers used for ChIP-qPCR analysis were the same as described above for *Igκ* and the following primers were designed for *Cd83*: Fw-*Cd83* 5'- cttacgccgctctgtttctc-3' and Rev-*Cd83* 5'- aacccagtgaggacttacct-3'.

**3D FISH.** *Wt* and *MAR<sub>SEμ</sub><sup>ΔΔ</sup>* naive and GC B cells from Peyer's patches were dropped onto poly-L-lysine slides and fixed at room temperature for 10 min with 2% paraformaldehyde. Cell permeabilization was performed at room temperature for 10 min in pepsine 0.02%/HCl 10mM. DNA denaturation was performed at 72°C for 5 min in formamide 70%/2X SSC followed by cold dehydration. Hybridization was performed with 200ng of DNA probes: *IgH* (BAC 199M11) labelled with biotin-11-dCTP (Life Technologies) or *Igκ* (RP23-34PD19), *Bcl6* (RP23-127N11) and *Cd83* (RP23-150M22) labelled with digoxigenin-11-dUTP (Roche), using random priming (BioPrime, Life Technologies) combined with 3μg of Cot-1 DNA (Invitrogen) and 2μg of salmon sperm DNA in 50% formamide/10% dextran sulfate/2X SSC. DNA probes were denatured at 95°C for 5 minutes and pre-annealed at 37°C for 1h and then hybridized overnight at 37°C in a humid chamber. The next day, cells were washed for 5 min in 1X SSC at 72°C and for 5 min at room temperature in 1X SSC. Cells were then incubated for 1h at room temperature with FITC-conjugated streptavidin (1/200, BD Biosciences) and rhodamine-conjugated anti-digoxigenin antibodies (1/200, Roche) diluted in 4XSSC/3%BSA and then washed three times during 5 minutes in 2X SSC. Finally, slides were mounted in Prolong Gold Antifade Mountant containing DAPI (ThermoFisher Scientific). Images were acquired along the Z-axis with an epifluorescence microscope (Nikon) and NIS-element software. Optical sections separated by 0.2μm were collected, stacked images were deconvoluted using Huygens software (SVI) and distances were measured using Volocity 3D Image Analysis software (Perkin Elmer) and normalized to nucleus diameter.

**Statistical analysis.** Statistical tests were performed using GraphPad Prism.



## References

- Álvarez-Prado, Á.F., P. Pérez-Durán, A. Pérez-García, A. Benguria, C. Torroja, V.G. de Yébenes, and A.R. Ramiro. 2018. A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J. Exp. Med.* doi:10.1084/jem.20171738.
- Cockerill, P.N., and W.T. Garrard. 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell.* 44:273–282.
- Cockerill, P.N., M.H. Yuen, and W.T. Garrard. 1987. The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *J. Biol. Chem.* 262:5394–5397.
- Dougier, H.-L., S. Reynaud, E. Pinaud, C. Carrion, L. Delpy, and M. Cogné. 2006. Interallelic class switch recombination can reverse allelic exclusion and allow trans-complementation of an *IgH* locus switching defect. *Eur. J. Immunol.* 36:2181–2191. doi:10.1002/eji.200535529.
- Downen, J.M., Z.P. Fan, D. Hnisz, G. Ren, B.J. Abraham, L.N. Zhang, A.S. Weintraub, J. Schujiers, T.I. Lee, K. Zhao, and R.A. Young. 2014. Control of cell identity genes occurs in insulated *neIgH* neighborhoods in mammalian chromosomes. *Cell.* 159:374–387. doi:10.1016/j.cell.2014.09.030.
- Fukita, Y., H. Jacobs, and K. Rajewsky. 1998. Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity.* 9:105–114.
- Gluch, A., M. Vidakovic, and J. Bode. 2008. Scaffold/matrix attachment regions (S/MARs): relevance for disease and therapy. *Handb Exp Pharmacol.* 67–103. doi:10.1007/978-3-540-72843-6\_4.
- Inlay, M.A., H.H. Gao, V.H. Odegard, T. Lin, D.G. Schatz, and Y. Xu. 2006. Roles of the Ig kappa *IgHt* chain intronic and 3' enhancers in Igk somatic hypermutation. *J. Immunol.* 177:1146–1151.
- Jardin, F., P. Ruminy, C. Bastard, and H. Tilly. 2007. The BCL6 proto-oncogene: a leading role during germinal center development and lymphomagenesis. *Pathol. Biol.* 55:73–83. doi:10.1016/j.patbio.2006.04.001.
- Jardin, F., and S.S. Sahota. 2005. Targeted somatic mutation of the BCL6 proto-oncogene and its impact on lymphomagenesis. *Hematology.* 10:115–129. doi:10.1080/10245330400026105.
- Liu, M., J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, and D.G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature.* 451:841–845. doi:10.1038/nature06547.
- Marquet, M., A. Garot, S. Bender, C. Carrion, P. Rouaud, S. Lecardeur, Y. Denizot, M. Cogné, and E. Pinaud. 2014. The E $\mu$  enhancer region influences H chain expression and B cell fate without impacting IgVH repertoire and immune response in vivo. *J. Immunol.* 193:1171–1183. doi:10.4049/jimmunol.1302868.
- Martin, O.A., A. Garot, S. Le Noir, J.-C. Aldigier, M. Cogné, E. Pinaud, and F. Boyer. 2018. Detecting Rare AID-Induced Mutations in B-Lineage Oncogenes from *H/IgH*-Throughput Sequencing Data Using the Detection of Minor Variants by Error Correction Method. *J. Immunol.* doi:10.4049/jimmunol.1800203.
- Park, S.-K., Y. Xiang, X. Feng, and W.T. Garrard. 2014. Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. *Genes Dev.* 28:1159–1164. doi:10.1101/gad.237479.114.



- Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R.S. Chaganti, R. Küppers, and R. Dalla-Favera. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 412:341–346. doi:10.1038/35085588.
- Peled, J.U., F.L. Kuang, M.D. Iglesias-Ussel, S. Roa, S.L. Kalis, M.F. Goodman, and M.D. Scharff. 2008. The biochemistry of somatic hypermutation. *Annu. Rev. Immunol.* 26:481–511. doi:10.1146/annurev.immunol.26.021607.090236.
- Pinaud, E., A.A. Khamlichi, C. Le Morvan, M. Drouet, V. Nalesso, M. Le Bert, and M. Cogné. 2001. Localization of the 3' *IgH* locus elements that effect long-distance regulation of class switch recombination. *Immunity*. 15:187–199.
- Proudhon, C., B. Hao, R. Raviram, J. Chaumeil, and J.A. Skok. 2015. Long-Range Regulation of V(D)J Recombination. *Adv. Immunol.* 128:123–182. doi:10.1016/bs.ai.2015.07.003.
- Ratliff, M.L., T.D. Templeton, J.M. Ward, and C.F. Webb. 2014. The *BrIgHt* Side of Hematopoiesis: Regulatory Roles of ARID3a/*BrIgHt* in Human and Mouse Hematopoiesis. *Front Immunol.* 5:113. doi:10.3389/fimmu.2014.00113.
- Rocha, P.P., M. Micsinai, J.R. Kim, S.L. Hewitt, P.P. Souza, T. Trimarchi, F. Strino, F. Parisi, Y. Kluger, and J.A. Skok. 2012. Close proximity to *IgH* is a contributing factor to AID-mediated translocations. *Mol. Cell*. 47:873–885. doi:10.1016/j.molcel.2012.06.036.
- Rouaud, P., C. Vincent-Fabert, A. Saintamand, R. Fiancette, M. Marquet, I. Robert, B. Reina-San-Martin, E. Pinaud, M. Cogné, and Y. Denizot. 2013. The *IgH* 3' regulatory region controls somatic hypermutation in germinal center B cells. *J. Exp. Med.* 210:1501–1507. doi:10.1084/jem.20130072.
- Sakai, E., A. Bottaro, L. Davidson, B.P. Sleckman, and F.W. Alt. 1999. Recombination and transcription of the endogenous Ig heavy chain locus is effected by the Ig heavy chain intronic enhancer core region in the absence of the matrix attachment regions. *Proc. Natl. Acad. Sci. U.S.A.* 96:1526–1531.
- Spilianakis, C.G., M.D. Lalioti, T. Town, G.R. Lee, and R.A. Flavell. 2005. Interchromosomal associations between alternatively expressed loci. *Nature*. 435:637–645. doi:10.1038/nature03574.
- van Steensel, B., and A.S. Belmont. 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*. 169:780–791. doi:10.1016/j.cell.2017.04.022.
- Wang, Z., A. Goldstein, R.T. Zong, D. Lin, E.J. Neufeld, R.H. Scheuermann, and P.W. Tucker. 1999. Cux/CDP homeoprotein is a component of NF- $\mu$ NR and represses the immunoglobulin heavy chain intronic enhancer by antagonizing the *brIgHt* transcription activator. *Mol. Cell. Biol.* 19:284–295.
- Weill, J.-C., and C.-A. Reynaud. 2008. DNA polymerases in adaptive immunity. *Nat. Rev. Immunol.* 8:302–312. doi:10.1038/nri2281.
- Williams, A., C.G. Spilianakis, and R.A. Flavell. 2010. Interchromosomal association and gene regulation in trans. *Trends Genet.* 26:188–197. doi:10.1016/j.tig.2010.01.007.
- Yi, M., P. Wu, K.W. Trevor, L. Claflin, and W.T. Garrard. 1999. Evidence that the *Igkappa* gene MAR regulates the probability of premature V-J joining and somatic hypermutation. *J. Immunol.* 162:6029–6039.
- Yokota, T., and Y. Kanakura. 2014. Role of tissue-specific AT-rich DNA sequence-binding proteins in lymphocyte differentiation. *Int. J. Hematol.* 100:238–245. doi:10.1007/s12185-014-1602-2.



**Figure 1.  $MARs_{E\mu}$  deletion displayed a *trans*-dominant effect on SHM**

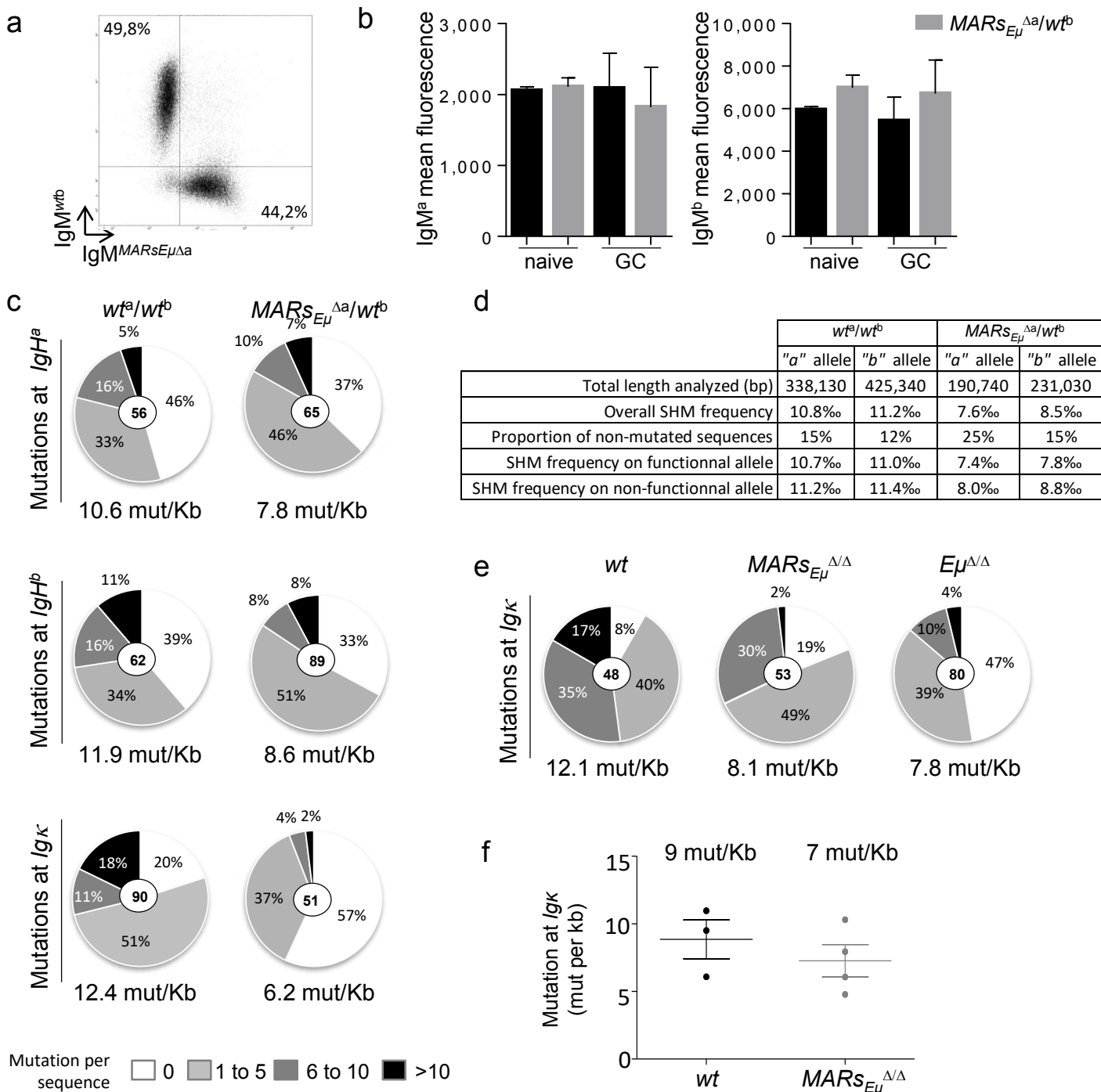
(a) Dot plot displaying percentages of IgM<sup>a</sup> and IgM<sup>b</sup>-expressing B cells (gated B220<sup>+</sup>) in Peyer's patches of  $MARs_{E\mu}^{\Delta a}/wt^b$  heterozygous mice. (b) Intracellular mean fluorescence intensities of IgM<sup>a</sup> and IgM<sup>b</sup> allotypes measured by flow cytometry in Peyer's patch naive and GC B cells from  $wt^a/wt^b$  and  $MARs_{E\mu}^{\Delta a}/wt^b$  mice. At least 5 mice per group in 2 independent experiments, mean±s.e.m. (c) Somatic hypermutation at *IgH* locus, quantified after cloning by classical Sanger method, downstream from *J<sub>H</sub>4* rearranged introns of spleen GC B cells sorted from  $wt^a/wt^b$  and  $MARs_{E\mu}^{\Delta a}/wt^b$  heterozygous mice. For each genotype, pie charts showed the percentage of mutated clones (proportional to the area in each slice) on *IgH<sup>a</sup>* alleles (top panel), *IgH<sup>b</sup>* alleles (middle panel) and both *Igκ* alleles (bottom panel). In each case, total number of clones was noted in the center and SHM frequency was indicated below pie chart. Data were obtained from at least 3 mice per group in at least 3 independent experiments. (d) Somatic hypermutation at both *IgH* alleles of  $wt^a/wt^b$  and  $MARs_{E\mu}^{\Delta a}/wt^b$  heterozygous mice. Mutations were quantified by next generation sequencing, downstream from *J<sub>H</sub>4* rearranged introns of sorted Peyer's patches GC B cells. SHM frequency was specified for both the functional and non-functional alleles. (e) Somatic hypermutation at *Igκ* locus, quantified downstream from *J<sub>k</sub>5* rearranged introns of Peyer's patches GC B cells sorted from *wt* (left panel),  $MARs_{E\mu}^{\Delta\Delta}$  (middle panel) and  $E\mu^{\Delta\Delta}$  (right panel) mice. Pie charts were annotated as in (c). Data were obtained from at least 3 mice per group and at least 3 independent experiments. (f) Left histograms displayed primary transcription of *Igκ* locus (downstream from *J<sub>k</sub>5* exon) evaluated by quantitative PCR (normalized to *Gapdh* transcripts) on Peyer's patch naive and GC B cells from *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  mice (*n*=10 mice in at least 3 independent experiments, mean±SEM). Right histograms displayed intracellular Igκ mean fluorescence intensities, measured by flow cytometry, in naive and GC B cells from Peyer's patches of *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  mice (*n*=8 mice in 2 independent experiments, mean±SEM).





# Figure 1

*wt<sup>a</sup>/wt<sup>b</sup>*  
 *MARs<sub>Eμ</sub><sup>Δa</sup>/wt<sup>b</sup>*



Mutation per sequence

- 0
- 1 to 5
- 6 to 10
- >10

NGS

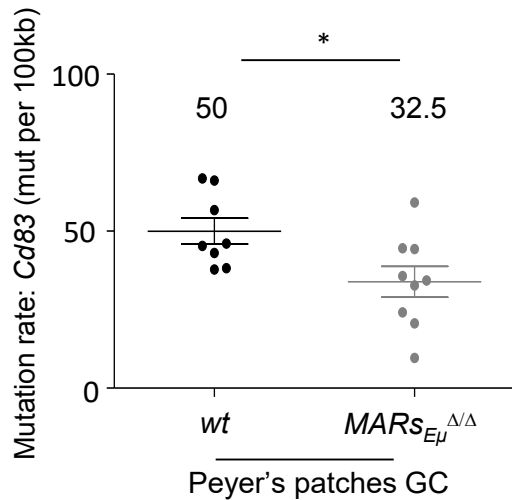
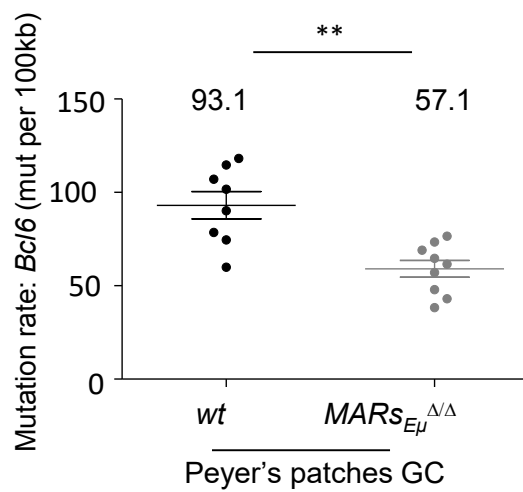


**Figure 2. *IgH*  $MARs_{E\mu}$  deletion decreased SHM on both AID off target genes *Bcl6* and *Cd83***

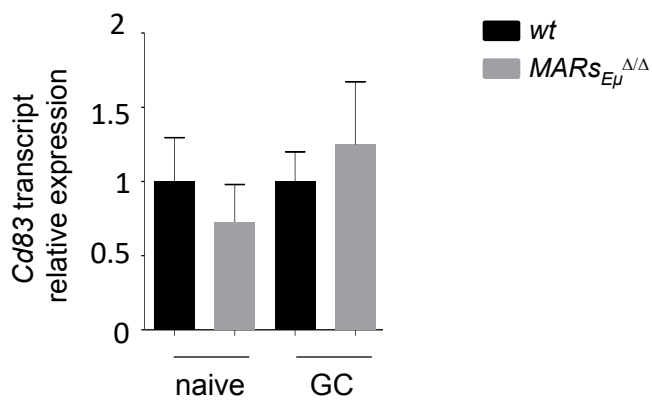
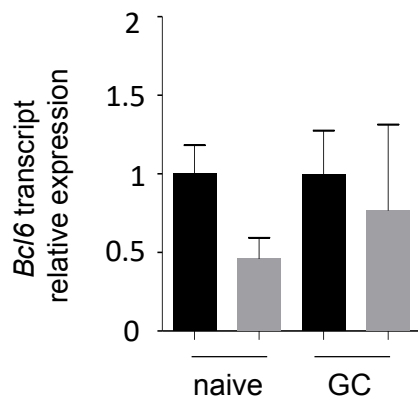
(a) Somatic hypermutation quantified by NGS on *Bcl6* (left graph) and *Cd83* (right graph) genes in Peyer's patch GC B cells sorted from wt and  $MARs_{E\mu}^{\Delta/\Delta}$  mice (6 months old). NGS data were represented as a graph showing the mutation rate (mean $\pm$ SEM). NGS experiments were obtained from at least 8 wt and  $MARs_{E\mu}^{\Delta/\Delta}$  mice. Each point on the graph refers to a pool of 2 mice. Statistical results were obtained using a Mann-Whitney test (two-tailed). (b) Histograms displayed *Bcl6* (left) and *Cd83* (right) primary transcription evaluated by quantitative PCR (normalized to Gapdh transcripts) on Peyer's patch naive and GC B cells from wt and  $MARs_{E\mu}^{\Delta/\Delta}$  mice (n=6 mice in 2 independent experiments, mean $\pm$ SEM).



a



b





**Figure 3. Ig genes loci are preferentially located in away from the periphery in the absence of the  $MARs_{E\mu}$**

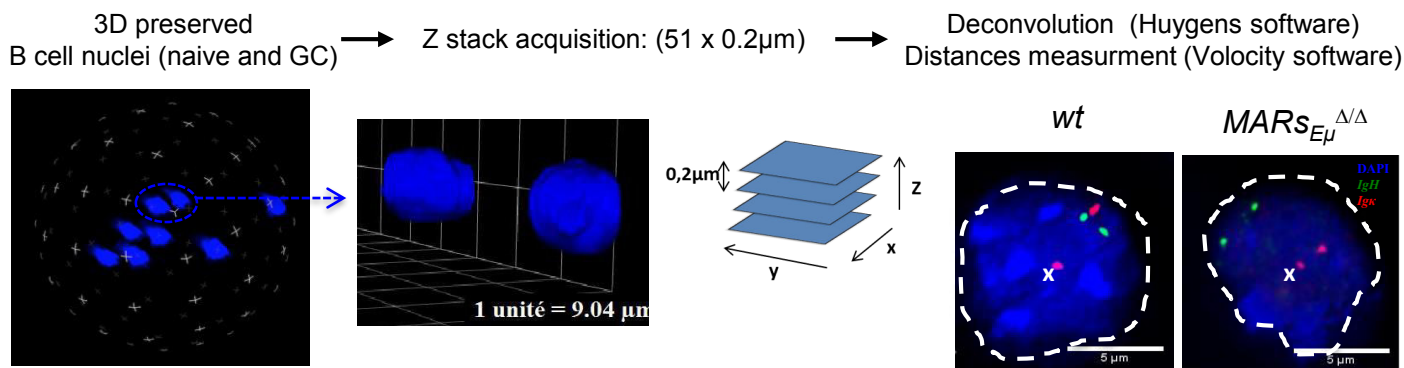
3D-FISH data were obtained from at least 3 independent experiments with at least 3 mice per group. (a) 3D-FISH experiment schematic overview and *IgH* and *IgK* loci representation (not to scale) detailing genomic regions covered by BACs used as FISH probes. Probes used for *Bcl6* and *Cd83* detection covered the entire gene (not represented). (b) Scatter plots displayed distances between nuclear periphery and *IgH* (left) or *IgK* (right) normalized to nucleus diameter, measured in individual nucleus by 3D-FISH in naive and GC B cells from Peyer's patches sorted from wt and  $MARs_{E\mu}^{\Delta\Delta}$  mice. Each point represents one distance measured in one nucleus (mean $\pm$ SEM). (c) As in b, scatter plots displayed distances between nuclear periphery and *Bcl6* (left) or *Cd83* (right).

Data are obtained from at least 60 (b) or 20 (c) nuclei. P-values were calculated using a Kolmogorov-Smirnov test.

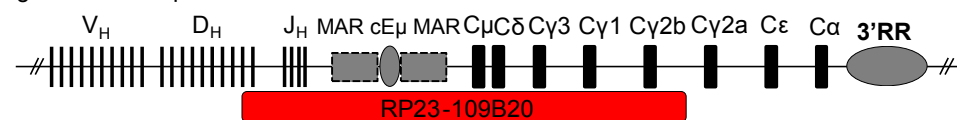




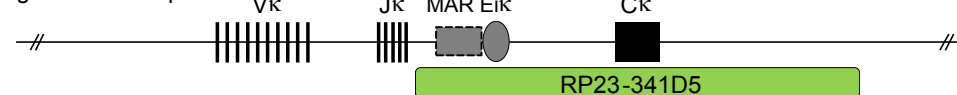
**a** 3D-FISH overview



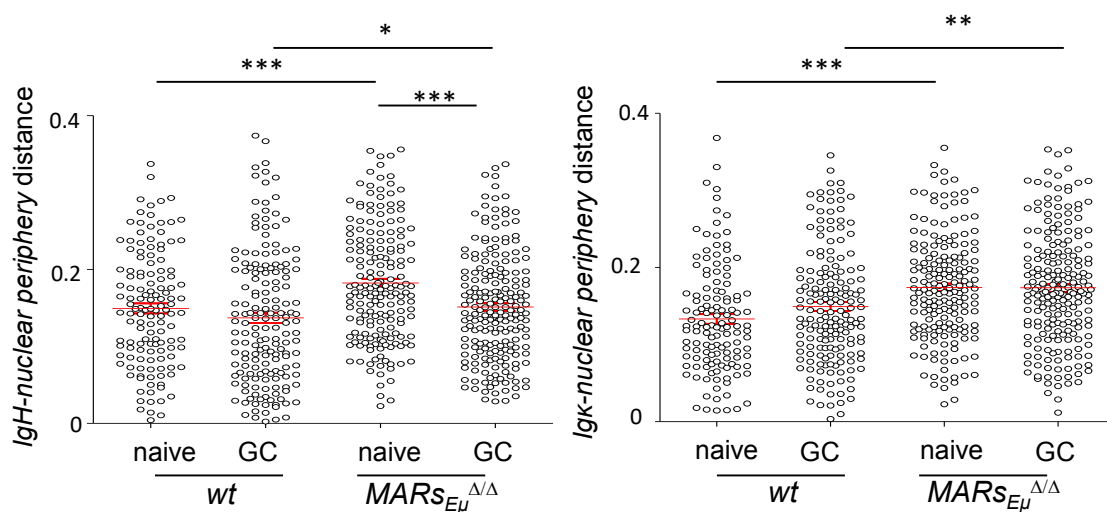
IgH Locus and probe



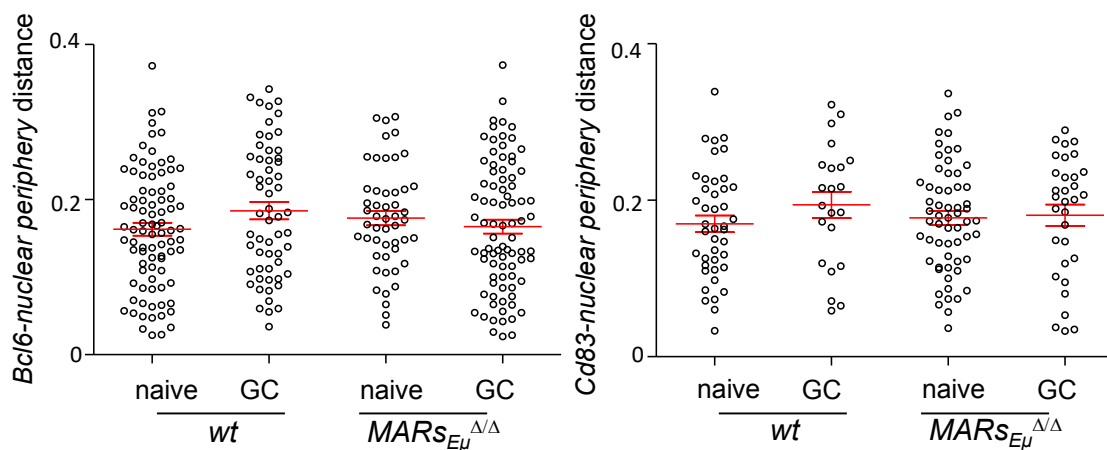
Igκ Locus and probe



**b**



**c**



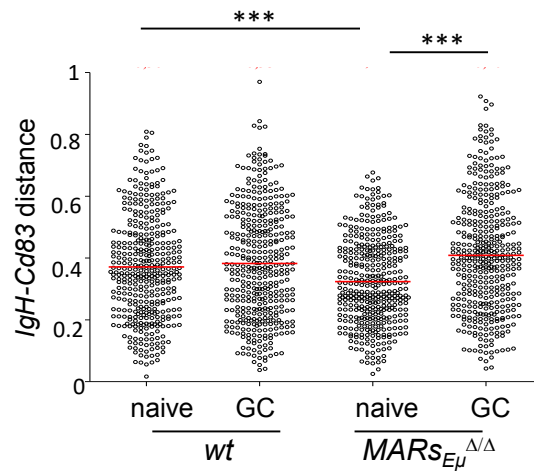
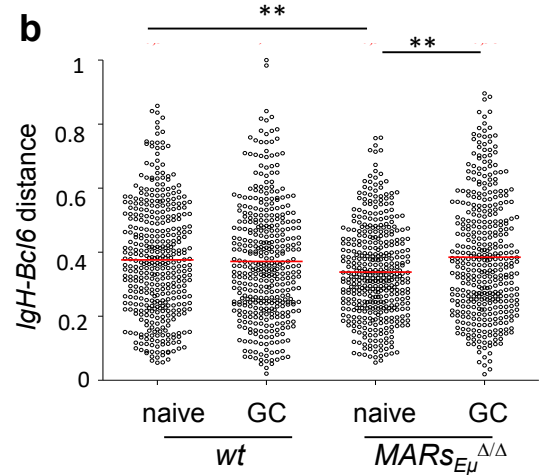
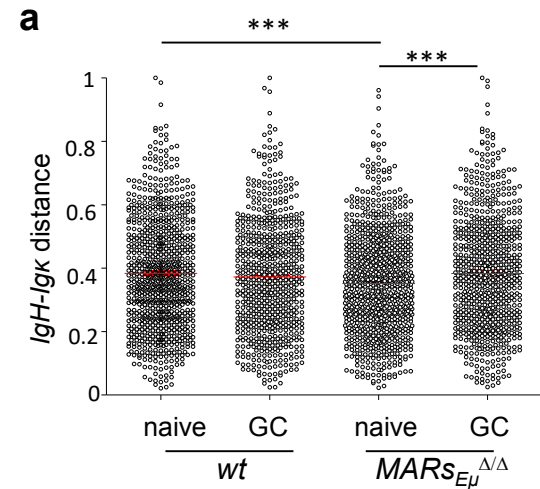


**Figure 4. *IgH*  $MAR_{E\mu}$  deletion disrupted AID targets inter genic distances**

3D-FISH data were obtained from at least 3 independent experiments with at least 3 mice per group. **(a and b)** Scatter plots displayed *IgH* – *Ig $\kappa$*  (a), *IgH* – Bcl6 (b left) or *IgH* – Cd83 (b right) inter genic distances normalized to nucleus diameter, measured in individual nucleus by 3D-FISH in naive and GC B cells from Peyer's patches sorted from wt and  $MAR_{E\mu}^{\Delta\Delta}$  mice. Each point represents one distance measured in one nucleus (mean $\pm$ SEM).

Data are obtained from at least 100 nuclei. P-values were calculated using a Kolmogorov-Smirnov test.





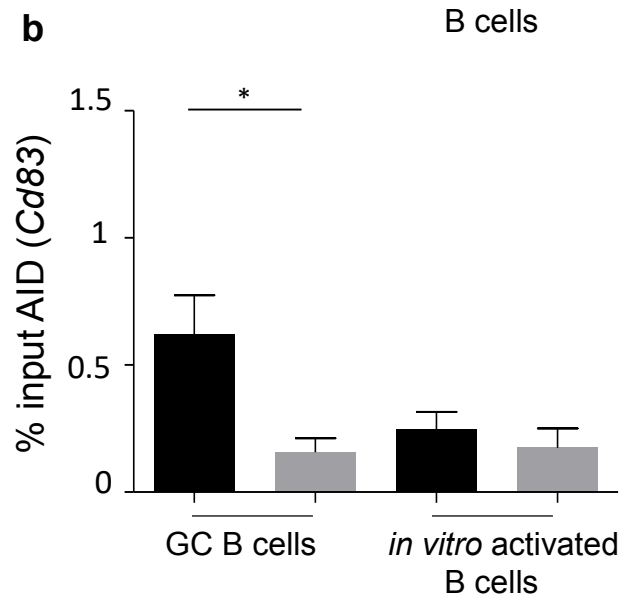
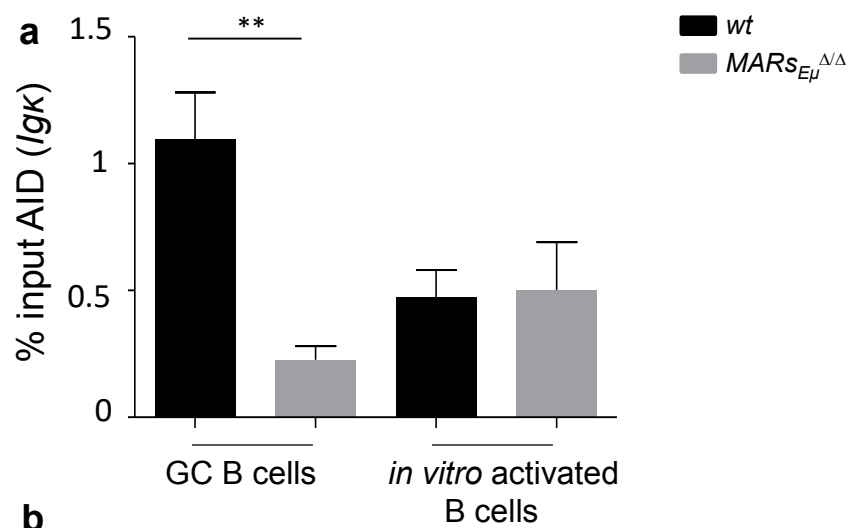


**Figure 5.  $MARs_{E\mu}$  deletion impaired AID recruitment**

(a and b) AID recruitment at *Ig $\kappa$*  locus (a) or *Cd83* genes (b) accessed by ChIP experiment using anti-AID antibody performed on Peyer's patch cells (n=5 samples per genotype) and on *in vitro* activated splenic B cells with LPS + IL4 (n=4 and 5 samples for *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  genotypes, respectively). Statistical results were obtained using a Mann-Whitney test (two-tailed).









## *Discussion et perspectives*



## Discussion et perspectives.

---

Grâce à l'élaboration de modèles murins, notre laboratoire s'est efforcé à déterminer le rôle des régions régulatrices au locus *IgH*, notamment lors du processus de SHM. Mon travail de thèse a contribué à l'étude du modèle de souris MARs<sub>E $\mu$</sub>  KO correspondant à la délétion des 2 régions d'attachement à la matrice nucléaire encadrant l'activateur intronique cE $\mu$ . Les études que nous avons menées ont pour objectif une meilleure compréhension de la régulation du processus de SHM. Les résultats inattendus observés, en faveur d'une régulation « en *trans* » médiée par les régions MARs<sub>E $\mu$</sub>  du locus *IgH*, soulignent la complexité des mécanismes régulateurs de la SHM, qui s'étend au-delà des hypermutations des gènes d'Ig en causant des « dommages collatéraux » à certains oncogènes du lignage B.

La région E $\mu$  au locus *IgH* est constituée de l'activateur intronique core E $\mu$  et de ces 2 régions MARs<sub>E $\mu$</sub>  flanquantes ; sa localisation entre les segments J<sub>H</sub> et C $\mu$  fait que cette région est préservée lors des événements de remaniement. De nombreuses études se sont intéressées à la région E $\mu$  et ont confirmé son rôle essentiel lors du développement B précoce, attribué à l'activité du seul core E $\mu$ . En effet, il a été montré que le core activateur E $\mu$  régule l'étape de recombinaison D vers J<sub>H</sub> en modulant la transcription germinale conduisant l'accessibilité des régions à recombiner (Afshar et al., 2006; Bolland et al., 2004b, 2007b; Chakraborty et al., 2009b; Chen et al., 1993; Perlot et al., 2005; Sakai et al., 1999a; Serwe and Sablitzky, 1993). La délétion combinée du core E $\mu$  et des MARs<sub>E $\mu$</sub>  (modèle E $\mu$  FL KO) n'aggrave pas le phénotype (Marquet et al., 2014). Par ailleurs, la délétion combinée (du core E $\mu$  et des MARs<sub>E $\mu$</sub> ) dans ce modèle E $\mu$  FL KO, entraîne un défaut de la répartition des populations B en périphérie en faveur d'une différenciation des cellules B de la zone marginale (B MZ). Ce défaut est à mettre en relation avec la diminution de l'expression de la chaîne lourde d'Ig aux stades précoces qui conditionne ensuite le destin des cellules B en périphérie. Aucune défaillance de la recombinaison VDJ (Sakai et al., 1999a) ou de l'expression de la chaîne lourde n'ayant été constaté en l'absence des régions MARs<sub>E $\mu$</sub>  (article 2), la régulation de ces événements peut logiquement être attribuée à la seule présence du core E $\mu$ , laissant suggérer un rôle négligeable pour les régions MARs<sub>E $\mu$</sub> .

Concernant les événements de remaniement se déroulant en périphérie dans les cellules B du CG, les différents modèles d'étude de la région E $\mu$  (présentant une délétion du cE $\mu$  seul ou en combinaison avec les MARs<sub>E $\mu$</sub> ) ont successivement validé et invalidé son implication lors de la CSR et de la SHM (Bottaro et al., 1998b; Li et al., 2010a; Perlot et al., 2005; Ronai et al., 2005b). La délétion du core E $\mu$  dans un modèle de souris KO (modèle le plus approprié car incluant la délétion propre du cE $\mu$  dans la lignée germinale) n'affecte pas les processus de CSR et de SHM (Perlot et al., 2005). Bien que la proportion des séquences mutées soit diminuée dans ce modèle (en lien avec la diminution des cellules B en périphérie), la suppression du core E $\mu$  n'affecte pas la fréquence de mutation (au locus *IgH*, dans des cellules B du CG triées de Plaques de Peyer) qui est similaire à celle obtenue dans un contexte *wt*. Une étude non exhaustive (Marquet & Pinaud, données non



publiées) de la SHM dans le modèle  $E_{\mu}$  FL KO indique cependant que l'absence du core  $E_{\mu}$  et des  $MAR_{E_{\mu}}$  réduit la fréquence des SHM dans la population cellulaire correspondante. Ces derniers résultats ont soulevé de nouvelles interrogations sur un rôle potentiel pour les régions  $MAR_{E_{\mu}}$ .

La création du modèle  $MAR_{E_{\mu}}$  KO nous a permis de confirmer l'implication des  $MAR_{E_{\mu}}$  dans le processus de SHM : ces régions participent à la régulation des mutations somatiques non seulement des cibles légitimes (les gènes d'Ig) mais également des cibles illégitimes de AID telles que *Bcl6* et *Cd83*.

Notre étude a débuté par l'analyse des mutations au locus *IgH* (en *cis*) dans un contexte *wt* et déficient pour les  $MAR_{E_{\mu}}$  (article 2). La délétion des  $MAR_{E_{\mu}}$  affecte le processus de SHM : la fréquence de mutation en amont de la région  $E_{\mu}$  (région intronique ciblée par la SHM mais non sujette à une sélection par l'antigène) est diminuée. Cette observation corrèle avec une étude antérieure rapportant que la fréquence de mutation au locus *Igk* se trouve diminuée en l'absence de la MAR située en amont de l'activateur intronique  $E_{i\kappa}$  (Yi et al., 1999b). Cependant aucun mécanisme moléculaire associé à cette diminution de SHM n'avait été proposé, les auteurs n'ayant constaté aucun défaut de transcription dans ce modèle.

Au locus *IgH*, notre étude conduit à plusieurs hypothèses quant au mode d'action des régions  $MAR_{E_{\mu}}$ .

(1) Les  $MAR_{E_{\mu}}$  semblent agir comme une barrière permettant d'isoler l'unité transcriptionnelle incluant le segment VDJ recombiné comme un élément « à cibler » par AID. Il est vraisemblable que la barrière soit établie par le biais de l'ancrage à la matrice nucléaire. Le fait que la fréquence de mutation soit augmentée en aval de la région  $E_{\mu}$ , en absence des régions  $MAR_{E_{\mu}}$ , est un argument en faveur du rôle de barrière de ces régions. Nos données montrant un ciblage accentué des régions en aval de  $E_{\mu}$  sont en accord avec une étude précédente montrant une augmentation des délétions internes dans la région  $S_{\mu}$  en absence des  $MAR_{E_{\mu}}$  (Sakai et al., 1999b). Cependant, il est probable que le phénomène d'isolement de l'unité VDJ associé à l'étape de SHM soit dynamique et transitoire car il a été montré, dans un contexte *wt*, que la région en aval de  $E_{\mu}$  est également ciblée par AID (Xue et al., 2006). Par ailleurs, il est possible que ce rôle de barrière par l'ancrage à la matrice favorise : i/ le ralentissement et l'empilement de complexes de transcription (ARN pol II en pause) (Maul et al., 2014) et ii/ l'apparition de contraintes de l'ADN autour de la machinerie de transcription (un surenroulement négatif en aval de l'ARN pol II), ce type de contrainte pouvant conduire à l'exposition de l'ADN sous forme simple brin (substrat de l'enzyme AID) (Parsa et al., 2012; Ronai et al., 2007; Wang et al., 2014).

(2) Notre étude suggère également que la transcription antisens dirigée vers l'amont de la région  $E_{\mu}$  (région ciblée par AID lors de la SHM) soit médiée par les  $MAR_{E_{\mu}}$ . Au cours du développement B post médullaire, des transcrits antisens ont été décrits au niveau des régions switch (avant la CSR) (Haddad et al., 2011; Pefanis et al., 2015; Perlot et al., 2008) et également dans la région intronique en amont de  $E_{\mu}$  (dans les cellules B du CG) (Perlot et al., 2008). La présence de ces transcrits antisens semble être d'une importance particulière,





puisque une étude révèle que les régions de transcription convergente (sens et antisens) corrélient avec le recrutement de AID (Meng et al., 2014). Une hypothèse séduisante serait que la SHM soit régulée par un équilibre entre la transcription sens et antisens. Notre étude suggère que l'apparition de transcrits antisens dépend de la présence des régions MAR<sub>εμ</sub>. Ces dernières constituent de véritables plateformes de recrutement pour différentes protéines (MAR-Binding proteins ou MAR-BP) en lien avec la structure du noyau (notamment la structure en réseau associée à SATB1 (Cai et al., 2003)). Il est possible que certaines MAR-BP soient capables de promouvoir la transcription antisens (Forrester et al., 1994; Jenuwein et al., 1997b; Ratliff et al., 2014b) et nous envisageons de rechercher plus précisément cette fonction potentielle pour les MAR-BP. Les transcrits antisens sont connus pour être faiblement produits et particulièrement instables (Zhao et al., 2009) ; par conséquent la diminution de ces transcrits dans un contexte déficient pour les MAR<sub>εμ</sub> reste difficile à mettre en évidence. Dans notre étude, nous avons donc multiplié nos expériences de RT-qPCR spécifiques de brin par la réalisation de plusieurs ADN complémentaires (ADNc) sens et antisens couplés à différentes sondes Taqman réparties dans les régions introniques d'intérêt. Pour appuyer ces observations, l'utilisation d'autres techniques plus « robustes » (comme le séquençage haut débit d'ARN, RNA-seq) est envisageable mais la fugacité des transcrits antisens (pris en charge par le RNA exosome) (Flynn et al., 2011; Pefanis et al., 2014) pose la question de la technique haut débit à utiliser. Une alternative pour améliorer la détection de ces transcrits serait de réaliser l'étude dans un contexte déficient pour le RNA exosome dans lequel la stabilité des transcrits non codants (lncRNA, eRNA, etc..) est prolongée (Li et al., 2016b; Pefanis et al., 2014).

(3) Notre étude indique que la délétion des MAR<sub>εμ</sub> interfère avec la progression du complexe de l'ARN pol II. Bien que le complexe de l'ARN pol II « en pause » ait été décrit comme favorisant la SHM (Maul et al., 2014; Wang et al., 2014), dans notre modèle MAR<sub>εμ</sub> KO, la présence du complexe en pause évaluée par CHIP n'est pas significativement diminuée. Ces résultats peuvent s'expliquer par le fait que le processus de SHM n'est que partiellement affecté en l'absence des MAR<sub>εμ</sub>. En revanche, le recrutement du complexe de l'ARN pol II « en élongation » est quant à lui significativement diminué dans les cellules B du CG issues d'animaux MAR<sub>εμ</sub> KO. Ces résultats surprenants pourraient résulter d'une diminution globale de l'activité du cεμ ; les enhancers sont en effet décrits pour recruter cette forme de l'ARN pol II de part et d'autre de leur core (Zentner et al., 2011).

(4) Notre étude par CHIP montre que les MAR<sub>εμ</sub> participent au recrutement de AID. Ce défaut de recrutement de AID en absence des MAR<sub>εμ</sub> est observé dans la région génomique en amont du cεμ où sont initiés les transcrits antisens. Ces observations, en accord avec l'hypothèse de Alt (Meng et al., 2014), semblent indiquer une corrélation entre transcription convergente et recrutement de AID au locus *IgH*. A ces observations s'ajoute le constat que la diminution du recrutement de AID et de la transcription antisens dans le modèle



MAR<sub>SEμ</sub> KO est uniquement observée dans les cellules B du CG (subissant la SHM) et non pas dans les cellules B stimulées *in vitro* (ne subissant pas la SHM).

L'abolition quasi-totale de la SHM dans les cellules B de CG issues de souris 3'RR KO pouvait laisser suggérer que la région régulatrice en 3' contrôlait à elle seule la régulation de la SHM au locus *IgH* (Rouaud et al., 2013). Notre étude met en évidence que les MAR<sub>SEμ</sub> sont des régions *cis*-régulatrices « partenaires » de la 3'RR permettant d'optimiser l'efficacité du ciblage de la SHM en modulant la transcription convergente. Il est par ailleurs envisageable que la transcription sens soit absolument indispensable (Fukita et al., 1998) et que la transcription antisens ne soit qu'un mécanisme additionnel comme le suggère la baisse relative de SHM (sans abrogation) dans les modèles dépourvus de MARs (Li et al., 2010b et article 2).

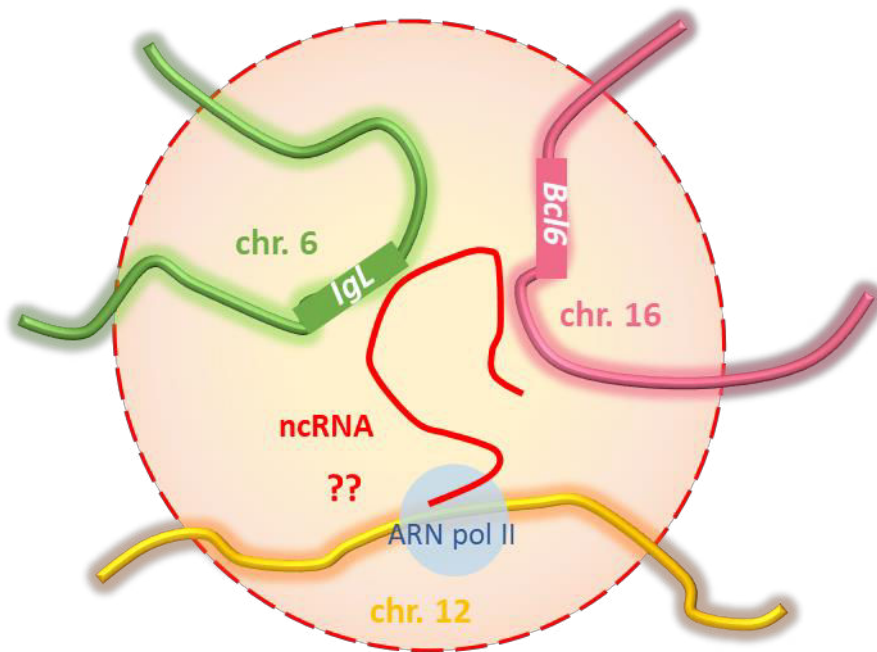
Le processus de SHM introduit également des mutations dans les régions variables des chaînes légères d'Ig. De façon similaire aux études précédentes, menées dans les modèles de délétions dans la région 3'RR (Garot et al., 2016; Rouaud et al., 2013), notre étude s'est étendue à l'analyse des mutations au locus *Igκ* des chaînes légères d'Ig. De façon inattendue, la fréquence de mutation au locus *Igκ* est diminuée en l'absence des MAR<sub>SEμ</sub> suggérant une action *trans*-régulatrice de ces régions sur le processus de SHM (article 3). Les modèles d'animaux hétérozygotes porteurs d'allèles *IgH* d'haplotypes différents (*IgH* « a » dans la souche 129/sv et *IgH* « b » dans la souche C57Bl/6) ont par le passé permis des comparaisons fiables et reproductibles pour nos modèles de délétion d'éléments régulateurs *IgH* (Garot et al., 2016; Perlot et al., 2005; Sakai et al., 1999a). Dans le cas du modèle MAR<sub>SEμ</sub> KO, les modèles hétérozygotes MAR<sub>SEμ</sub><sup>Δa/wtb</sup> ont confirmé l'effet *trans*-régulateur lié à la suppression des régions MAR<sub>SEμ</sub>, que nous avons qualifié de *trans*-dominant. La suppression de la région régulatrice sur un seul allèle *IgH* diminue la SHM sur le second allèle *IgH* (ce dernier étant cependant pourvu de MAR<sub>SEμ</sub> intactes) et simultanément diminue la SHM aux loci des chaînes légères *Igκ*. Chez ces modèles, les proportions respectées des allèles productifs et non productifs confirment un réel défaut de SHM en invalidant tout biais de sélection.

Après avoir révélé la fonction *trans*-régulatrice des régions MAR<sub>SEμ</sub> sur la SHM, il était légitime de poser la question de leur capacité potentielle à influencer l'introduction des mutations aux niveaux des cibles illégitimes (appelées gènes « off target »). Nous avons alors déterminé la fréquence de mutation de 2 cibles illégitimes de AID : *Bcl6* et *Cd83* dont la fréquence de mutation peut être étudiée dans un contexte *wt* (ces 2 gènes sont physiologiquement réparés de façon peu fidèle) (Álvarez-Prado et al., 2018; Liu et al., 2008 et article 1). De façon similaire aux résultats obtenus pour les gènes d'Ig, nous observons une diminution de la fréquence de mutation pour les 2 cibles « off targets » *Bcl6* et *Cd83* en l'absence des régions MAR<sub>SEμ</sub> (article 3). Ces résultats confirment une fois de plus l'action en *trans* des régions MAR<sub>SEμ</sub> lors de la SHM : de telles régions semblent être impliquées dans le ciblage de la machinerie de mutation à distance au niveau de gènes localisés sur d'autres chromosomes que ceux portant les loci des gènes d'Ig (ici les chromosomes 16 pour *Bcl6*



et 13 pour *Cd83*). L'apparition de mutations dans le gène *Bcl6* a précédemment été impliquée dans la survenue de lymphomes (Jardin and Sahota, 2005; Pasqualucci et al., 2001). L'étude des éléments *cis*- et *trans*-régulateurs impliqués dans l'apparition de ces mutations présente un intérêt certain pour une meilleure compréhension de l'apparition des cancers des lymphocytes B. Notre outil (DeMinEr) dédié à la détection des mutations rares pourrait être utilisé de façon plus large pour des applications liées à l'identification de mutations à partir d'échantillons tumoraux. Par ailleurs, la sensibilité de détection de DeMinEr en fait un outil idéal pour étendre notre recherche de mutations à d'autres cibles illégitimes de AID.

Nos résultats nous ont conduit à l'hypothèse d'une régulation « spatiale » du processus de SHM médiée par les MARs<sub>EM</sub>, dans les noyaux des cellules B du CG. Si des changements dynamiques de position des gènes d'Ig ont été largement décrits au cours du développement B précoce, peu d'études rapportent leur position au sein de cellules B effectuant la SHM. Notre étude s'est alors poursuivie par l'analyse du positionnement des cibles de AID (légitimes et illégitimes) dans les noyaux des cellules B naïves et de CG issues d'animaux *wt* et MARs<sub>EM</sub> KO. Les résultats présentés dans l'article 3 semblent en faveur d'une perte d'organisation nucléaire en l'absence des régions MARs<sub>EM</sub>. Sous réserve que certains de nos résultats soient confirmés par une augmentation de l'échantillonnage, il semblerait que le déroulement normal de la SHM nécessite un positionnement des cibles de AID en périphérie du noyau. Bien qu'une étude propose qu'il n'existe pas, pour un gène cible illégitime, de corrélation entre sa fréquence de mutation et sa proximité avec le locus *IgH* (Gramlich et al., 2012) (pour exemple *Bcl6* la cible illégitime la plus mutée n'est pas la plus proche de *IgH*) ; il n'est pas exclu que le phénomène de SHM nécessite une relocalisation des cibles légitimes (et peut être illégitimes) de AID dans une région du noyau des cellules B favorable aux mutations. Nous proposons que cette relocalisation soit médiée par les MARs<sub>EM</sub>. Par ailleurs notre hypothèse de repositionnement corrèle avec une étude récente indiquant une relocalisation des gènes d'Ig dans un compartiment nucléaire dédié (à proximité du nucléole) lors du développement tardif des cellules B (d'origine humaine) (Pichugin et al., 2017). Cependant chez la souris, selon la souche, la présence des NOR (Nucleolar Organizer Regions) sur des chromosomes différents peut influencer le positionnement de gène se trouvant à proximité de ces régions NOR (Strongin et al., 2014). Dans la continuité de nos études de positionnement, il est envisageable d'inclure un immunomarquage de certains compartiments nucléaires (nucléoles, nucléopores...) de façon à préciser la zone nucléaire ciblée pour la relocalisation des cibles légitimes et illégitimes de AID. D'autres études s'intéressant à la prise en charge des cassures double brin de l'ADN ont montré que le choix de la voie de réparation varie selon la position. Ainsi les cassures double brin induites au niveau des LAD (Lamin Associated Domain) vont préférentiellement être réparées par la voie de NHEJ (ou AEJ) qui semblent pourtant plus « error prone » que la voie de la recombinaison homologue (Lemaître et al., 2014). Bien que ces voies n'interviennent pas dans la prise en charge des cassures simple brin de l'ADN (à l'origine des SHM), il est envisageable que la périphérie nucléaire soit préférentiellement associée à une réparation de type « error prone » ; une destination



**Figure 24 : Hypothèse d'un mode d'action en *trans* pour les ARN non codants.**

Les ARN non codants faciliteraient les interactions interchromosomales rapprochant des cibles localisées sur différents chromosomes à proximité de leur site de transcription.

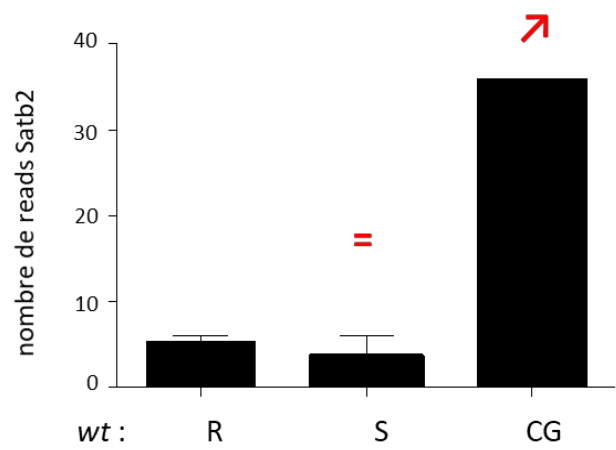
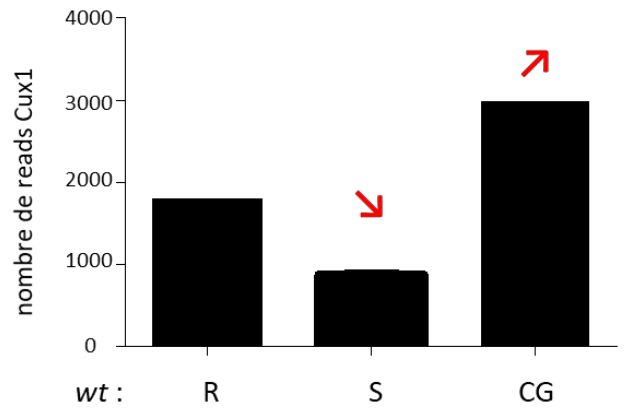
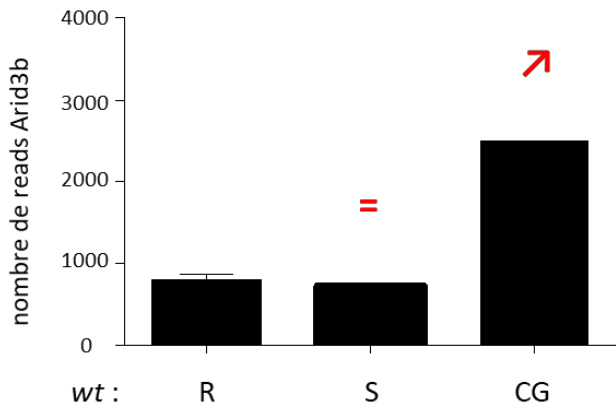
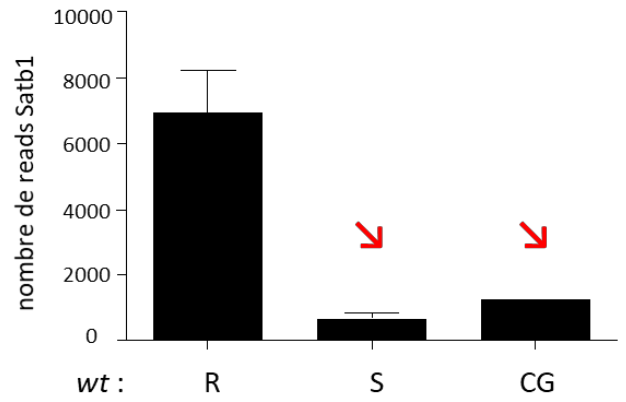
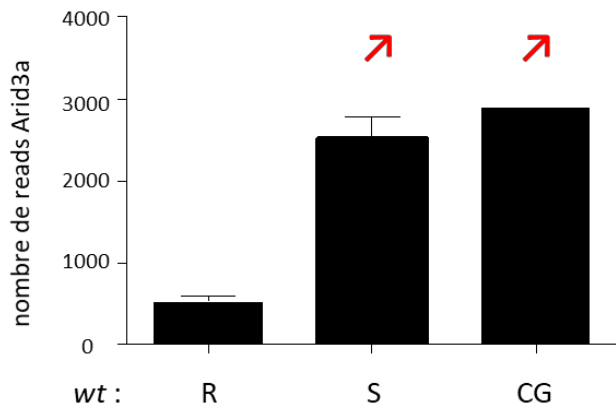
périnucléaire pour des gènes en cours de SHM serait cohérente avec nos données suggérant que les cibles de AID se localisent plus en périphérie du noyau dans un contexte *wt*. Toutefois, notre analyse des distances à la périphérie n'est basée que sur la prise en compte du contour des noyaux marqués au DAPI ; nous envisageons de préciser et confirmer la localisation périnucléaire en utilisant un immunomarquage des lamines. Une étude récente menée dans la lignée cellulaire BL2 montre que l'induction des mutations dans cette lignée est associée à une diminution de l'interaction entre *IgH* et la protéine lamine B (observée par CHIP Seq) (Klymenko et al., 2018). Ces résultats sont en contradiction avec nos données ; une telle différence pourrait s'expliquer par le fait que le modèle utilisé est une lignée cellulaire cultivée *in vitro* (qui n'est pas capable d'hypermuter les gènes d'Ig de façon spontanée) dont les mécanismes régulateurs peuvent différer de ceux de cellules fraîchement isolées d'un animal.

Au-delà d'une régulation « spatiale » de la SHM, notre étude n'exclue pas un potentiel rôle de la transcription antisens dans le recrutement de AID au niveau de cibles autres que *IgH*. Les zones de transcription convergente favorables au recrutement de la machinerie de mutation concernent aussi bien les cibles légitimes que illégitimes de AID (Meng et al., 2014). L'étude de la fonction de la transcription antisens dans le ciblage de AID sur les gènes d'Ig et les cibles illégitimes peut être facilitée dans un contexte déficient pour le RNA exosome qui évite la dégradation rapide des ARN non codants (souvent antisens) (Pefanis et al., 2014). Une étude détaillée du transcriptome des cellules B de CG de nos modèles serait particulièrement utile et pertinente, nous avons tenté de réaliser une telle étude par RNA-seq mais les résultats obtenus ne sont pas assez robustes pour déterminer avec fiabilité la présence de tels transcrits (trop rares) dans les régions d'intérêts. De nouvelles méthodes de séquençage (telle que le NET-seq, détectant les transcrits naissants associés à l'ARN pol II) pourraient présenter un intérêt méthodologique dans nos modèles. Par ailleurs, des hypothèses proposent que parmi les ARN non codants, certains eRNA ou lncRNA exercent un rôle *trans*-régulateur (Hacisuleyman et al., 2014; Mousavi et al., 2013; Quinodoz and Guttman, 2014; Tsai et al., 2010) (**Figure 24**).

Dans le but de déterminer le lien entre les régions MAR<sub>EM</sub> au locus *IgH* et leur fonction régulatrice du processus de SHM, nous souhaitons étendre nos recherches à l'étude des MAR-BP et déterminer si de telles protéines pourraient contribuer à ce mécanisme. Parmi les MAR-BP décrites, nous supposons que celles dont l'expression varie, lors de la transition entre les stades cellules B au repos (ne subissant pas la SHM) et du CG (subissant la SHM), peuvent être des facteurs régulateurs de la SHM. Dans une vision simpliste, une protéine surexprimée dans les cellules B du CG pourrait être activatrice de la SHM, à l'inverse une protéine sous exprimée pourrait être inhibitrice. Nous avons alors fait un 1<sup>er</sup> crible à partir de nos données de RNA-seq obtenues à partir d'ARN extraits :

- de cellules B de rate au repos (condition R, n=3)
- de cellules B de rate stimulées *in vitro* avec du LPS (condition S, n=3)





**Figure 25 : Données de RNA-seq et MAR-BP.**

Les données de RNA-seq présentées ici ont été obtenues à partir d'ARN extrait de cellules B *wt* dans différentes conditions :

- R : cellules B de rate au repos
- S : cellules B de rate stimulées *in vitro* (LPS pendant 2 jours)
- CG : cellules B activées de CG de plaques de Peyers

- de cellules B de CG de plaques de Peyer (condition CG, n=1).

Il faut noter que nous avons également effectué des expériences de RNA-seq dans les mêmes conditions (R, S et CG) dans un contexte déficient pour les MARs<sub>Eμ</sub>. De façon attendue, le profil d'expression des gènes codant les MAR-BP est similaire entre les animaux *wt* et MARs<sub>Eμ</sub> KO.

Nos résultats montrent que, les transcrits correspondant aux protéines ARID3A, ARID3B et CUX1 sont augmentés dans la condition CG par rapport à la condition R. Cependant, il est intéressant de noter que les transcrits du gène *Arid3a* sont aussi bien augmentés dans la condition CG que dans la condition S. La situation est différente pour les transcrits du gène *Arid3b*, qui se trouvent augmentés uniquement dans la condition CG. Quant au gène codant CUX1, alors que la quantité de transcrite est augmentée dans la condition CG, elle est diminuée dans la condition S. Nous nous sommes également intéressés aux 2 MAR-BP de la famille « special AT-rich » SATB1 et SATB2. Alors que le gène codant SATB2 semble très peu exprimé quelle que soit la condition (R, S ou CG), les transcrits du gène codant la protéine SATB1 se trouvent diminués dans les cellules B stimulées *in vitro* ou B du CG (**Figure 25**).

Les cellules B stimulées *in vitro* n'ont pas la capacité de muter leur gènes d'Ig dans les conditions standards de stimulation (en présence de LPS) (Maul et al., 2014). Pour ces cellules (condition S), le profil des transcrits correspondant aux protéines ARID3B et CUX1 diffère de celui associé aux cellules B du CG (qui hypermutent leur gènes d'Ig) : ces résultats suggèrent une potentielle implication de ces protéines lors du processus de SHM. Cependant, ces résultats doivent être confirmés car obtenus à partir d'un seul échantillon *wt* en condition CG (une autre puce de RNA-seq est en cours) et ne nous permettent pas d'exclure un rôle pour ARID3A ou SATB1 dans le processus de SHM. Une étude récente montre que la délétion d'une portion de la région située en aval du dernier segment J<sub>H</sub> (incluant une partie de la MAR<sub>Eμ</sub> 5') n'affecte en rien la SHM (Castiblanco et al., 2018). Dans ce modèle, il est important de noter que les sites de fixation pour les MAR-BP ont été conservés, il n'est donc pas impossible que ces protéines interviennent dans la régulation de la SHM. En revanche, la protéine ARID3C ne semble pas être une candidate potentielle intervenant dans la régulation de la SHM puisqu'aucun transcrite correspondant à cette protéine n'a pu être détecté dans nos 3 conditions (R, S et CG).

Par ailleurs, nos données de quantification des transcrits des MAR-BP doivent être complétées par des vérifications de l'expression de chacune de ces protéines. Ces vérifications sont indispensables sachant que d'une part des données de la littérature décrivent l'expression de CUX1 uniquement aux stades précoces de développement B (Wang et al., 1999b) et que d'autre part des études récentes montrent que le profil d'expression de la protéine ARID3C suit celui de ARID3A (Tidwell et al., 2011). Les profils d'expression des MAR-BP dans les différentes sous populations de cellules B seront à concilier à l'aide des nouvelles techniques haut débit et des nouveaux anticorps disponibles. La difficulté à élucider ces profils dans un contexte

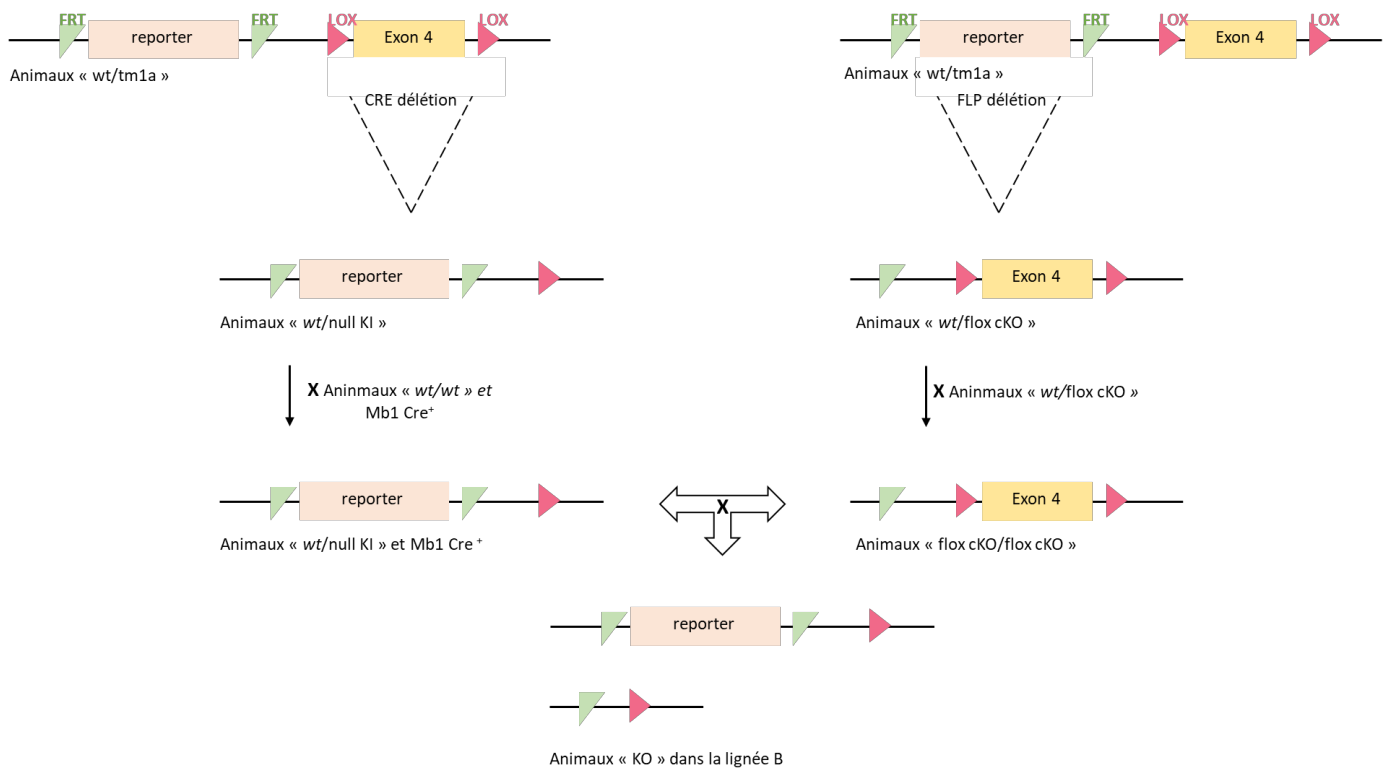


Figure 26 : Stratégie de croisement pour l'obtention des modèles de souris KO conditionnels pour ARID3A et SATB1.

physiologique (à l'échelle de l'organisme entier) réside dans le fait que le KO de la plupart de ces protéines soit létal (Alvarez et al., 2000; Dobrev et al., 2006; Kurkewich et al., 2016; Sinclair et al., 2001; Webb et al., 2011). Cette létalité pourrait probablement également expliquer pourquoi le potentiel rôle des MAR-BP dans la régulation de la SHM n'ait jusqu'alors pas été exploré.

Pour contourner la létalité associée au KO, des modèles de souris KO conditionnels peuvent être utilisés. Il est possible d'établir des modèles de souris cKO pour ARID3A et SATB1 à partir d'animaux hétérozygotes « *wt/tm1a* » disponibles via le « KOMP KO mouse project ». Nous avons donc débuté des croisements à partir des animaux hétérozygotes « *wt/tm1a* » selon la stratégie décrite dans la figure 26. A l'issue de ces croisements, nous obtiendrons des animaux cKO pour lesquels la délétion des protéines ARID3A et SATB1 est restreinte aux cellules du lignage B. Dans un 1<sup>er</sup> temps, l'analyse des mutations au locus *IgH* à partir de cellules B de CG triées issues des animaux cKO permettra de déterminer si ces 2 MAR-BP (ARID3A et SATB1) sont impliquées dans le processus de SHM. En fonction des résultats obtenus, notre étude pourra s'étendre à l'analyse de la transcription et du recrutement de AID voir à la quantification des mutations au locus des chaînes légères et aux cibles illégitimes de AID.

Pour d'autres MAR-BP dont l'expression semble être augmentée dans les cellules B du CG (et dont les cKO ne sont pas disponibles au KOMP, c'est le cas pour ARID3B et CUX1) nous avons envisagé une stratégie de knock down combinée à une reconstitution de moelle osseuse, telle que décrite par l'équipe de Sébastien Storck (Girelli Zubani et al., 2017). Cette stratégie repose sur la collecte de cellules immatures issues de la moelle osseuse de souris *wt* qui seront transduites *in vitro* par un outil retroviral (Sh-ARN) adapté au KD d'une MAR-BP d'intérêt. Une fois les cellules de moelle transduites, les précurseurs lymphoïdes seront injectés chez des animaux immunodéficients (*RAG2<sup>-/-</sup>* ; *γc<sup>-/-</sup>*) dans le but d'induire une reconstitution. La reconstitution de moelle permet de reproduire le développement B et d'obtenir, dans l'animal, des cellules B matures transduites et non-transduites qui nous permettront d'évaluer le processus de SHM dans des conditions plus optimales. Des stratégies identiques sont également envisagées pour la surexpression des protéines dont l'expression semble diminuée dans les cellules B du CG (SATB1 et SATB2).

Une des limites à l'étude du mécanisme de SHM est l'obtention d'une population cellulaire adéquate (cellules B activées du CG) en nombre suffisant. L'optimisation de modèles *in vitro* à partir de cellules fraîchement collectées *in vivo* pourrait représenter un avantage majeur, notamment pour les expérimentations nécessitant un grand nombre de cellules (4C, ChIP et ChIP-seq...). Un modèle utilisant des cellules B naïves de souris a particulièrement retenu notre attention. Ce modèle, basé sur un système de co-culture cellulaire, permet d'obtenir des cellules de type CG (nommées iCG pour induced CG) caractérisées par l'expression des marqueurs typiques du CG (GL7, Fas et PNA). Dans ce système, des cellules B naïves extraites de la rate sont cultivées sur une couche de fibroblastes nourriciers (sécrétant le facteur de survie BAFF et exprimant à leur



surface la molécule CD40L) en présence d'IL4. Dans ce modèle, l'expression forcée de AID conduit à l'introduction de mutation (à une fréquence faible) dans les gènes d'Ig (Nojima et al., 2011). Il serait pertinent d'utiliser ce système qualifié de « induced CG » pour affiner et compléter nos études sur nos modèles animaux.



# *Annexes*





## Annexes.

---

### **Article 1 : Physiological and druggable skinpping of immunoglobulin variable exons in plasma cells.**

Ashi Mohamad Omar, Srour Nivine, Lambert Jean-Marie, Marchalot Anne, Martin Ophélie, Sandrine Le Noir, Pinaud Eric, Ayala Maria Victoria, Sirac Christophe, Saulière, Jérôme, Moreaux Jérôme, Cogné Michel and Delpy Laurent.

Cellular and Molecular Immunology, accepté pour publication, août 2018.





## ARTICLE

## Physiological and druggable skipping of immunoglobulin variable exons in plasma cells

Mohamad Omar Ashi<sup>1</sup>, Nivine Srour<sup>4</sup>, Jean-Marie Lambert<sup>1</sup>, Anne Marchalot<sup>1</sup>, Ophélie Martin<sup>1</sup>, Sandrine Le Noir<sup>1</sup>, Eric Pinaud<sup>1</sup>, Maria Victoria Ayala<sup>1</sup>, Christophe Sirac<sup>1</sup>, Jérôme Saulière<sup>1</sup>, Jérôme Moreaux<sup>2</sup>, Michel Cogné<sup>1,3</sup> and Laurent Delpy<sup>1</sup>

The error-prone V(D)J recombination process generates considerable amounts of nonproductive immunoglobulin (Ig) pre-mRNAs. We recently demonstrated that aberrant Ig chains lacking variable (V) domains can be produced after nonsense-associated altered splicing (NAS) events. Remarkably, the expression of these truncated Ig polypeptides heightens endoplasmic reticulum stress and shortens plasma cell (PC) lifespan. Many questions remain regarding the molecular mechanisms underlying this new truncated Ig exclusion (TIE-) checkpoint and its restriction to the ultimate stage of B-cell differentiation. To address these issues, we evaluated the extent of NAS of Ig pre-mRNAs using an Ig heavy chain (IgH) knock-in model that allows for uncoupling of V exon skipping from TIE-induced apoptosis. We found high levels of V exon skipping in PCs compared with B cells, and this skipping was correlated with a biallelic boost in IgH transcription during PC differentiation. Chromatin analysis further revealed that the skipped V exon turned into a pseudo-intron. Finally, we showed that hypertranscription of Ig genes facilitated V exon skipping upon passive administration of splice-switching antisense oligonucleotides (ASOs). Thus, V exon skipping is coupled to transcription and increases as PC differentiation proceeds, likely explaining the late occurrence of the TIE-checkpoint and opening new avenues for ASO-mediated strategies in PC disorders.

Cellular & Molecular Immunology \_\_\_\_\_; <https://doi.org/10.1038/s41423-018-0160-6>

## INTRODUCTION

During the V(D)J recombination process, combinatorial and junctional diversity ensures that the immunoglobulin (Ig) primary repertoire is adapted to the vast heterogeneity of antigens. However, random nucleotide additions or deletions generate numerous out-of-frame V(D)J junctions that lead to the appearance of premature termination codons (PTCs) in Ig mRNAs.<sup>1,2</sup> The nonsense-mediated mRNA decay (NMD) pathway is very active in B-lineage cells and ensures the rapid degradation of PTC-containing Ig mRNAs, limiting the synthesis of truncated Ig polypeptides.<sup>3–6</sup>

Whereas NMD basically protects cells from truncated protein synthesis, the activation of another RNA surveillance pathway referred as nonsense-associated altered splicing (NAS) can exert opposite effects and yield internally deleted mRNAs and proteins.<sup>7–9</sup> We previously observed that the presence of nonsense codons within the variable (V) exon enhanced exon skipping and the production of Igk light chains without V domains ( $\Delta V$ -kLCs).<sup>5,10</sup> Remarkably, these truncated Ig chains exhibited toxic effects and induced endoplasmic reticulum (ER) stress-associated apoptosis in antibody-secreting cells.<sup>10</sup> Hence, the production of  $\Delta V$ -kLCs blunts plasma cell (PC) differentiation through activation of an antigen-independent process known as the truncated Ig Exclusion (TIE-) checkpoint. This novel PC

checkpoint modifies the Ig repertoire by eliminating terminally differentiated cells harboring biallelic Igk rearrangements with a PTC-containing V exon ( $V^{PTC}$ ) on the nonproductive allele. Thus, aberrantly rearranged Ig alleles are frequent but not innocuous, and their translation into truncated Ig chains is often possible after skipping of  $V^{PTC}$  exons.

Because the TIE-checkpoint provokes PC death as a consequence of deregulated proteostasis, its activation per se precludes the analysis of V exon skipping. Accordingly, alternatively spliced mRNAs (alt-mRNAs) encoding structurally abnormal Ig chains are hardly detectable in PCs.<sup>10</sup> To overcome this issue and decipher how exon skipping events are regulated during PC differentiation, we utilized a mouse model harboring an additional frameshift-inducing V exon (frVk) at the Ig heavy (IgH) chain locus.<sup>11,12</sup> Using this dedicated model, skipping of the supplemental frVk exon does not encode truncated Ig chains provided in-frame VDJ junctions. Therefore, by uncoupling Ig exon skipping to the activation of the TIE-checkpoint, we could depict the regulation of NAS of IgH pre-mRNAs in vivo. As a new strategy for specifically affecting the fate of PCs transcribing a given V exon, we also assessed whether V exon skipping was achievable with splice-switching antisense oligonucleotides (ASOs). The proof of the concept for ASO-mediated V exon skipping should thus open new

<sup>1</sup>Unité Mixte de Recherche Centre National de la Recherche Scientifique 7276, INSERM U1262—Contrôle de la Réponse Immune B et Lymphoproliférations, Université de Limoges, Limoges, France; <sup>2</sup>Institute of Human Genetics, CNRS-UM UMR9002, Department of Biological Haematology, CHU Montpellier, University of Montpellier, UFR Medecine, Montpellier, France and <sup>3</sup>Institut Universitaire de France, Université de Limoges, Limoges, France

Correspondence: Laurent Delpy (laurent.delpy@unilim.fr)

<sup>4</sup>Present address: Lady Davis Institute for Medical Research, McGill University, 3755 Cote Ste-Catherine Road, Montreal, QC H3T 1E2, Canada

These authors contributed equally: Mohamad Omar Ashi, Nivine Srour.

Received: 30 April 2018 Accepted: 7 July 2018



avenues for “inducible-TIE” therapeutic approaches in PC dyscrasias.

## RESULTS

### Determining exon skipping of IgH transcripts in vivo

To evaluate the extent of nonsense-associated Ig exon skipping during PC differentiation, experiments were performed using heterozygous IgH<sup>wt/frVκ</sup> knock-in mice.<sup>11,12</sup> In this model, the expression of VDJ-rearranged IgH<sup>wt</sup> alleles drives B-cell development, whereas the supplemental frVκ exon leads to the appearance of PTCs that trigger NMD degradation of full-length IgH<sup>frVκ</sup> mRNAs.<sup>6</sup> By contrast, the activation of the NAS pathway can yield functional IgH chains by eliminating the frVκ exon from transcripts with in-frame VDJ junctions (Fig. 1a).

In IgH<sup>wt/frVκ</sup> mice with a mixed B6/129 background, allotypic differences allowed for us to distinguish between IgM<sup>b</sup> expression from the IgH<sup>wt</sup> allele and IgM<sup>a</sup> produced upon exon skipping of VDJ-rearranged IgH<sup>frVκ</sup> pre-mRNAs. The secretion of total IgM, IgM<sup>a</sup>, and IgM<sup>b</sup> was assessed in the sera of B6, F1 (B6/129), and IgH<sup>wt/frVκ</sup> mice using standard or allotype-specific ELISA assays. Sera from B6 mice served as negative controls for anti-IgM<sup>a</sup> assays. As expected, similar serum concentrations of total IgM and IgM<sup>b</sup> were found in B6 and IgH<sup>wt/frVκ</sup> mice (Fig. 1b, d). Interestingly, we found significant amounts of IgM<sup>a</sup> in sera from IgH<sup>wt/frVκ</sup> mice, albeit diminished compared with F1 (B6/129) mice in which ~50% of the B-lineage cells express a functional IgHa allotype (Fig. 1c). Thus, NAS is involved in the processing of nonproductive IgH pre-mRNAs, eliciting active translation of alternatively spliced (alt-) mRNAs in PCs harboring biallelically VDJ-rearranged IgH alleles.

### High rate of NAS of IgH pre-mRNAs in PCs

To assess the regulation of NAS of IgH pre-mRNAs at the molecular level, alt-mRNA amounts were quantified in B and PCs. For this quantification, IgH<sup>frVκ</sup> knock-in mice were crossed with DH-LMP2A mice harboring a replacement of JH segments by the Epstein–Barr virus *LMP2A* gene under the control of a DH promoter.<sup>13</sup> The expression of *LMP2A* drives B-cell development and PC differentiation.<sup>13,14</sup> In heterozygous IgH<sup>DH-LMP2A/frVκ</sup> mice, VDJ-rearranged IgH transcripts arise solely from the nonproductive IgH<sup>frVκ</sup> allele, facilitating the analysis of NAS of IgH<sup>frVκ</sup> pre-mRNAs. B cells (B220<sup>+</sup>/CD138<sup>-</sup>) and PCs (B220<sup>-</sup>/CD138<sup>+</sup>) were isolated from the spleens of IgH<sup>DH-LMP2A/frVκ</sup> mice (Fig. 2a), and RT-PCR analysis revealed high alt-mRNA amounts in PCs whereas a faint band was detected in B cells (Fig. 2b). Next, the rate of NAS of IgH<sup>frVκ</sup> transcripts was determined by qPCR and corresponded to alt-mRNA/pre-mRNA ratios. Likewise, this NAS rate was strongly increased in PCs compared with B cells (Fig. 2c). We then analyzed the VDJ repertoire of alt-mRNAs, as described.<sup>15</sup> Highly diverse productive CDR3 junctions were obtained among the 11,506 clonotypes identified using IMGT/HighV-QUEST<sup>16</sup> (Fig. 2d). In addition, VDJ junctions involved numerous VH and DH families and all the JH segments (Supplementary Fig. 1). The near absence (0.35%) of nonproductive CDR3 junctions reflected the strong degradation of PTC-containing Igμ mRNAs<sup>6</sup> and the coupling of the NMD and NAS pathways.<sup>17</sup> This repertoire analysis revealed that alt-mRNAs arose from polyclonal PCs rather than rare clones selected after splice site mutations. Hence, a marked upregulation of NAS of IgH pre-mRNAs occurs during PC differentiation.

### Elevated expression of VDJ-rearranged IgH alleles and RNA splicing genes during PC differentiation

Many studies have reported that the rate of transcription by RNA polymerase II (RNAPII) can regulate alternative splicing.<sup>18,19</sup> To address this question, the transcription of nonproductive IgH<sup>frVκ</sup> alleles was evaluated by quantifying pre-mRNA levels in B cells (B220<sup>+</sup>/GL7<sup>-</sup>/FAS<sup>-</sup>/CD138<sup>-</sup>), germinal center (GC-) B cells (B220<sup>+</sup>/GL7<sup>+</sup>/FAS<sup>+</sup>/CD138<sup>-</sup>), and PCs (B220<sup>-</sup>/CD138<sup>+</sup>) isolated from

spleens of SRBC-immunized IgH<sup>wt/frVκ</sup> mice (Fig. 3a). Consistent with the increase in Ig gene transcription during PC differentiation,<sup>20</sup> we found approximately threefold more IgH<sup>frVκ</sup> pre-mRNA levels in PCs than in B cells (Fig. 3b). Next, we performed a comparative analysis of RNAPII binding on productive and nonproductive IgH alleles in LPS-stimulated B cells isolated from IgH<sup>wt/frVκ</sup> mice. Supporting a biallelic IgH transcription pattern, chromatin immunoprecipitation (ChIP) experiments revealed similar phosphorylated serine 2 (Ser2P) and 5 (Ser5P) RNAPII levels on both IgH alleles (Fig. 3c, d). The nuclear location of IgH alleles was further analyzed in B cells, LPS-stimulated B cells, GC-B cells, and PCs by DNA-FISH using IgH-specific probes (Fig. 3e). We observed that asymmetric heterochromatin recruitment of one IgH allele was extremely rare (<20%) in the vast majority of B-lineage cells, including terminally differentiated PCs.

Next, the expression of genes involved in RNA splicing (Gene ontology ID GO:0008380, *n* = 360) was assessed in purified follicular (FO-) B cells, GC-B cells, plasmablasts (PBs), and PCs using available microarray dataset GSE26408 and GenomicScape platform ([www.genomicscape.com](http://www.genomicscape.com)).<sup>21</sup> Multiclass analysis revealed that the expression profile of RNA splicing regulators is highly modified during PC differentiation, with 249 (~70%) genes differentially expressed, and among them, 106 (>40%) were significantly overexpressed in PCs (fold change > 1.5 and *P* < 0.05) (Fig. 4; Supplementary Table 1). These data underline specific gene expression patterns related to RNA splicing at the different stages of B-cell maturation with a significant enrichment of genes related to the spliceosomal complex in PCs as identified by gene set expression analysis.

Thus, biallelic hypertranscription of IgH alleles together with huge modifications of the RNA splicing environment are triggered during PC differentiation, and this was correlated with the reinforcement of NAS of IgH pre-mRNAs.

### Delineating the chromatin profile of the alternative frVκ exon

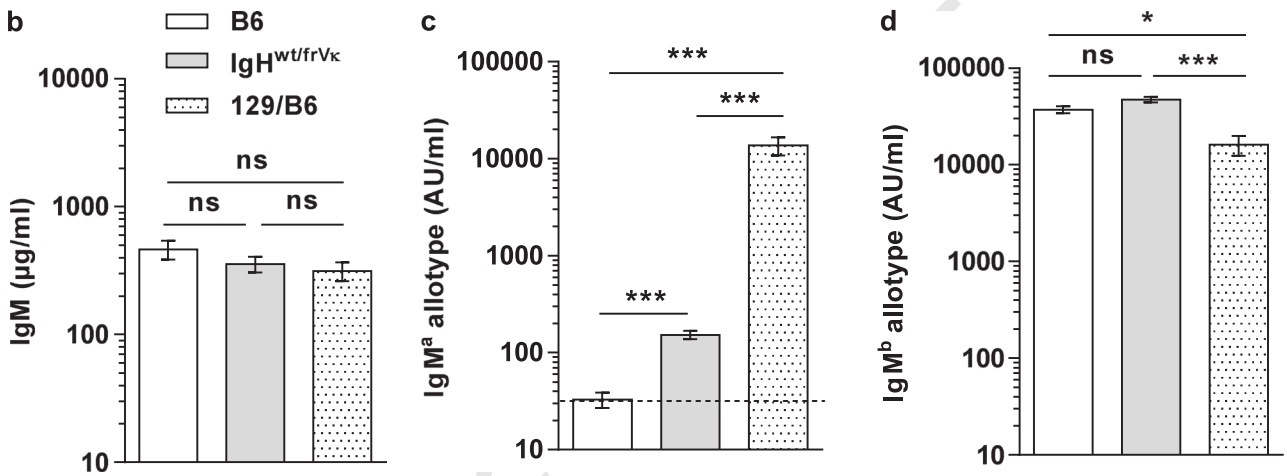
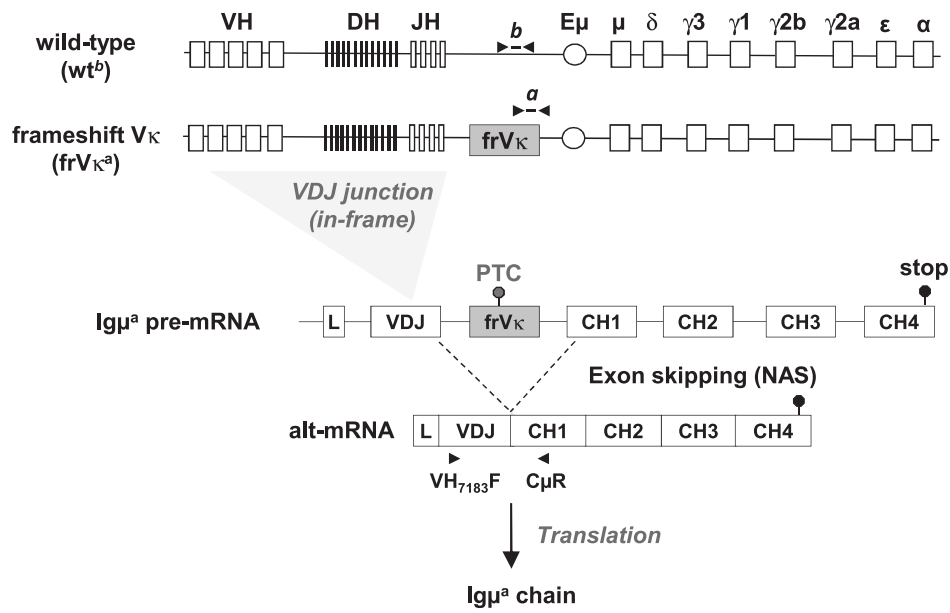
To evaluate the interplay between NAS and chromatin structure, a comparative analysis of specific histone H3 marks was conducted among the alternative frVκ exon, constitutive CH1μ exon, and intronic “JH-Eμ” sequence (Fig. 5a). ChIP experiments were performed using anti-H3K4me3, H3K9me3, H3K9ac, and H3K36me3 antibodies in LPS-stimulated B cells isolated from IgH<sup>wt/frVκ</sup> mice. H3K4me3, H3K9ac, and H3K36me3 marks are associated with transcriptional activation, whereas H3K9me3 is linked to transcriptional repression.<sup>22,23</sup> Remarkably, the alternative frVκ exon and the “JH-Eμ” intron exhibited highly similar chromatin profiles for all epigenetic marks investigated (Fig. 5b–e). By contrast, alternative frVκ and constitutive CH1μ exons displayed striking differences. The levels of H3K4me3 and H3K9ac increased (Fig. 5b, c), whereas H3K9me3 amounts decreased in the alternative frVκ exon compared with the constitutive CH1μ exon (Fig. 5e). The comparative analysis of H3K36me3 levels did not reveal any significant changes (Fig. 5d). Supporting the idea that epigenetic histone marks can influence the pattern of alternative splicing by favoring exclusion or inclusion of alternative exons,<sup>22,24–26</sup> skipped frVκ exons exhibited an “intron-like” chromatin signature.

### ASO-mediated exon skipping of Ig transcripts

In recent years, the considerable improvement in terms of stability and cellular penetration of ASOs has driven the development of these molecules for medical purposes in various genetic diseases and, in particular, for the treatment of neuromuscular disorders.<sup>27–29</sup> By simply monitoring the production of IgM<sup>a</sup> following exon skipping events, the IgH<sup>wt/frVκ</sup> mouse model represents a unique tool facilitating the analysis of ASO-mediated Ig exon skipping approaches in primary B and PCs. To address this issue, LPS-stimulated B cells were treated with a specific “vivo-morpholino” ASO targeting the frVκ exon donor splice site (ASO-frVκ<sub>ds</sub>) (Fig. 6a)



**a** Mouse IgH locus



**Fig. 1** Assessment of exon skipping of nonproductive VDJ-recombined IgH transcripts in vivo. **a** Map of wild-type (*b* allotype) and targeted (*a* allotype) IgH loci. IgH<sup>wt/frVκ</sup> knock-in mice were created by insertion of an additional frameshift-inducing Vκ exon (frVκ) between J<sub>H</sub>4 and Eμ. Skipping of frVκ exon occurs after NAS of nonproductive IgH<sup>frVκ</sup> pre-mRNAs, allowing for the expression of complete IgH<sup>a</sup> chains upon in-frame VDJ rearrangement. Primers and probes used to detect specifically IgH<sup>wt</sup> (*b*) and IgH<sup>frVκ</sup> (*a*) alleles by qPCR are represented (black arrows and short rectangles). **b-d** To determine the amounts of total IgM (**b**), IgM<sup>a</sup> (**c**), and IgM<sup>b</sup> (**d**) allotypes, ELISA assays were performed in sera from 6 to 8-week-old IgH<sup>wt/frVκ</sup> (wt<sup>b</sup>/frVκ<sup>a</sup>), B6/129 F1 (wt<sup>b</sup>/wt<sup>b</sup>) and B6 (wt<sup>b</sup>/wt<sup>b</sup>) mice (*n* = 3–8/group). **c** The negative threshold of IgM<sup>a</sup> expression was obtained using sera from B6 mice (dotted line). Bars represent the mean expression ± SEM. Unpaired two-tailed Student's *t* test was performed to determine significance (ns not significant; \**P* < 0.05; \*\*\**P* < 0.001)

or with an irrelevant ASO as a control (ASO-ctrl). Remarkably, the passive administration of ASO-frVκ<sub>dss</sub> strongly increased the secretion of IgM<sup>a</sup> in culture supernatants (Fig. 6b). We also observed that both the frequency of IgM<sup>a</sup>-positive cells (Fig. 6c, d) and the mean fluorescence intensity (MFI) of IgM<sup>a</sup> (Fig. 6e) were increased after ASO-frVκ<sub>dss</sub> treatment compared with the ASO-ctrl. Based on previous observations showing that biallelic VDJ-rearrangements are retrieved in ~40% of B-lineage cells<sup>30</sup> and that ~22% of VDJ-rearranged IgH<sup>frVκ</sup> transcripts exhibited in-frame CDR3 junctions,<sup>6</sup> we estimated a theoretical maximum of IgM<sup>a</sup>-expressing cells of ~8.8%. Interestingly, the frequency of PB-expressing IgM<sup>a</sup> reached ~7.5% after ASO-frVκ<sub>dss</sub> treatment, indicating that passive administration of ASOs was highly efficient (~85% of the theoretical maximum) and targeted nearly all antibody-secreting cells (Fig. 6d). Likewise, we found that ASO-

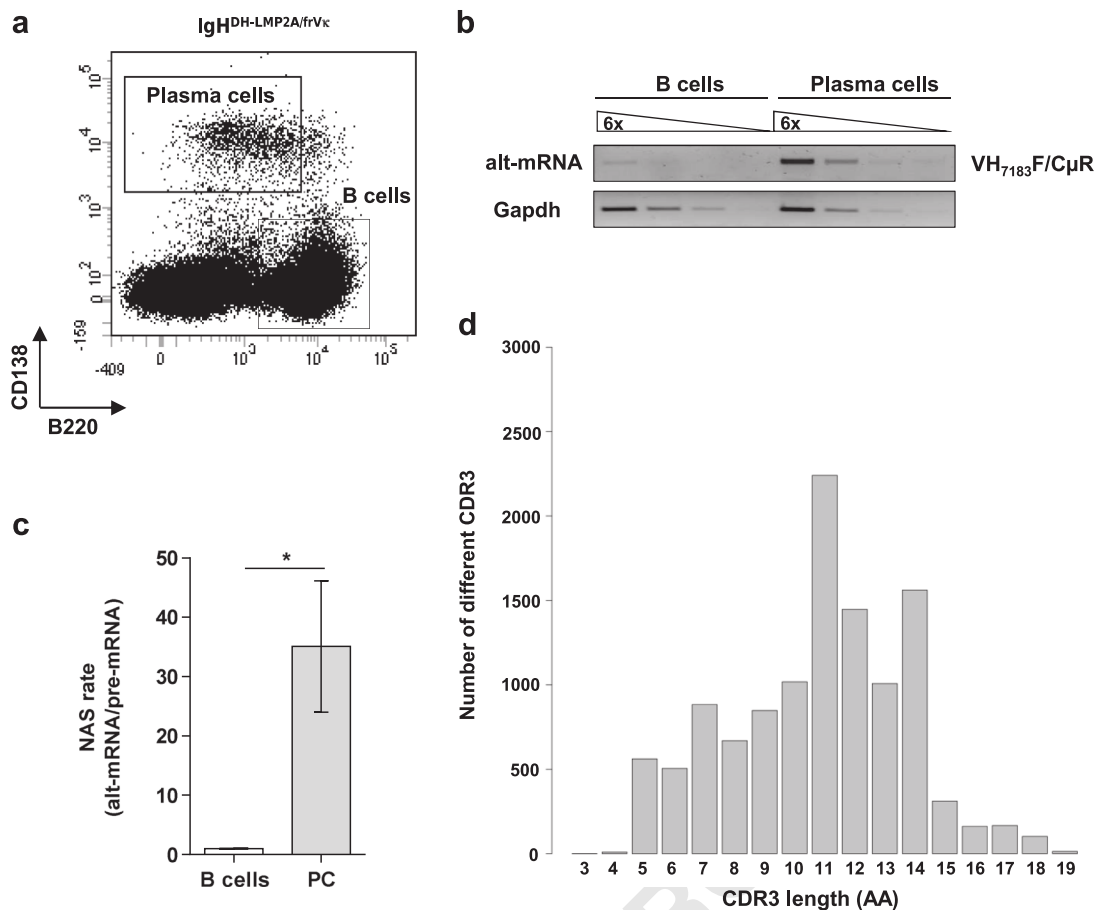
frVκ<sub>dss</sub> treatment drastically increased the rate of frVκ exon skipping by monitoring alt-mRNA/pre-mRNA ratios (Fig. 6f). Of note, we observed similar proliferative responses (Supplementary Fig. 2a) and Igk light chain production (Supplementary Fig. 2b) in LPS-stimulated B cells treated or not with ASOs, indicating that ASO treatment did not induce adverse toxic effects. Hence, passive administration of ASO targeting V exon donor splice sites strongly induced exon skipping in PCs.

**DISCUSSION**

Alternative splicing regulates gene expression and affects more than 90% of multiexonic pre-mRNAs in humans.<sup>31–33</sup> Likewise, ASO-mediated splicing modulation can be exploited for numerous therapeutic purposes.<sup>34</sup> In this study, we provide evidence that







**Fig. 2** Comparative analysis of NAS of IgH pre-mRNAs in B cells and PCs. **a** B cells and PCs were isolated from the spleens of IgH<sup>DH-LMP2A/frVκ</sup> mice after staining with anti-B220 and anti-CD138 mAbs. Representative dot plot and gates used for cell sorting are depicted. The purity of sorted populations was always above 90%. **b** To identify alternatively spliced mRNAs (alt-mRNAs), RT-PCR were performed using VH<sub>7183</sub>F/CμR primers. One representative experiment out of three performed is shown. **c** Alt-mRNA and IgH<sup>frVκ</sup> pre-mRNA levels were assessed by qPCR using VH<sub>7183</sub>F/CμR primers and the probe *a* respectively, as described in Fig. 1a. NAS rates corresponding to the relative alt-mRNA/pre-mRNA ratios were determined in sorted B cells (empty bar) and PCs (gray bar) (*n* = 5). **d** VDJ repertoire analysis of alt-mRNAs showing polyclonal CDR3 junctions. RACE-PCR, high throughput sequencing and CDR3 length analysis were performed as described in methods and in Fig. S1. Unpaired two-tailed Student's *t* test was performed to determine significance (\**P* < 0.05)

the biallelic boost of Ig gene transcription accompanying PC differentiation promotes NAS of VDJ-rearranged IgH pre-mRNAs and facilitates ASO-mediated skipping of V exons.

Several studies including ours have shown that the pattern of IgH gene transcription is mostly biallelic in B cells and PCs.<sup>6,20,35–37</sup> Accordingly, we found that ~80% of mature B cells and PCs exhibited no IgH allele in heterochromatin. This observed frequency is consistent with previous 3D RNA FISH data showing that >80% of IgH and Igk genes are biallelically transcribed in PCs.<sup>20</sup> ChIP experiments further revealed similar Ser2P and Ser5P RNAPII binding on both IgH alleles in LPS-stimulated B cells. In agreement with a drastic increase in Ig gene transcription during PC differentiation, we found that the amount of nonproductive IgH pre-mRNAs was ~3-fold higher in PCs compared with B cells. Altogether, biallelic hypertranscription of Ig genes generates high levels of nonproductively V(D)J-rearranged Ig pre-mRNAs in antibody-secreting cells.

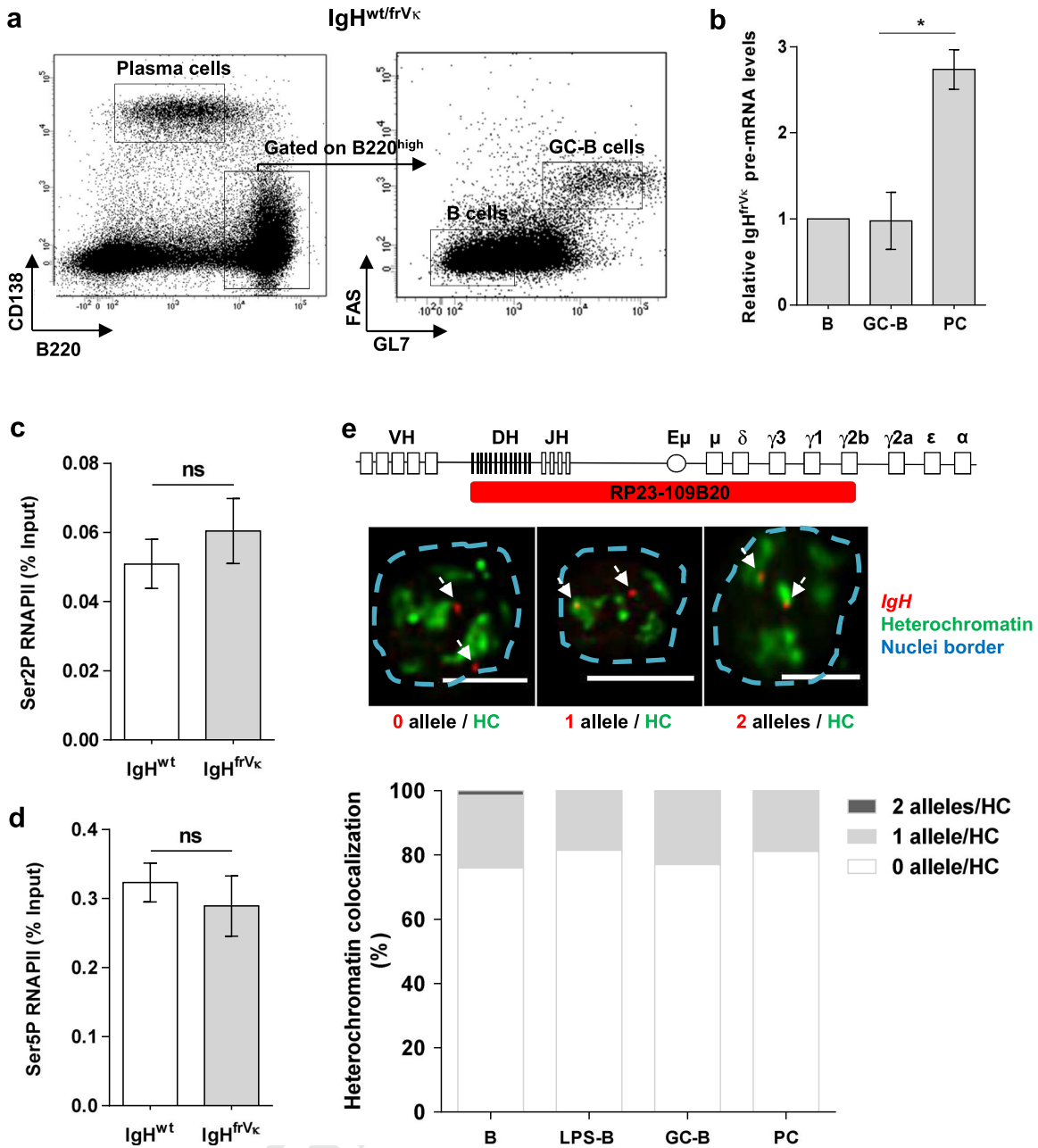
Because the TIE-checkpoint eliminates PC-expressing truncated Ig chains,<sup>10</sup> it remains very challenging to evaluate the extent of exon skipping during splicing of Ig pre-mRNAs. Using our dedicated IgH<sup>frVκ</sup> model allowing for skipping of V exon without production of shortened Ig polypeptides, we provided evidence that NAS of IgH pre-mRNAs is greatly enhanced in PCs compared with B cells. These data strongly suggest that NAS and conventional alternative splicing are governed by similar rules

and controlled by the transcription elongation rate.<sup>19,26,38–40</sup> The transcriptomic analysis in Fig. 4 also indicates that PC differentiation entails major changes with regard to the expression of genes involved in RNA splicing. Further investigations are necessary to define the role of differentially expressed RNA splicing regulators. Therefore, increasing Ig gene expression in PCs with biallelic V(D)J-rearrangements favors antibody production but detrimentally induces activation of the TIE-checkpoint after nonsense-associated exon skipping events.

Several studies have identified a specific histone mark signature around alternative exons.<sup>22,24–26</sup> Interestingly, trimethylation and acetylation of H3K9 have been associated with inclusion or exclusion of alternative exons, respectively.<sup>19,25</sup> High levels of H3K9me3 promotes heterochromatin protein 1γ (HP1γ) recruitment and provokes a local slowdown of RNAPII, supporting the inclusion of CD44 variant exons.<sup>25</sup> By contrast, a local increase in H3K9ac is associated with the exclusion of NCAM (neural cell adhesion molecule) exon 18 during the depolarization of human neuronal cells.<sup>19</sup> Remarkably, we observed a massive enrichment in H3K9ac and an opposite decrease in H3K9me3 at the skipped frVκ exon compared with the constitutive CH1μ exon. Altogether, this chromatin profile revealed that the alternative frVκ exon is turned into a pseudo-intron.

This study provides evidence that NAS of Ig pre-mRNAs is coupled to transcription and increases as PC differentiation





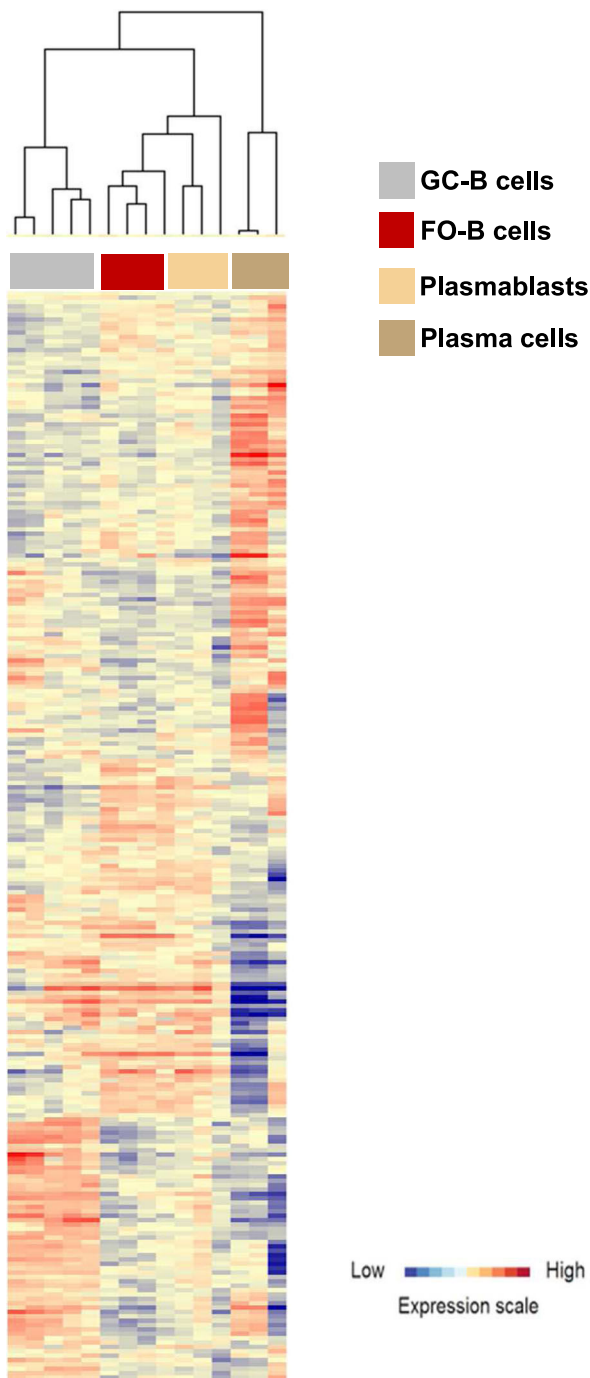
**Fig. 3** Biallelic transcription of IgH alleles during PC differentiation. **a** B cells, germinal center (GC-) B cells, and PCs were isolated from the spleens of SRBC-immunized  $IgH^{wt/frV\kappa}$  mice after staining with anti-B220, anti-GL7, anti-FAS, and anti-CD138 mAbs. Representative dot plots and gates used for cell sorting are depicted. The purity of each sorted population was always above 90%. **b** Relative  $IgH^{frV\kappa}$  pre-mRNA levels were determined by qPCR as described in Fig. 1a (probe *a*) after normalization to *Gapdh* mRNA expression. The value obtained for B cells served as a reference and was set to 1 in each FACS sorting experiment ( $n = 3$ ). **c, d** Resting B cells were isolated from the spleens of  $IgH^{wt/frV\kappa}$  mice and stimulated with LPS (1  $\mu$ g/ml) for 3 days. ChIP experiments were performed using phospho-specific antibodies against phosphorylated Ser2 (Ser2P) and Ser5 (Ser5P) RNAPII ( $n = 4$ ). Relative RNAPII enrichments (percent input) were analyzed on productive and nonproductive IgH alleles by qPCR, as described in Fig. 1a (probes *b* and *a* to detect  $IgH^{wt}$  and  $IgH^{frV\kappa}$  alleles, respectively). Background signals from mock samples with irrelevant antibody staining were subtracted. **e** Map of the IgH locus detailing genomic regions covered by BAC used as FISH probes. Representative 3D-FISH images of B-cell nuclei from B6 mice, reporting no allele or 1 or 2 alleles within heterochromatin (scale bars: 5  $\mu$ m). IgH alleles are indicated by white arrows. Bars represent the percentage of IgH alleles within the heterochromatin, determined in resting B cells ( $n = 137$ ), LPS-stimulated B cells ( $n = 97$ ), GC-B cells ( $n = 52$ ) and PC ( $n = 53$ ) nuclei. For each cell type, data were obtained from three individual mice. Unpaired two-tailed Student's *t* test was performed to determine significance (ns not significant; \* $P < 0.05$ )

proceeds. We propose that the late occurrence of TIE-induced apoptosis eliminating PCs with biallelic Ig gene rearrangements results from the combined effects of enhanced NAS activation during splicing of nonproductive Ig pre-mRNAs and the intrinsic sensitivity of antibody-secreting cells towards prolonged ER stress. Altogether, these data reinforce the idea that nonproductively V

(D)J-recombined alleles can be considered as drivers rather than passengers at late steps of B-cell differentiation.

Finally, we validated that ASO-mediated Ig exon skipping can be easily achieved by targeting the donor splice site of V exons. Thus, the use of ASOs to induce massive production of detrimental truncated Ig chains represents an attractive therapeutic approach





**Fig. 4** Analysis of RNA splicing gene expression profiles during PC differentiation. Clustergram of GO RNA splicing gene list significantly differentially expressed between follicular B cells (FO B cells,  $n = 3$ ), germinal center B cells (GC B cells,  $n = 5$ ), plasmablasts (PBs,  $n = 4$ ), and plasma cells (PCs,  $n = 3$ ) (fold change > 1.5 and FDR < 0.05)

that should be of great interest in patients with monoclonal gammopathies.

**MATERIALS AND METHODS**

**Mice**  
Igh<sup>frVk</sup> mice harboring an extra frameshift-inducing Vk exon between J<sub>H</sub>4 and E<sub>μ</sub> have been described elsewhere previously.<sup>6</sup> Heterozygous Igh<sup>wt/frVk</sup> mice were backcrossed to C57BL/6 (B6) background to distinguish between Igh<sup>WT</sup> (*b* allotype) and Igh<sup>frVk</sup>

(a allotype) alleles. F1 (B6<sup>b</sup>/129<sup>a</sup>) mice were used as a control. Igh<sup>DH-LMP2A</sup> mice harbor a replacement of JH segments by the Epstein-Barr virus *LMP2A* gene under the control of a DH promoter.<sup>13</sup> The expression of LPM2A mimics the BCR tonic signal, allowing for B-cell development and PC differentiation.<sup>13,14</sup> Two- to three-month-old mice were used in all experiments and maintained in our animal facilities, at 21–23 °C with a 12-h light/dark cycle. Experiments were performed according to the guidelines of our institutional review board for animal experimentation (No. CREEAL 6-07-2012).

**ASO treatments**

Vivo-morpholino ASOs (ASO-frVk<sub>ds</sub>: 5'-ATGCTCGAGACT-TACCCGTTTGATT-3') and an irrelevant ASO (ASO-Ctrl: 5'-CCTTTACCTCAGTTACAATTATA-3') were designed and purchased at Gene Tools, LLC. Splenic B cells isolated from Igh<sup>wt/frVk</sup> mice were purified by negative selection using anti-CD43 magnetic beads (Miltenyi Biotec) and stimulated (0.8×10<sup>6</sup> cells/ml) with 1 μg/ml lipopolysaccharide (LPS) (LPS-EB Ultrapure; InvivoGen) in RPMI 1640 with 10% fetal calf serum. At day 2, cells were harvested and incubated with 10 μM AON (4 h in PBS). After incubation, the cells were stimulated with LPS for 2 days in the presence of AON (1 μM) in culture media.

**Flow cytometry and cell sorting**

Erythrocyte-depleted spleen cells were labeled with anti-mouse CD138 (281-2; BD Pharmingen), B220 (RA3-6B2; BioLegend), anti-mouse- and B-cell Activating Ag (GL7; BD Pharmingen) and anti-mouse CD95 (Jo2) mAbs (BD Pharmingen). Intracellular staining was performed using anti-mouse IgM<sup>a</sup> (DS-1; BD Pharmingen) and IgM<sup>b</sup> (AF6-78; BD Pharmingen) mAbs, according to the manufacturer's instructions. Cells were analyzed on an LSR II Fortessa apparatus using FACS-Diva software (BD Pharmingen). B cells, GC-B cells, and PCs were sorted from the spleens of B6 or Igh<sup>wt/frVk</sup> mice 7 days after intraperitoneal (i.p.) injection of sheep red blood cells (SRBCs, BioMérieux SA-France). Cell sorting was performed using a FACS-Aria III (BD Biosciences), and the gates used are depicted in dot plots (purity was above 90%).

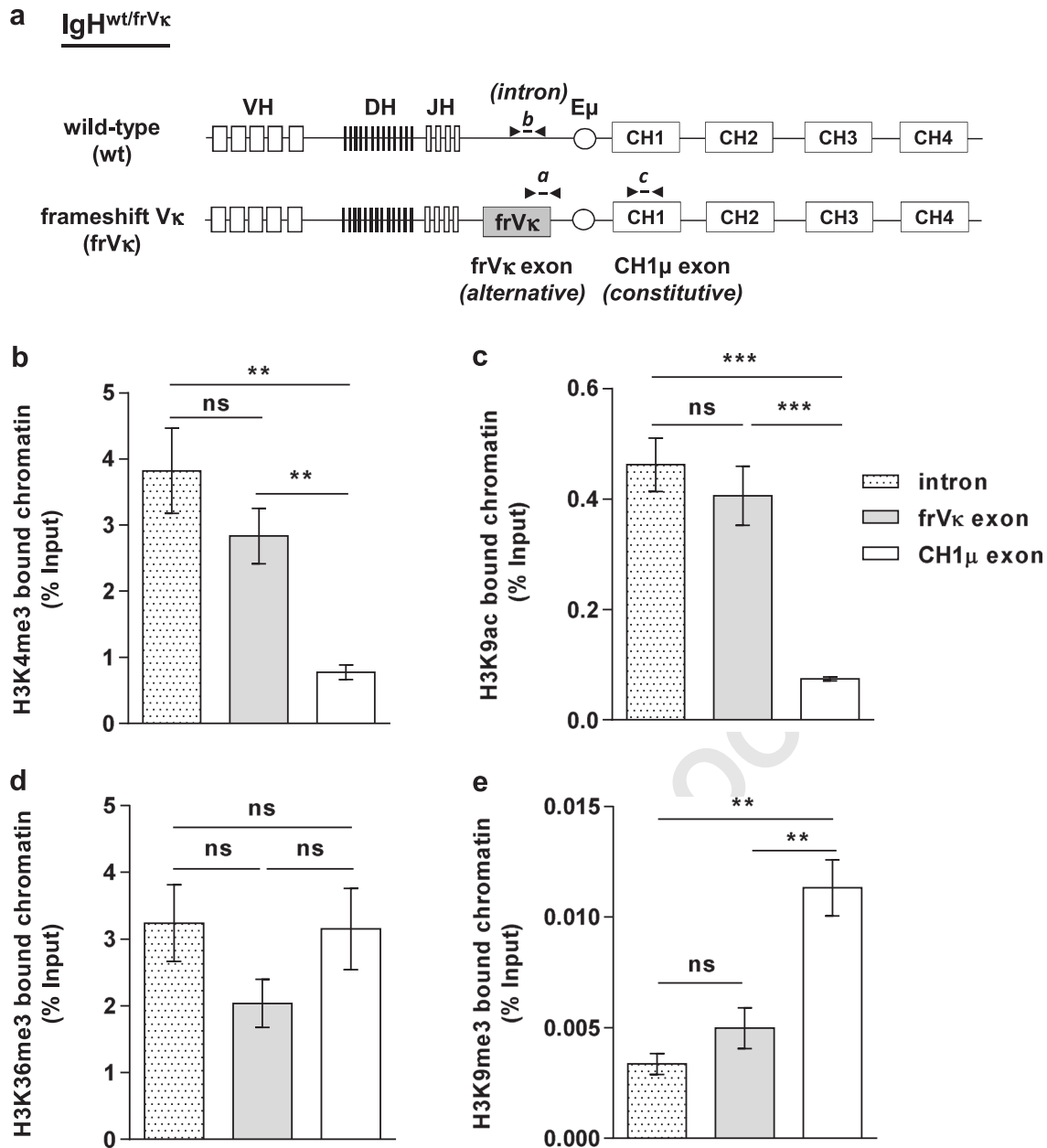
**ELISA assays**

Sera from Igh<sup>wt/frVk</sup> (wt<sup>b</sup>/frVk<sup>a</sup>), B6 (wt<sup>b</sup>/wt<sup>b</sup>), and B6/129 F1 (wt<sup>b</sup>/wt<sup>a</sup>) mice were collected and analyzed using total IgM or allotype-specific IgM<sup>a</sup> and IgM<sup>b</sup> ELISA assays. Ig titers were determined in polycarbonate 96-multiwell plates (Maxisorp, Nunc) using unlabeled IgM (Southern Biotech), IgM<sup>a</sup> (DS-1; BD Pharmingen), and IgM<sup>b</sup> (AF6-78; BioLegend) Abs and alkaline phosphatase-conjugated goat anti-mouse IgM Abs (Southern Biotech), as described.<sup>41,42</sup> Ig amounts were revealed by the addition of p-nitrophenyl phosphate (Sigma-Aldrich) and blocked with the addition of 3 M NaOH; optic density was measured at 405 nm.

**Chromatin immunoprecipitation experiments**

ChIP experiments were performed using anti-H3K4me3 (Millipore, 07-473), anti-H3K9ac (Millipore, 06-942), anti-H3K9me3 (Millipore, 05-1242), anti-H3K36me3 (Abcam, ab9050), anti-RNA Pol II ser2P (Abcam, ab5095), and anti-RNA Pol II ser5P (Abcam, ab5131), as previously described.<sup>43</sup> In brief, 1×10<sup>7</sup> LPS-stimulated B cells from Igh<sup>wt/frVk</sup> mice were harvested at day 3, washed twice in PBS and cross-linked at 37 °C for 15 min in 15 ml of PBS with 1% formaldehyde. The reaction was quenched with 0.125 M glycine. After lysis, chromatin was sonicated to 0.5–1 kb using a Vibracell 75043 (Thermo Fisher Scientific). After dilution in ChIP buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, and 167 mM NaCl), chromatin was precleared by rotating for 2 h at 4 °C with 100 μl of 50% protein A/G slurry (0.2 mg/ml sheared salmon sperm DNA, 0.5 mg/ml BSA, and 50% protein A/G; Sigma), 0.3–0.5×10<sup>6</sup> cell equivalents were saved as input, and 3–5×10<sup>6</sup> cell equivalents were incubated overnight with specific or





**Fig. 5** Analysis of the alternative frVκ exon chromatin profile. **a** Chromatin marks were analyzed comparatively at the JH4-Eμ intron, the alternative frVκ exon, and the constitutive CH1μ exon. Primers and probes used in qPCR are depicted. **b–e** ChIP assays were performed in splenic B cells isolated from IgH<sup>wt/frVκ</sup> mice 3 days after stimulation with LPS (1 μg/ml). Immunoprecipitation experiments were conducted using anti-H3K4me3 (**b**), H3K9ac (**c**), H3K36me3 (**d**), and H3K9me3 (**e**) mAbs. Background signals from mock IP using irrelevant antibodies were subtracted, and relative enrichments were expressed after normalization to total input DNA (percent of input). Data are the mean ± SEM ( $n = 5$ ), and unpaired two-tailed Student's  $t$  test was performed to determine significance (ns not significant; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ )

control antibodies. Immune complexes were precipitated by the addition of protein A/G. Cross-linking was reversed by overnight incubation (70 °C) in TE buffer with 0.02% SDS, and genomic DNA was obtained after phenol/chloroform extraction.

#### PCR and RT-PCR

Total RNA was prepared using Tri-reagent (Invitrogen) procedures. RT-PCR was carried out on 1–3 μg of DNase I (Invitrogen)-treated RNA using Superscript III (Invitrogen). Priming for reverse transcription was done with random hexamers. Quantitative PCR was performed on cDNA samples equivalent to 10 ng of RNA per reaction, using TaqMan or SYBR-Green Universal MasterMix (Applied Biosystems) on a StepOnePlus Real-Time PCR system

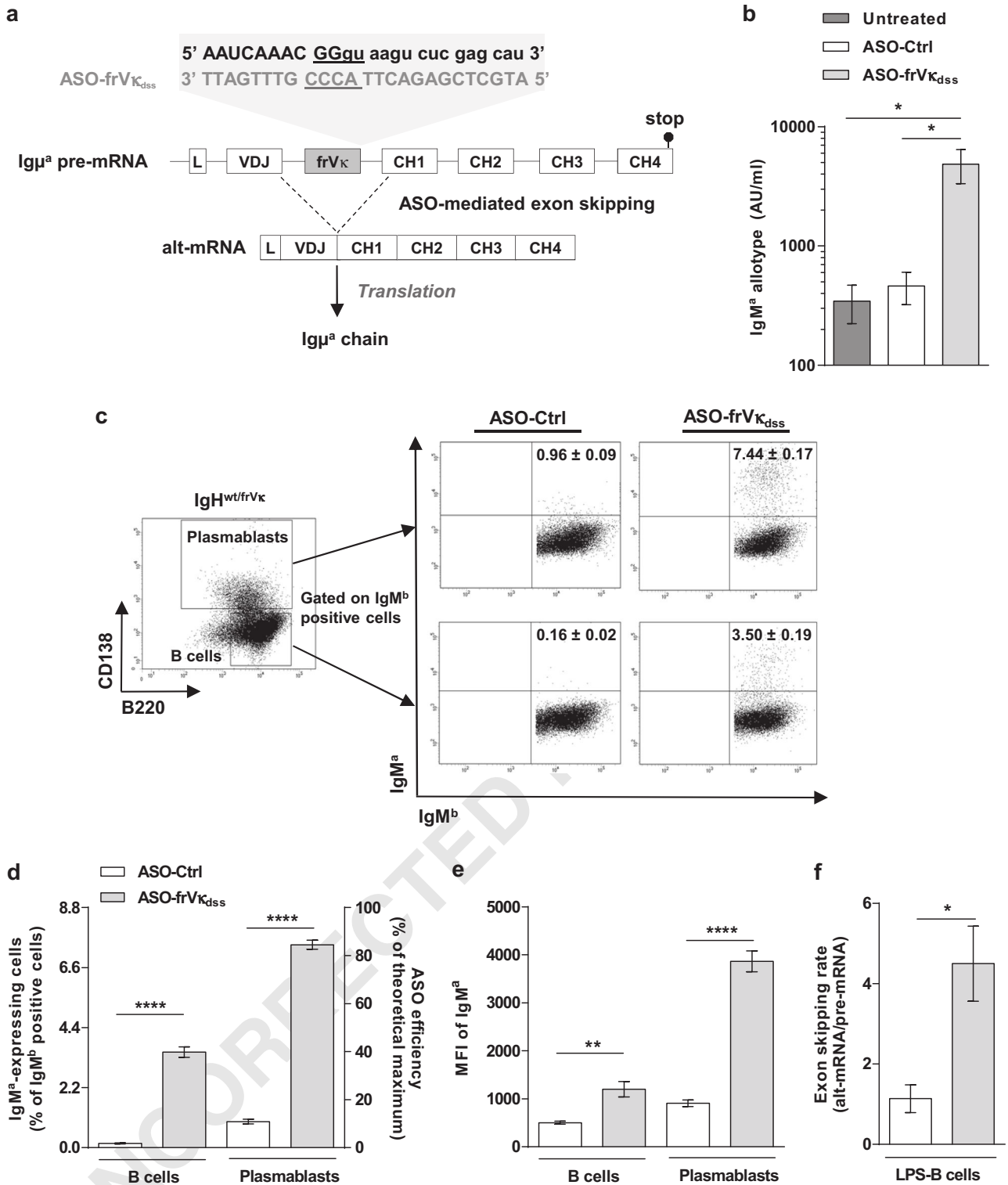
(Applied Biosystems). Transcripts were quantified according to the standard  $2^{-\Delta\Delta Ct}$  method after normalization to *Gapdh* (Mm99999915\_g1). Primers and probes used for Q-PCR are listed in Supplementary Table 2.

#### 3D-DNA FISH experiments

A DNA probe specific to the IgH locus was prepared with the RP23-109B20 BAC construct (Life Technologies) labeled by random priming with digoxigenin-11-dUTP (Roche). A DNA probe specific to heterochromatin compartment was prepared with the pYsat plasmid (kindly provided by Dr. Nial Dillon) containing multiple copies of the 234 bp major satellite ( $\gamma$ -satellite) repeat labeled by random priming with Alexa Fluor 488-dUTP (Life







**Fig. 6** Significant IgM<sup>a</sup> production upon passive administration of ASOs. **a** An ASO sequence targeting the donor splice site of the frV<sub>K</sub> exon (AON-frV<sub>K</sub><sub>dss</sub>) was designed and synthesized as “*vivo*-morpholino ASO” (Gene Tools, LLC). **b** Splenic B cells isolated from IgH<sup>wt/frV<sub>K</sub></sup> mice were stimulated with LPS (1 μg/ml) for 4 days and treated with ASO-frV<sub>K</sub><sub>dss</sub> or irrelevant ASO (ASO-Ctrl) during the last 2 days, as described in the Methods. At day 4, supernatants were harvested, and IgM<sup>a</sup> levels were determined. **c** Cells were labeled with anti-B220, anti-CD138, anti-IgM<sup>b</sup>, and anti-IgM<sup>a</sup> mAbs. Representative dot plots and gates used for FACS analysis are depicted. **d** The left y-axis represents the percentage of IgM<sup>a</sup>-expressing cells in IgM<sup>b</sup>-positive B cells and plasmablasts treated with ASO-frV<sub>K</sub><sub>dss</sub> (gray bars) or ASO-Ctrl (empty bars). The right y-axis shows ASOs’ efficiency as percentage of theoretical maximum of IgM<sup>a</sup>-positive cells. **e** The mean fluorescence intensity (MFI) of IgM<sup>a</sup> was determined on IgM<sup>b</sup>-positive B cells and PBs. **f** Alt-mRNA/pre-mRNA ratios were determined in IgH<sup>DHLMP2A/frV<sub>K</sub></sup> LPS-stimulated B cells, as described in Fig. 2c. Passive ASO administration was performed as described in panel (b). Data are representative of five independent experiments. Bars represent the mean expression ± SEM. Unpaired two-tailed Student’s *t* test was performed to determine significance (\**P* < 0.05; \*\**P* < 0.01 \*\*\*\**P* < 0.0001)



Technologies). Before hybridization, the probes were denatured in 2× SSC/50% formamide/10% dextran sulfate hybridization buffer at 95 °C for 5 min. Resting B cells, LPS-stimulated B cells, GC-B cells, and PCs were dropped onto poly-L-lysine slides and fixed with PBS/4% paraformaldehyde for 10 min at room temperature (RT). After washing steps with PBS, cells were permeabilized using pepsin 0.02%/HCl 10 mM for 15 min at RT. Cells were washed with PBS, post-fixed with PBS/1% paraformaldehyde for 5 min at RT, washed with PBS/0.2 M MgCl<sub>2</sub>, dehydrated by successive incubations in ethanol solutions (70, 90, 100%, 2 min each), denatured for 5 min in 2× SSC/70% formamide at 72 °C, dehydrated by successive incubations in cold ethanol solutions (70, 90, 100%, 2 min each) and hybridized overnight at 37 °C using 200 ng of each labeled probe in hybridization buffer in a dark humid chamber under coverslips. The next day, the slides were washed three times with 1× SSC, and a blocking step was performed by incubating slides in 4× SSC/3% BSA for 30 min followed by incubation with rhodamine-conjugated anti-digoxigenin secondary antibody (Roche) diluted 1/200 in 4× SSC/3% BSA for 60 min. Slides were washed four times in 2× SSC and mounted under coverslips using ProLong Gold antifade reagent containing DAPI (Life Technologies). Images were acquired along the z-axis with an epifluorescence microscope (Nikon Eclipse Ni-E) with a ×100 oil objective. Fifty-one optical sections separated by 0.2 μm were collected, and stacked images were deconvoluted using Huygens software (SVI) and analyzed using Volocity 3D Image Analysis software (PerkinElmer). Separation of alleles was measured in 3D from the center of mass of each signal. Volumetric pixel size was 0.064 μm in the xy and 0.2 μm in the z-direction.

#### Bioinformatics analysis

Gene expression profiles of purified mouse follicular (FO-) B cells, GC-B cells, PBs, and PCs were obtained from the publicly available dataset [GSE 26408](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26408) (Gene Expression Omnibus database).<sup>44</sup> Affymetrix Mouse Genome 430A 2.0 Array gene expression data were analyzed with the GenomicScape bioinformatics platform ([www.genomicscape.com](http://www.genomicscape.com)).<sup>21</sup> Clustering was performed and visualized with Cluster and TreeView.<sup>45</sup> Differentially expressed genes between cell populations were identified with the significance analysis of microarray statistical method (fold change ≥ 1.5, false discovery rate ≤ 0.05).<sup>46</sup>

#### High throughput sequencing of the IgH repertoire

Repertoire sequencing was performed as previously described.<sup>47</sup> In brief, RNA (500 ng) was extracted from sorted PCs of IgH<sup>DH-LMP2A/frVκ</sup> mice, and RACE-PCR (rapid amplification of cDNA-ends by polymerase chain reaction) was performed using a CH1μ reverse primer.<sup>48</sup> Sequencing adapter sequences were added by primer extension, and amplicons were sequenced on an MiSeq sequencing apparatus (Illumina). Repertoire was analyzed using IMGT/High-V-Quest<sup>49</sup> and the bcRep R package<sup>50</sup> available on the IMGT Web site (<https://www.imgt.org>).

#### Statistical analysis

The results are expressed as the mean ± standard error of the mean (SEM), and overall differences between variables were evaluated by two-tailed unpaired Student's *t* test using Prism GraphPad software (San Diego, CA).

#### ACKNOWLEDGEMENTS

We thank the staff of our animal facility as well as C. Carrion for technical assistance with microscopy and cell cytometry. We also thank K. Rajewsky (The Max Delbrück Center for Molecular Medicine, Berlin, Germany) and S. Casola (Institute of Molecular Oncology Foundation, Milano, Italy) for providing DH-LMP2A mice. We are grateful to N. Diallon (MRC Clinical Sciences Centre, London, UK) for providing the pySat plasmid and M. Alizadeh (UMR S 917, Rennes, France) for repertoire sequencing. This work was supported by grants from Fondation ARC (PJA 20161204724/PGA120150202338),

INCa (PLBIO15-256), ANR (2017-CE15-0024-01), Ligue Contre le Cancer (comités Corrèze, Haute-Vienne), Fondation Française pour la Recherche contre le Myélome et les Gammopathies monoclonales (FFRMG) and Comité d'Organisation de la Recherche sur le Cancer du Limousin (CORC). M.O.A. and N.S. were funded by Société Française d'Hématologie (SFH) and Région Limousin, respectively.

#### AUTHOR CONTRIBUTIONS

M.O.A. and N.S. designed and performed experiments, analyzed data and wrote the paper. J.-M.L. and A.M. contributed data to Fig. 6. O.M., S.L.N., and E.P. contributed 3D DNA-FISH data in Fig. 3. J.M. performed gene expression analysis in Fig. 4. M.V.A., J.S., C.S., and M.C. helped with the experiments and interpretation of the data. L.D. conceived the project, designed experiments, analyzed data and wrote the paper.

#### ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41423-018-0160-6>) contains supplementary material.

**Competing interests:** The authors declare no competing interests.

**Data availability** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### REFERENCES

- Li, S. & Wilkinson, M. F. Nonsense surveillance in lymphocytes? *Immunity* **8**, 135–141 (1998).
- Deiss, T. C. et al. Immunogenetic factors driving formation of ultralong VH CDR3 in *Bos taurus* antibodies. *Cell Mol. Immunol.* <https://doi.org/10.1038/cmi.2017.117> (2017).
- Baumann, B., Potash, M. J. & Köhler, G. Consequences of frameshift mutations at the immunoglobulin heavy chain locus of the mouse. *EMBO J.* **4**, 351–359 (1985).
- Jäck, H. M., Berg, J. & Wabl, M. Translation affects immunoglobulin mRNA stability. *Eur. J. Immunol.* **19**, 843–847 (1989).
- Chemin, G. et al. Multiple RNA surveillance mechanisms cooperate to reduce the amount of nonfunctional Ig kappa transcripts. *J. Immunol.* **184**, 5009–5017 (2010).
- Tinguey, A. et al. Cross talk between immunoglobulin heavy-chain transcription and RNA surveillance during B cell development. *Mol. Cell. Biol.* **32**, 107–117 (2012).
- Valentine, C. R. The association of nonsense codons with exon skipping. *Mutat. Res.* **411**, 87–117 (1998).
- Maquat, L. E. NASTY effects on fibrillin pre-mRNA splicing: another case of ESE does it, but proposals for translation-dependent splice site choice live on. *Genes Dev.* **16**, 1743–1753 (2002).
- Chang, Y.-F., Chan, W.-K., Imam, J. S. & Wilkinson, M. F. Alternatively spliced T-cell receptor transcripts are up-regulated in response to disruption of either splicing elements or reading frame. *J. Biol. Chem.* **282**, 29738–29747 (2007).
- Srouf, N. et al. A plasma cell differentiation quality control ablates B cell clones with biallelic Ig rearrangements and truncated Ig production. *J. Exp. Med.* **213**, 109–122 (2016).
- Delpy, L., Le Bert, M., Cogné, M. & Khamlichi, A. A. Germ-line transcription occurs on both the functional and the non-functional alleles of immunoglobulin constant heavy chain genes. *Eur. J. Immunol.* **33**, 2108–2113 (2003).
- Reynaud, S. et al. Interallelic class switch recombination contributes significantly to class switching in mouse B cells. *J. Immunol.* **174**, 6176–6183 (2005).
- Casola, S. et al. B cell receptor signal strength determines B cell fate. *Nat. Immunol.* **5**, 317–327 (2004).
- Lechouane, F. et al. B-cell receptor signal strength influences terminal differentiation. *Eur. J. Immunol.* **43**, 619–628 (2013).
- Wesemann, D. R. et al. Microbial colonization influences early B-lineage development in the gut lamina propria. *Nature* **501**, 112–115 (2013).
- Aouinti, S. et al. IMGT/StatClonotype for pairwise evaluation and visualization of NGS IG and TR IMGT clonotype (AA) diversity or expression from IMGT/HighV-QUEST. *Front. Immunol.* **7**, 339 (2016).
- Pan, Q. et al. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.* **20**, 153–158 (2006).
- de la Mata, M. et al. RNA polymerase II elongation at the crossroads of transcription and alternative splicing. *Genet. Res. Int.* **2011**, 309865 (2011).



19. Schor, I. E., Rascovan, N., Pelisch, F., Alló, M. & Kornblihtt, A. R. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl Acad. Sci. USA* **106**, 4325–4330 (2009).
20. Park, S.-K., Xiang, Y., Feng, X. & Garrard, W. T. Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. *Genes Dev.* **28**, 1159–1164 (2014).
21. Kassambara, A. et al. GenomicScape: an easy-to-use web tool for gene expression data analysis. Application to investigate the molecular events in the differentiation of B cells into plasma cells. *PLoS Comput. Biol.* **11**, e1004077 (2015).
22. Kolasinska-Zwiercz, P. et al. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41**, 376–381 (2009).
23. Meda, F., Folci, M., Baccarelli, A. & Selmi, C. The epigenetics of autoimmunity. *Cell. Mol. Immunol.* **8**, 226–236 (2011).
24. Luco, R. F. et al. Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
25. Saint-André, V., Batsché, E., Rachez, C. & Muchardt, C. Histone H3 lysine 9 trimethylation and HP1 $\gamma$  favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* **18**, 337–344 (2011).
26. Zhou, H.-L., Luo, G., Wise, J. A. & Lou, H. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res.* **42**, 701–713 (2014).
27. Benchaouir, R., Robin, V. & Goyenvall, A. Gene and splicing therapies for neuromuscular diseases. *Front. Biosci. Landmark Ed.* **20**, 1190–1233 (2015).
28. Groen, E. J. N., Talbot, K., Gillingwater, T. H. Advances in therapy for spinal muscular atrophy: promises and challenges. *Nat. Rev. Neurol.* <https://doi.org/10.1038/nrneurol.2018.4> (2018).
29. Stein, C. A. & Castanotto, D. FDA-approved oligonucleotide therapies in 2017. *Mol. Ther. J. Am. Soc. Gene Ther.* **25**, 1069–1075 (2017).
30. Mostoslavsky, R., Alt, F. W. & Rajewsky, K. The lingering enigma of the allelic exclusion mechanism. *Cell* **118**, 539–544 (2004).
31. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126**, 37–47 (2006).
32. Johnson, K., Angelin-Duclos, C., Park, S. & Calame, K. L. Changes in histone acetylation are associated with differences in accessibility of V(H) gene segments to V-DJ recombination during B-cell ontogeny and development. *Mol. Cell. Biol.* **23**, 2438–2450 (2003).
33. Wang, E. T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
34. Aartsma-Rus, A. & van Ommen, G.-J. B. Progress in therapeutic antisense applications for neuromuscular disorders. *Eur. J. Hum. Genet.* **18**, 146–153 (2010).
35. Daly, J., Licence, S., Nanou, A., Morgan, G. & Mårtensson, I.-L. Transcription of productive and nonproductive VDJ-recombined alleles after IgH allelic exclusion. *EMBO J.* **26**, 4273–4282 (2007).
36. Delpy, L., Sirac, C., Le Morvan, C. & Cogné, M. Transcription-dependent somatic hypermutation occurs at similar levels on functional and nonfunctional rearranged IgH alleles. *J. Immunol.* **173**, 1842–1848 (2004).
37. Holwerda, S. J. B. et al. Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. *Nucleic Acids Res.* **41**, 6905–6916 (2013).
38. Nogués, G. et al. Control of alternative pre-mRNA splicing by RNA Pol II elongation: faster is not always better. *IUBMB Life* **55**, 235–241 (2003).
39. Pandit, S., Wang, D. & Fu, X.-D. Functional integration of transcriptional and RNA processing machineries. *Curr. Opin. Cell Biol.* **20**, 260–265 (2008).
40. Shukla, S. & Oberdoerffer, S. Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochim. Biophys. Acta* **1819**, 673–683 (2012).
41. Pinaud, E. et al. Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* **15**, 187–199 (2001).
42. Sirac, C. et al. Role of the monoclonal kappa chain V domain and reversibility of renal damage in a transgenic model of acquired Fanconi syndrome. *Blood* **108**, 536–543 (2006).
43. Rouaud, P. et al. The IgH 3' regulatory region controls somatic hypermutation in germinal center B cells. *J. Exp. Med.* **210**, 1501–1507 (2013).
44. Green, M. R. et al. Signatures of murine B-cell development implicate Yy1 as a regulator of the germinal center-specific program. *Proc. Natl Acad. Sci. USA* **108**, 2873–2878 (2011).
45. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
46. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
47. Li, S. et al. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat. Commun.* **4**, 2333 (2013).
48. Boice, M. et al. Loss of the HVEM tumor suppressor in lymphoma and restoration by modified CAR-T cells. *Cell* **167**, 405–418.e13 (2016).
49. Alamyar, E., Duroux, P., Lefranc, M.-P. & Giudicelli, V. IMGT(\*) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* **882**, 569–604 (2012).
50. Bischof, J. & Ibrahim, S. M. bcRep: R package for comprehensive analysis of B cell receptor repertoire data. *PLoS ONE* **11**, e0161569 (2016).



**Ashi et al.**

### **Supplementary Information**

Supplementary Fig. 1 shows VDJ repertoire analysis of IgH<sup>frV $\kappa$</sup>  alt-mRNAs. Supplementary Fig. 2 shows absence of adverse toxic effect following ASO treatment. Supplementary Table 1, a separate excel file, shows SAM Multiclass analysis of genes related to RNA splicing during plasma cell differentiation. Supplementary Table 2 contains list of primers and probes used in CHIP, standard and real time PCR assays.

### **LEGENDS OF SUPPLEMENTARY INFORMATION**

#### **Supplementary Figure 1: VDJ repertoire analysis of IgH<sup>frV $\kappa$</sup> alt-mRNAs**

PCs (B220<sup>-</sup>/CD138<sup>+</sup>) were sorted from spleens of IgH<sup>DH-LMP2A/frV $\kappa$</sup>  mice and RACE-PCR was performed using a CH1 $\mu$  reverse primer as described in methods. The VDJ repertoire of IgH<sup>frV $\kappa$</sup>  alt-mRNAs was analyzed using IMGT/High-V-Quest (Alamyar et al., 2012). **a, c** Pie charts represent VH and JH<sub>1-4</sub> family usage, respectively. The total number of clonotypes identified (**a**) and the frequency of JH usage (**c**) are indicated. **b** Histograms of DH family usage in IgH<sup>DH-LMP2A/frV $\kappa$</sup>  mice.

#### **Supplementary Figure 2: Absence of adverse toxic effect following ASO treatment**

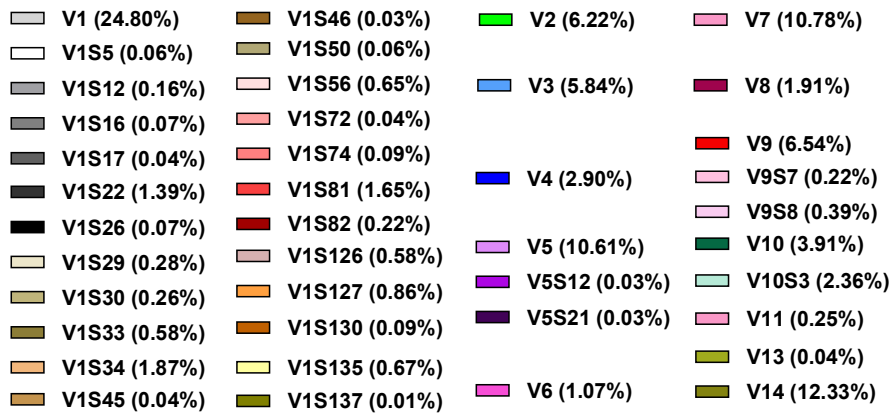
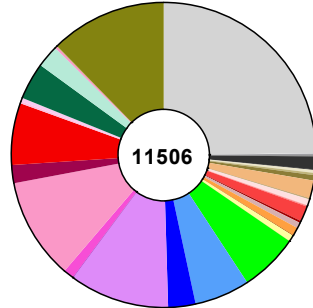
Splenic B cells (n=5) isolated from IgH<sup>wt/frV $\kappa$</sup>  mice were stimulated with LPS (1 $\mu$ g/ml) for 4 days and treated or not with AON (1 $\mu$ M) during the last 2 days, as described in methods. **a** The proliferative index was determined by counting viable cells after Trypan blue exclusion. **b** The production of Ig $\kappa$  was analyzed in 96h supernatants to evaluate off-target effects of ASO treatments in antibody secreting cells. Bar graphs are expressed as mean  $\pm$  SEM (ns: not significant).



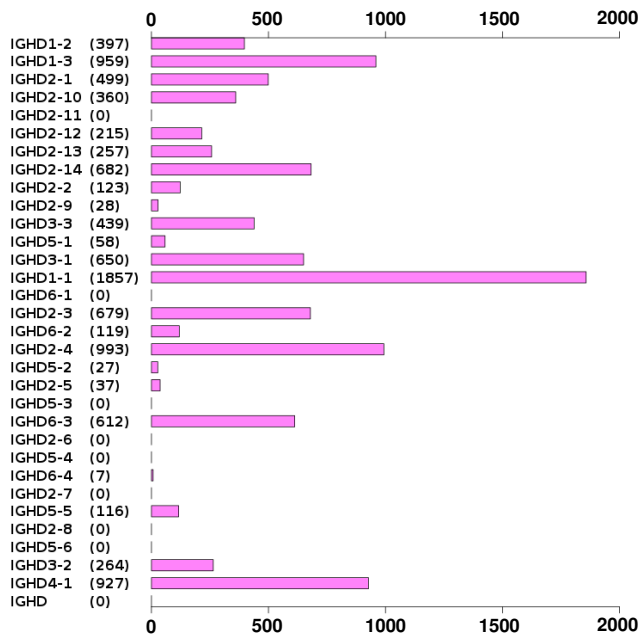


Supplementary Information

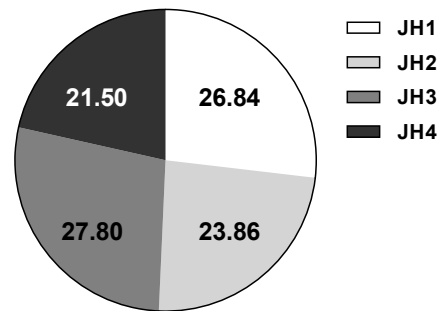
a



b



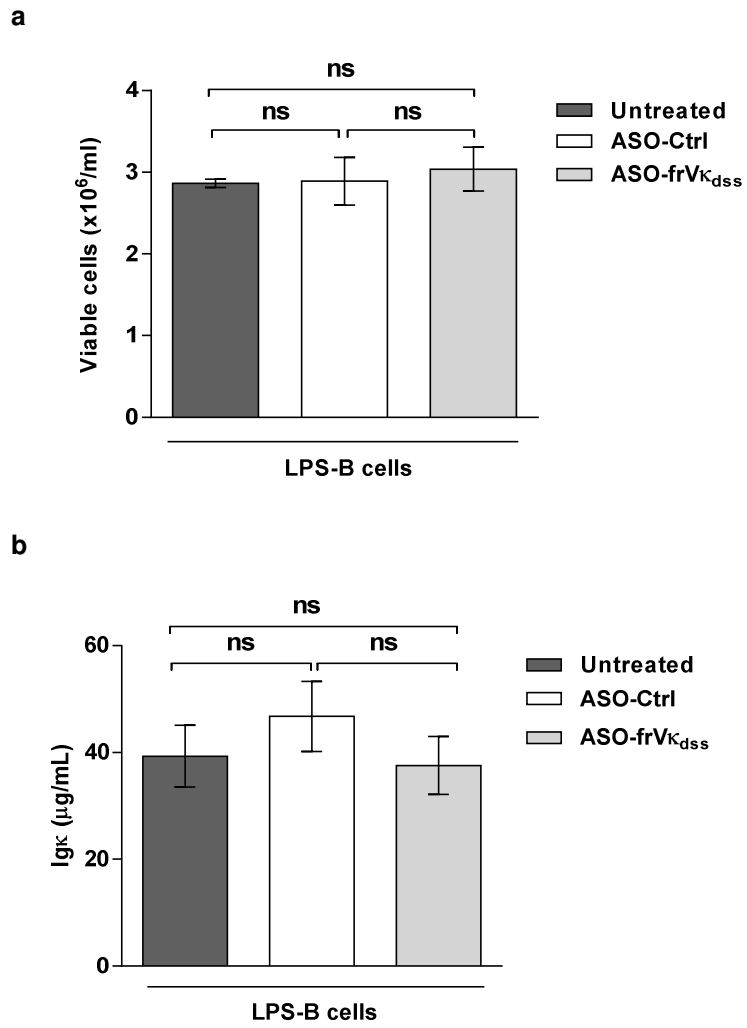
c



Supplementary Figure 1 (Ashi et al.)



Supplementary Information



Supplementary Figure 2 (Ashi et al.)



## Supplementary Information

**Supplementary Table 2. Primers and probes used in CHIP, standard and real time PCR assays**

Primers and probes		Standard	Real time	CHIP	Letter code <sup>a</sup>
<b>VH<sub>7183</sub>-for</b>	5' CGG TAC CAA GAA SAM CCT GTW CCT GCA AAT GAS C 3'	x	x		
<b>C<math>\mu</math>-rev</b>	5' AGA CGA GGG GGA AGA CAT TT 3'	x	x	x	
<b>IgH<sup>frVk</sup>-for</b>	5' GGG CAC CAA GCT GGA AAT C 3'		x	x	a
<b>IgH<sup>frVk</sup>-rev</b>	5' CGC CCC CAC TAG ATG CAT 3'		x	x	a
<b>IgH<sup>frVk</sup>-probe</b>	5' AAC GGG TAA GTC TCG 3'		x	x	a
<b>IgH<sup>wt</sup>-for</b>	5' TTC TGA GCA TTG CAG ACT AAT CTT G 3'		x	x	b
<b>IgH<sup>wt</sup>-rev</b>	5' CCT AGA CAG TTT ATT TCC CAA CTT CTC 3'		x	x	b
<b>IgH<sup>wt</sup>-probe</b>	5' CCC TGA GGG AGC CG 3'		x	x	b
<b>CH1<math>\mu</math>-for</b>	5' CTT CCC AAA TGT CTT CCC CC 3'			x	c
<b>CH1<math>\mu</math>-rev</b>	5' TGC GAG GTG GCT AGG TAC TTG 3'			x	c
<b>Gapdh Ex5-for</b>	5' TGT TCC TAC CCC CAA TGT GT 3'	x			
<b>Gapdh Ex6-rev</b>	5' GGT CCT CAG TGT AGC CCA AG 3'	x			
<b>Gapdh-probe</b>	Mm99999915_g1, Applied Biosystem		x		

<sup>a</sup> Letter code used in Fig. 1 A and Fig. 5 A. Degenerate nucleotide positions are coded as follows: M=A or C, S=C or G, Y=C, W=A or T.



**Article 2 : Radiolabelled polymeric IgA: from distribution to a new molecular imaging tool in colorectal cancer lung metastases.**

Helene Carpenet, Armelle Cuvillier, Aurélie Perraud, Ophélie Martin, Gaël Champier, Marie-Odile Jauberteau, Jacques Monteil and Isabelle Quelven.

Oncotarget, 2017, Vol. 8, (No. 49), pp: 85185-85202.





## Radiolabelled polymeric IgA: from biodistribution to a new molecular imaging tool in colorectal cancer lung metastases

Helene Carpenet<sup>1,2</sup>, Armelle Cuvillier<sup>3</sup>, Aurélie Perraud<sup>2</sup>, Ophélie Martin<sup>4</sup>, Gaël Champier<sup>3</sup>, Marie-Odile Jauberteau<sup>2</sup>, Jacques Monteil<sup>1,2,4</sup> and Isabelle Quelven<sup>1,2,4</sup>

<sup>1</sup>Nuclear Medicine Department, Dupuytren University Hospital, 87042 Limoges, France

<sup>2</sup>EA 3842 – Cellular Homeostasis and Diseases, Faculty of Medicine, University of Limoges, 87025 Limoges, France

<sup>3</sup>B Cell Design SAS, 87000 Limoges, France

<sup>4</sup>UMR CNRS 7276 – CRIBL, University of Limoges, 87025 Limoges, France

**Correspondence to:** Isabelle Quelven, **email:** isabelle.quelven@unilim.fr  
Armelle Cuvillier, **email:** armelle.cuvillier@b-cell-design.com

**Keywords:** anti-CEA IgA, radiolabelling, mucosal biodistribution, metastases, colorectal cancer imaging

**Received:** December 21, 2016

**Accepted:** July 03, 2017

**Published:** July 27, 2017

**Copyright:** Carpenet et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### ABSTRACT

**By radiolabelling monomeric (m) and polymeric (p) IgA with technetium 99m (<sup>99m</sup>Tc), this study assessed IgA biodistribution and tumour-targeting potency. IgA directed against carcinoembryonic antigen (CEA), a colorectal cancer marker, was selected to involve IgA mucosal tropism.**

**Ig was radiolabelled with <sup>99m</sup>Tc-tricarbonyl after derivatisation by 2-iminothiolane. <sup>99m</sup>Tc-IgA was evaluated by *in vitro* analysis. The biodistributions of radiolabelled anti-CEA mIgA, pIgA and IgG were compared in normal mice. Anti-CEA pIgA tumour uptake was studied in mice bearing the WiDr caecal orthotopic graft.**

**IgA radiolabelling was obtained with a high yield, was stable in PBS and murine plasma, and did not alter IgA binding functionality ( $K_d \approx 25$  nM). Biodistribution studies in normal mice confirmed that radiolabelled pIgA – and to a lesser extent, mIgA – showed strong and fast mucosal tropism and a shorter serum half-life than IgG. In caecal tumour model mice, evaluation of the anti-CEA-pIgA biodistribution showed a high uptake in lung metastases, confirmed by histological analysis. However, no radioactivity uptake increase in the tumoural caecum was discerned from normal intestinal tissue, probably due to high IgA caecal natural tropism. In microSPECT/CT imaging, <sup>99m</sup>Tc-IgA confirmed its diagnostic potency of tumour in mucosal tissue, even if detection threshold by *in vivo* imaging was higher than *post mortem* studies. Contribution of the FcαRI receptor, studied with transgenic mouse model (Tsg SCID-CD89), did not appear to be determinant in <sup>99m</sup>Tc-IgA uptake.**

**Pre-clinical experiments highlighted significant differences between <sup>99m</sup>Tc-IgA and <sup>99m</sup>Tc-IgG biodistributions. Furthermore, tumoural model studies suggested potential targeting potency of pIgA in mucosal tissues.**

### INTRODUCTION

Colorectal cancer (CRC) is the most commonly diagnosed cancer worldwide. For the past few years, almost 1.4 million new cases have been diagnosed every year (representing 9.7% of all cancers), with 450,000 new cases being reported in Europe (where CRC is the second-

most common cancer, representing 13% of all cancers). CRC is also the second-most common cause of cancer-related deaths in Europe, with an estimated 215,000 deaths every year. Cancer survival rates are based on initial staging of the cancer at diagnosis. Optical colonoscopy remains the gold-standard investigation for early detection of CRC. Accurate staging remains critical for



determining the most relevant treatment, and also offers prognostic and outcome indicators. However, despite the above, according to a review by Bradley no definitive consensus on the optimal imaging staging strategy has yet been established for CRC [1]. More specific and sensitive non-invasive methods are needed. One such method is immunoscintigraphy, a useful technique that has continued to expand since the 1980s commensurate with monoclonal expansion. Monoclonal antibody (mAb) generation in 1975 by Köhler and Milstein combined to nuclear imaging technology progress led to development of monoclonal antibodies labelled with radioactive isotopes to target tumours [2]. Concerning staging of CRC, the main interest in immunoscintigraphy lies in its sensitivity for detecting distant metastasis.

Many antibodies or antibody fragments have been developed for detecting primary tumours and distant metastasis. In cases of cancer with mucosal localisation, satumomab (Oncoscint<sup>®</sup>) directed against glycoprotein TAG-72 radiolabelled with 111-Indium has been used in the US since 1992 for CRC diagnosis [3]. Arcitumomab (CEA-Scan<sup>®</sup>), a F(ab')<sub>2</sub> directed against carcinoembryonic antigen (CEA) labelled with 99m-Technetium (<sup>99m</sup>Tc), received FDA and EMEA approval in 1996 for diagnosis of CRC recurrence [4, 5]. Another <sup>99m</sup>Tc-radiolabelled Fab, nofetumomab (Verluma<sup>®</sup>) – directed against pancarcinoma glycoprotein antigen epithelial cell adhesion molecule (EpCAM) – was FDA-approved for the diagnosis of small cell lung cancer. Today, satumomab is the only mAb still used in the US for the immunoscintigraphy of mucosal tumour localisation. Their diagnostic value remains controversial due to the low contrast between normal and tumoural tissue, especially compared with 18-fluoro-deoxyglucose positron emission tomography (FDG-PET) [6]. However, FDG is not a specific tracer as high FDG uptake in inflammatory or infectious lesions, and variable physiological concentrations in normal tissues/organs, can be confused with malignant tissue [7]. An approach to overcome the limitations of FDG-PET involves targeting molecules to localise specific tumour markers, with preferential tropism for the tissues of interest, such as mucous membranes.

Antibodies directed against CEA appear to be good candidates to target CRC. Indeed, CEA, a 180-kDa glycoprotein, is part of a large family of 12 related members, whose functions can vary widely and include cell adhesion, participation in intracellular signal cascades, cell migration, inflammation, and angiogenesis. CEA is synthesised mainly by the digestive tract and is polarised at the apical pole in the epithelial wall. This oncofetal antigen is quite absent from healthy adult tissues. CEA is frequently expressed, with a high serum concentration, in adenocarcinomas of various origins, such as breast, medullary thyroid, lung, and ovarian cancers, as well as CRC. CEA was identified in CRC for the first time in 1965 [8]. Its expression is upregulated in approximately 90% of advanced CRC cases [9]. One in two patients diagnosed

with CRC already has metastasis, thereby negating the use of a first-line surgical strategy; furthermore, many cases do not respond to chemotherapy. In such advanced cases, the prognosis is very poor, and the average survival time is 5 to 9 months. Although not a specific marker of CRC, CEA has prognostic value in patients with metastatic cancer. It is elevated in 58% to 65% of cases, and the first sign of recurrence can precede clinical or radiological recidivism by 1.5 to 6 months. Several clinical studies are underway to assess the relevance of targeting CEA by immunoscintigraphy for the detection of primary tumour and metastasis [10, 11]. Nevertheless, published data are only based on the Immunoglobulin (Ig) G or IgG fragment, and no study has used IgA. In humans, IgA is the most heavily produced isotype (66 mg/kg/day) and the second-most prevalent circulating isotype, after IgG. Long regarded as an anti-inflammatory antibody involved in maintaining homeostatic balance at the level of the mucous membrane, it has been demonstrated recently that IgA can enable or inhibit different inflammatory responses [12]. IgA efficacy to recruit effector cells, such as monocytes, macrophages, granulocytes and Kupffer cells, was demonstrated for antibody dependent cell-mediated cytotoxicity (ADCC) [13]. IgA interaction with myeloid cells is mediated with the IgA Fc receptor, FC $\alpha$ RI (CD89). Importantly, mice lacking to express any human FC $\alpha$ RI homologue, a transgenic mouse model (Tsg SCID-CD89) was described to express human FC $\alpha$ RI on mouse myeloid lineage [14] and mimics effector cells recruitment and cell cytotoxicity mechanisms observed in human [15, 16] in presence of human or humanised IgA.

IgA is expressed in three different forms: monomeric (mIgA) (in the blood: 1 to 3 g/L), dimeric (dIgA)/polymeric (pIgA) (in mucous membranes) and secretory (SIgA) (in the mucosal organs). The B lymphocytes are present in the serous compartment of mucosal organs and predominantly produce dIgA (two molecules of IgA complexed by the J chain peptide). The dimers, specifically recognised by the polymeric Ig receptor (pIgR) expressed on the basal side of epithelial cells, are endocytosed and translocated to the apical side of the cells. At the time of exocytosis, the extracellular part of the pIgR is cleaved and remains complexed to the dimer, thereby forming the secretory component, the SIgA [12]. Moreover, dIgA has a cell signalling capacity greater than that of monomeric IgA because the dimer carries twice as many paratopes [16]. Knowledge of IgA and its applications is limited partly due to difficulties in the identification of IgA-producing B cells, and with respect to stable production of IgA antibodies. IgA-secreting B lymphocytes represent less than 1% of normal mouse splenocytes (even fewer are found in mucosal lymphoid compartments: 0.01% in the *lamina propria* and 0.1% in Peyer's patches) [17]. The recently developed HAMIGA<sup>™</sup> technology allows this limitation to be bypassed [18]. By replacing the S $\mu$  domain with a human alpha 1 constant gene downstream of variable gene segments, the population of IgA-secreting



lymphocytes B in the spleen was increased significantly (62% of B220<sup>+</sup> membrane IgA cells in the spleen of HAMIGA<sup>TM</sup> mice versus undetectable level in wild-type mice) [18] which allows to easily sort highly specific monoclonal humanised IgA. From immunised transgenic human  $\alpha 1$  mice, IgA1 were produced in hybridoma, B lymphocyte hybrid cells, using its dedicated glycosylation pathways of antibodies.

Concerning IgA application in nuclear medicine, <sup>99m</sup>Tc remains the most widely used isotope for diagnosis, due to its suitable nuclear and chemical properties, and good availability [19]. This radionuclide has favourable physical characteristics for high-efficiency detection in molecular imaging (with a  $\gamma$ -ray of 140 keV) and for radiation protection due to a short half-life ( $T_{1/2} = 6$ h). This short half-life is not optimal for IgG (serum half-life 20 days) radiolabelling but seems suitable considering the shorter half-life of IgA (3–6 days). In a previous study, we showed that indirect radiolabelling using a bifunctional chelating agent, such as the tricarbonyl core  $[\text{Tc}(\text{CO})_3(\text{H}_2\text{O})_3]^+$ , and a spacer such as 2-iminothiolane (2-IT), provides good radiolabelling yields with high specific activity, and preserves IgG structure and immunoaffinity with limited amounts of antibody [20].

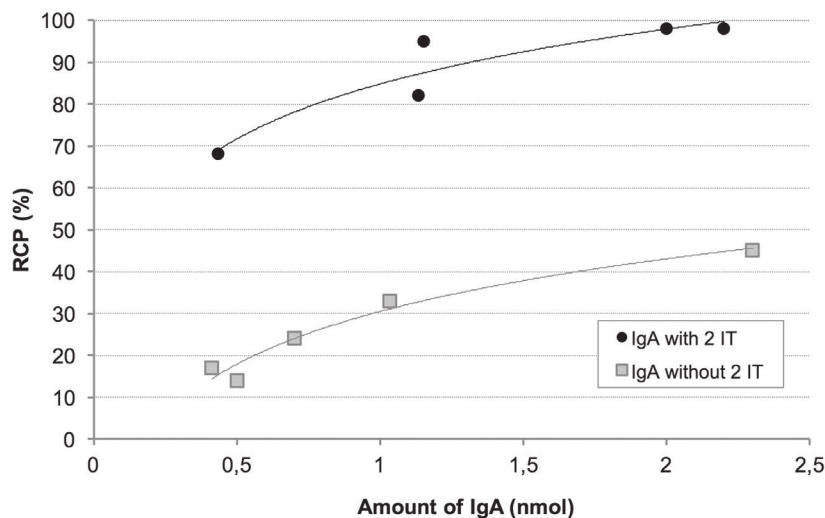
Here, by radiolabelling with <sup>99m</sup>Tc, we report the natural pharmacokinetic and biodistribution of different isotypes of antibodies, IgA directed against CEA, compared with IgG in healthy mice and in a xenograft mouse model of human colon carcinoma. This model enables us to evaluate the mucosal tropism of IgA antibodies and to test the ability of IgA antibodies, directed against a well-characterised human tumour-associated antigen, to detect early-stage mucosal metastasis foci.

Biodistribution was firstly evaluated by *post mortem* organs counting. This most sensitive technique is the gold standard to evaluate accurately molecule biodistribution, but needs animals sacrifice. MicroSPECT/CT studies, which are less sensitive but allow *in vivo* monitoring, were compared to gold standard method. Using the transgenic mouse model expressing the high affinity receptor FC $\alpha$ RI (Tsg SCID-CD89), we also evaluate the contribution of FC $\alpha$ RI in the distribution and the capacity of IgA to detect tumour *in vivo* by microSPECT/CT imaging.

## RESULTS

### IgA radiolabelling and <sup>99m</sup>Tc-IgA-SH *in vitro* stability

To optimise IgA radiolabelling with the tricarbonyl core, increasing quantities of antibody were tested by derivatisation with 2-IT (Figure 1). Native and functionalised IgA (IgA-SH) radiolabelling yields were correlated with the IgA amounts. The maximum quantity usable prior to IgA precipitation (maximum concentration = 2 mg/mL) is 2.2 nmol of IgA. With this amount of native IgA, only 45% radiochemical purity (RCP) was reached. With derivatised IgA, much higher yields were achieved (RCP = 98%) for this maximum IgA quantity, and an RCP > 85% was obtained from only 1 nmol of IgA. These results confirmed that the derivatisation of IgA is necessary to obtain high radiolabelling of RCP. After determining the sulphhydryl groups by using the micromethod and Ellman's reagent, a correlation between the radiolabelling yields and number of thiol moieties on



**Figure 1: Comparison of native and 2-IT derivatised radiolabelled IgA yields according to IgA amount.** Radiochemical purity (RCP) analysis was performed using instant thin layer chromatography–silica gel (ITLC-SG)/normal saline.



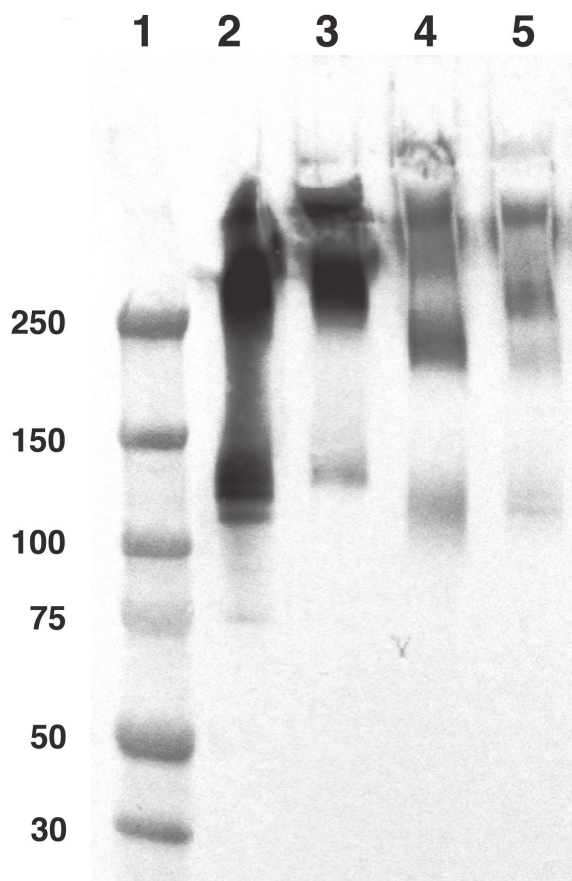
IgA was clearly seen, confirming derivatisation (data not shown). With derivatised IgA, on average,  $1.7 \pm 0.4$  ( $n = 3$ ) sulphhydryl groups were grafted to each IgA molecule; with native IgA, the number was below the limit of quantification. With these optimised conditions (1.4 nmol of derivatised IgA), radiolabelling yields  $> 96.6 \pm 2.1\%$  ( $n = 13$ ) were obtained. The specific activity corresponded to 115 MBq/nmol (511 MBq/mg).

Stability testing of  $^{99m}\text{Tc}$ -IgA-SH was performed over 24 h in phosphate-buffered saline (PBS), with storage at 4°C or 25°C, and in murine serum (1/10 dilution) at 37°C. In PBS, no appreciable loss in RCP was seen (at 24 h:  $\Delta\text{RCP} = 4\%$  and  $8\%$ , at 4°C and 25°C, respectively). However, after incubation in murine serum (1/10 dilution) at 37°C, a gradual decrease in radiolabelled IgA RCP was observed (at 18 h and 24 h:  $\Delta\text{RCP} = 13\%$  and  $23\%$ , respectively).

### Integrity of $^{99m}\text{Tc}$ -IgA-SH

To demonstrate full preservation of molecular integrity during the radiolabelling process, unlabelled and radiolabelled IgA samples of monomeric and polymeric mixture (mpIgA) were analysed by Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 2).

Western blot profiles of mpIgA showed major bands at 120–140 kDa and 300–400 kDa, corresponding to the molecular weight of monomeric and polymeric forms, respectively, for both unlabelled IgA (lanes 2–3) and radiolabelled IgA (lanes 4–5). As expected, the molecular structure of IgA was preserved after radiolabelling; no low-molecular-weight band corresponding to free heavy chain release appeared for either radiolabelled antibody. A slight decrease in intensity was observed between



**Figure 2: Molecular structure comparison of unlabelled and radiolabelled anti-carcinoembryonic antigen (CEA) or irrelevant monomeric and polymeric mixture (mpIgA).** Anti-CEA or irrelevant mpIgA molecular structure was evaluated by Western blotting. Molecular weights in kDa are specified on the left. 1. Molecular weight Marker, 2. Unlabelled anti-CEA mpIgA-SH (monomeric  $\approx 150$  kDa and polymeric  $> 250$  kDa) (1.5 mg), 3. Unlabelled irrelevant mpIgA-SH (1.5 mg), 4. Radiolabelled anti-CEA mpIgA (1.5 mg), 5. Radiolabelled irrelevant mpIgA-SH (1.5 mg).





unlabelled and radiolabelled antibodies due to Ig loss during the purification after incubation with 2-IT.

### Affinity of $^{99m}\text{Tc}$ -IgA-SH

Immunoaffinity evaluation by radioimmunoassay (RIA) allowed us to observe that the radiolabelled anti-CEA IgA-SH has a high affinity for the target antigen (Figure 3A). The specific binding of  $^{99m}\text{Tc}$ -anti-CEA IgA-SH was confirmed by displacement with unlabelled anti-CEA IgA. The dissociation constant (Kd) and maximum number of antibody bound of  $^{99m}\text{Tc}$ -anti-CEA IgA-SH were estimated using a saturation binding assay and Scatchard analysis (Figure 3C). The assays provided a value of 20.3 nM for the Kd and 3.2 nM for the binding capacities ( $B_{\max}$ ). Specific binding was also confirmed by lack of fixation of an irrelevant IgA ( $^{99m}\text{Tc}$ -anti-peanut IgA-SH).

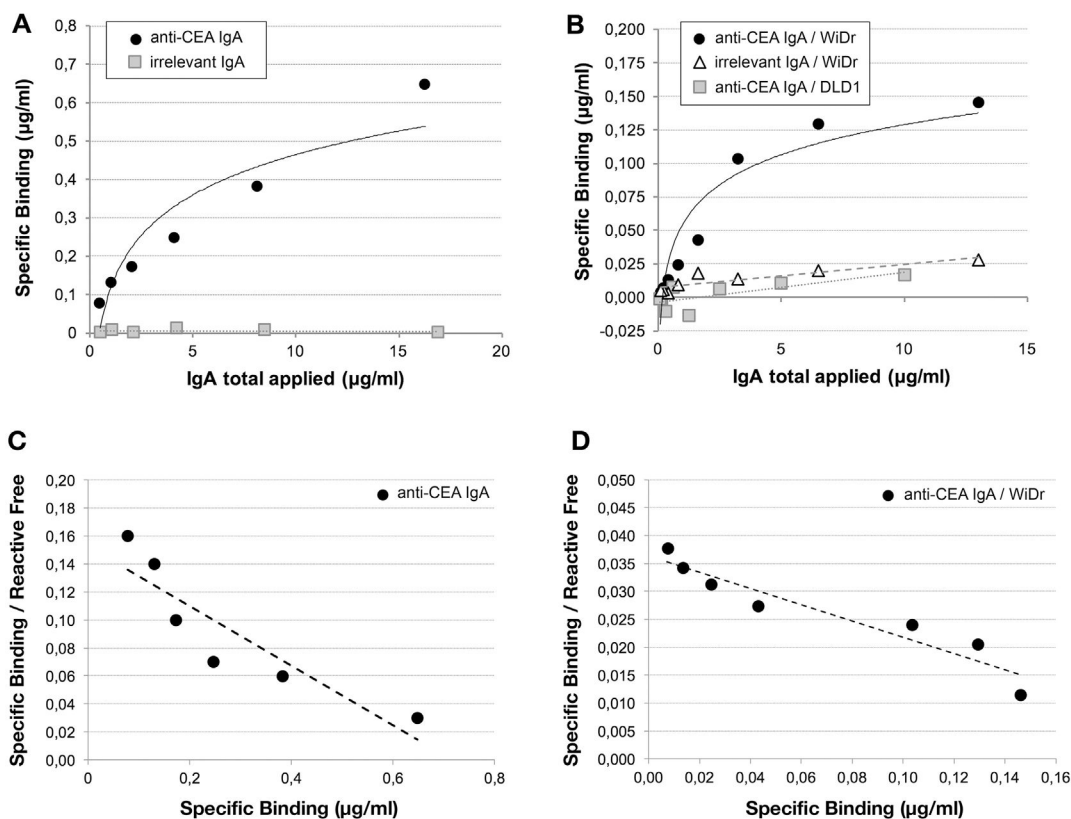
The immunoaffinity of  $^{99m}\text{Tc}$ -anti-CEA IgA-SH was also assessed using a saturation binding assay on the WiDr colon adenocarcinoma cell line, which expresses the fully folded membranal antigen (Figure 3B). This assay showed specific saturable binding. Using the Scatchard representation data, the calculated Kd and  $B_{\max}$  were 30.5 nM and 1.2 nM, respectively (Figure 3D). The controls were performed with  $^{99m}\text{Tc}$ -anti-CEA IgA-

SH on non-expressing CEA cells (DLD1; another colon adenocarcinoma cell line) and with irrelevant IgA on WiDr cells. As expected, no binding was observed under either of these conditions.

### Biodistribution of $^{99m}\text{Tc}$ -anti-CEA IgA-SH and $^{99m}\text{Tc}$ -anti-CEA IgG-SH in normal mice

Comparison of the *in vivo* biodistribution kinetics of  $^{99m}\text{Tc}$ -anti-CEA polymeric IgA-SH,  $^{99m}\text{Tc}$ -anti-CEA monomeric IgA-SH and  $^{99m}\text{Tc}$ -anti-CEA IgG-SH in healthy BALB/c mice are presented for all studied organs in Tables 1–3, respectively, and for selected organs in Figure 4.

Global analysis (i.e. whole-organ study) of radioactivity uptake (percentage of the injected dose per gram of tissue: %ID/g) showed a significant ( $p < 0.002$ ) difference among the three Ig types at 8 h, 24 h and 48 h. However, although no significant difference was observed in global analysis at 4 h among the three Ig types, an important difference in the radioactivity uptake was observed between the caecum and faeces. At this early time, high uptake was measured, in the caecum and faeces, of  $^{99m}\text{Tc}$ -anti-CEA pIgA-SH ( $27.0 \pm 4.9$  and  $30.6 \pm 9.3$  %ID/g, respectively) and, to a lesser



**Figure 3:** *In vitro* binding of  $^{99m}\text{Tc}$ -anti-CEA IgA-SH and  $^{99m}\text{Tc}$ -irrelevant IgA-SH. Radiolabelled IgA specific binding was evaluated by RIA protocol (A) and on cells (B). For cells studies, WiDr (CEA pos) and DLD1 (CEA neg) cell lines were tested. For the two experiments, Kd and  $B_{\max}$  were evaluated by a Scatchard plot (C: RIA; D: cells).



**Table 1: Biodistribution of <sup>99m</sup>Tc-anti-CEA IgA polymeric-SH in healthy Balb-c mice at 4, 8, 18, 24 and 48h, expressed as the percentage of injected (intravenous; IV) dose per gram, %ID/g (values represent means ± S.D. of the %ID/g)**

	<b>4h (n=4)</b>	<b>8h (n=4)</b>	<b>18h (n=4)</b>	<b>24h (n=8)</b>	<b>48h (n=4)</b>
<b>Heart</b>	5.4±2.5	3.0±0.3	2.4±0.1	1.9±0.8	1.2±0.4
<b>Brain</b>	0.33±0.09	0.22±0.04	0.10±0.03	0.11±0.04	0.05±0.01
<b>Lungs</b>	7.8±2.7	5.2±0.7	3.4±1.1	3.3±1.1	2.3±1.0
<b>Kidneys</b>	21.4±6.9	16.0±5.1	11.6±6.1	11.0±6.6	7.9±4.6
<b>Spleen</b>	16.5±6.0	12.2±3.5	6.9±2.7	6.5±1.1	3.9±1.7
<b>Stomach</b>	2.7±1.5	1.1±0.4	1.5±1.3	0.7±0.3	0.5±0.2
<b>Small intestine</b>	5.2±1.0	3.2±0.6	1.5±0.7	1.1±0.4	0.6±0.2
<b>Caecum</b>	27.0±4.9	11.2±4.5	2.8±1.7	3.2±2.1	1.4±0.7
<b>Colon</b>	6.7±2.0	3.4±1.3	1.5±0.8	1.1±0.4	0.7±0.4
<b>Liver</b>	66.4±19.4	49.0±18.4	19.2±3.9	24.4±11.0	16.9±7.6
<b>Muscle</b>	1.6±0.5	1.4±0.6	0.9±0.5	0.6±0.2	0.5±0.3
<b>Blood cells</b>	3.3±0.8	2.0±0.5	4.6±2.0	2.1±1.2	1.3±0.5
<b>Plasma</b>	19.2±4.3	10.0±1.3	1.3±0.5	2.8±1.9	0.9±0.4
<b>Faeces</b>	30.6±9.3	6.9±4.4	3.9±2.6	2.1±1.2	0.9±0.6
<b>Mesenteric Ganglion</b>	4.9±3.1	3.0±1.6	1.05±0.3	1.8±1.7	0.5±0.2

**Table 2: Biodistribution of <sup>99m</sup>Tc-anti-CEA IgA monomeric-SH in healthy Balb-c mice at 4, 8, 18, 24 and 48h, expressed as the percentage of injected (IV) dose per gram, %ID/g (values represent means ± S.D. of the %ID/g)**

	<b>4h (n=4)</b>	<b>8h (n=4)</b>	<b>18h (n=3)</b>	<b>24h (n=7)</b>	<b>48h (n=4)</b>
<b>Heart</b>	6.0±0.8	3.4±0.3	3.4±1.4	1.7±0.8	1.6±0.3
<b>Brain</b>	0.44±0.06	0.28±0.04	0.17±0.11	0.12±0.04	0.07±0.02
<b>Lungs</b>	7.8±1.6	5.3±0.7	5.8±3.0	3.4±1.2	3.0±1.3
<b>Kidneys</b>	20.0±4.0	15.8±6.0	16.2±9.2	12.1±6.9	9.8±3.5
<b>Spleen</b>	27.2±4.0	18.6±6.5	52.6±14.1	34.8±19.5	41.4±8.7
<b>Stomach</b>	2.7±1.7	0.9±0.4	0.9±0.7	1.0±0.7	0.8±0.5
<b>Small intestine</b>	1.9±1.8	3.1±0.5	2.2±1.2	1.9±0.6	1.5±0.4
<b>Caecum</b>	16.7±4.0	9.8±4.8	3.9±2.2	3.0±1.3	2.1±1.3
<b>Colon</b>	4.7±1.2	3.0±1.3	1.8±1.0	1.9±1.1	2.8±2.7
<b>Liver</b>	50.6±12.4	34.7±11.5	32.3±8.7	24.1±8.1	21.3±4.4
<b>Muscle</b>	1.7±0.6	1.5±0.8	1.7±1.2	0.8±0.6	0.5±0.1
<b>Blood cells</b>	4.6±2.4	3.1±1.7	1.5±0.8	1.3±0.7	2.2±0.6
<b>Plasma</b>	29.6±5.7	14.3±1.5	6.5±1.8	4.7±1.6	2.0±0.9
<b>Faeces</b>	14.3±4.6	14.1±13.9	4.7±2.9	2.4±1.5	2.6±1.9
<b>Mesenteric Ganglion</b>	4.5±2.9	3.0±0.7	2.0±0.8	1.4±0.3	0.8±0.2



**Table 3: Biodistribution of <sup>99m</sup>Tc-anti-CEA IgG-SH in healthy Balb-c mice at 4, 8, 18, 24 and 48h, expressed as the percentage of injected (IV) dose per gram, %ID/g (values represent means ± S.D. of the %ID/g)**

	4h (n=4)	8h (n=4)	18h (n=4)	24h (n=4)	48h (n=4)
Heart	9.8±4.8	8.9±1.7	4.0±2.1	3.6±1.6	2.7±1.3
Brain	0.61±0.39	0.58±0.07	0.28±0.22	0.23±0.08	0.21±0.10
Lungs	16.5±9.3	16.0±6.0	7.9±3.4	6.9±2.0	5.0±1.3
Kidneys	28.2±13.9	29.5±3.3	16.3±7.2	15.6±6.6	10.9±5.0
Spleen	37.6±16.2	28.3±8.5	18.9±13.2	17.9±8.7	12.6±4.2
Stomach	3.3±2.7	3.9±0.3	1.3±0.4	1.6±0.7	1.1±0.3
Small intestine	5.3±2.3	4.9±0.7	3.5±1.6	3.2±0.6	1.6±0.4
Caecum	6.1±3.6	7.9±1.9	3.7±2.1	4.4±1.4	2.0±0.3
Colon	3.8±2.0	4.5±0.2	2.2±1.0	3.2±1.2	1.5±0.5
Liver	29.8±16.3	45.5±16.1	18.6±7.3	21.4±4.5	15.7±2.0
Muscle	1.6±1.0	3.3±0.9	1.2±0.7	1.2±0.7	0.9±0.5
Blood cells	6.6±2.7	6.8±4.5	1.9±0.9	1.6±0.5	1.6±0.7
Plasma	45.3±21.0	29.3±8.0	13.7±7.9	12.1±4.1	8.0±1.8
Faeces	5.2±3.7	6.5±2.5	2.9±0.9	8.1±2.7	1.8±0.5
Mesenteric Ganglion	3.6±3.6	5.9±1.9	2.1±1.4	2.8±2.3	2.0±1.5

extent, <sup>99m</sup>Tc-anti-CEA mIgA-SH (16.7 ± 4.0 and 14.3 ± 4.6 %ID/g, respectively). In contrast, the uptake of <sup>99m</sup>Tc-anti-CEA IgG-SH was lower (6.1 ± 3.6 and 5.2 ± 3.8 %ID/g, respectively). Eight hours' post-injection, significant differences among the three Ig types were observed in the plasma and blood-rich tissues (heart, spleen, lungs). Higher radioactivity was observed with IgG (in plasma: 29.3 ± 8.0 and 12.1 ± 4.1 %ID/g at 8 h and 24 h, respectively), than with mIgA (14.3 ± 1.5 and 4.7 ± 1.6 %ID/g, respectively) or pIgA (10.0 ± 1.3 and 2.8 ± 1.9 %ID/g, respectively). Furthermore, beyond 18 h, strong radioactivity uptake was observed for mIgA in the spleen (at 18 h: 52.6 ± 14.1 %ID/g; at 48 h: 41.4 ± 8.7 %ID/g), whereas no uptake was seen for pIgA or IgG. High radioactivity uptake was also recorded in the liver for all three Ig types. This uptake was more important for pIgA at 4 h (pIgA: 66.4 ± 19.4 %ID/g, mIgA: 50.6 ± 12.4 %ID/g, IgG: 29.8 ± 16.3 %ID/g) but the difference among the three Ig types was not significant.

#### **Biodistribution of <sup>99m</sup>Tc-anti-CEA pIgA-SH and <sup>99m</sup>Tc-irrelevant pIgA-SH in a mouse colorectal tumour model**

According to the biodistribution results in normal mice (pIgA-SH higher uptake in the caecum), biodistribution studies of <sup>99m</sup>Tc-anti-CEA pIgA-SH at 4 h and 8 h were performed in a mouse CRC orthotopic

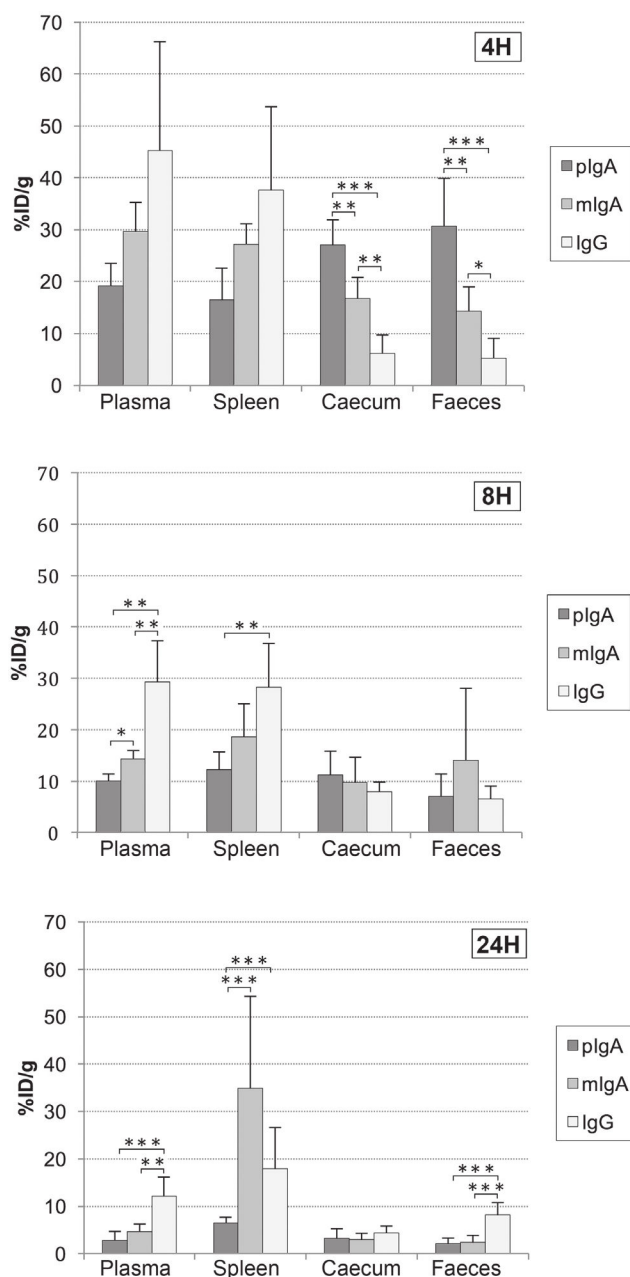
graft model (Table 4) and compared with irrelevant IgA (<sup>99m</sup>Tc-anti-peanut pIgA-SH) and healthy nude mice to evaluate the *in vivo* CEA affinity of IgA. All of the mice included in the studies had macroscopic colorectal tumours in the caecum observed after dissection. A notable accumulation of radioactivity was found in the "faeces-removed" tumoural caecum (15.8 ± 5.5 and 18.5 ± 8.5 %ID/g at 4 h and 8 h, respectively), but this radioactivity uptake was not significantly different from that of the two controls (<sup>99m</sup>Tc-irrelevant pIgA-SH: 16.7 ± 6.9 and 11.5 ± 6.5 %ID/g; nude mouse control: 7.9 ± 2.5 and 19.9 ± 1.4 %ID/g). The pattern of distribution of <sup>99m</sup>Tc-anti-CEA pIgA-SH between the different organs in the mouse colorectal tumour model was similar to that observed in healthy mouse. However, a significant uptake in the lungs of <sup>99m</sup>Tc-anti-CEA pIgA-SH was observed after as early as 4 h after injection, and doubled at 8 h (at 4 h: 42.8 ± 59.4 %ID/g; at 8 h: 97.6 ± 29.5 %ID/g). This uptake was significantly more intense (p<0.001) than the low fixation observed in the lungs at 4 h or 8 h after the injection of irrelevant <sup>99m</sup>Tc-pIgA-SH, as well as in the nude mouse control. For some animals (n = 1–3, depending on group), a tumour was present in the peritoneal cavity opposite the injection site. On these cases, uptake was measured and found to be significantly higher with anti-CEA pIgA than with irrelevant pIgA (2.9 ± 1.4 and 1.5 ± 0.5 %ID/g, respectively; p = 0.0479).



## Pathological analysis of the caecum and lungs

The caecum and lungs in the mouse colorectal tumour model were sliced and analysed using coloration techniques and immunohistochemistry (Figure 5). HES analysis confirmed the presence of tumour mass in the caecal wall, particular in the sub-mucosa (Figure 5A and 5B). Moreover, the tumour cells could secrete mucus, as revealed by Alcian blue staining (black arrows in Figure 5C). The tumours were vascularised, and CD31-positive endothelial cells

were detected within the tumour mass (black arrows in Figure 5D). These data confirmed that the model mimics a mucosal human adenocarcinoma. CD31 analysis of the tumour mass revealed that tumour cells were present within tumour vessels (black arrows in Figure 5E). HES analysis in the lungs highlighted the presence of thick and large cells within the pulmonary parenchyma, corresponding to lung metastases (Figure 5F) and explaining the strong uptake of anti-CEA pIgA in this organ (at 4 h:  $42.8 \pm 59.4$  %ID/g; at 8 h:  $97.6 \pm 29.5$  %ID/g).



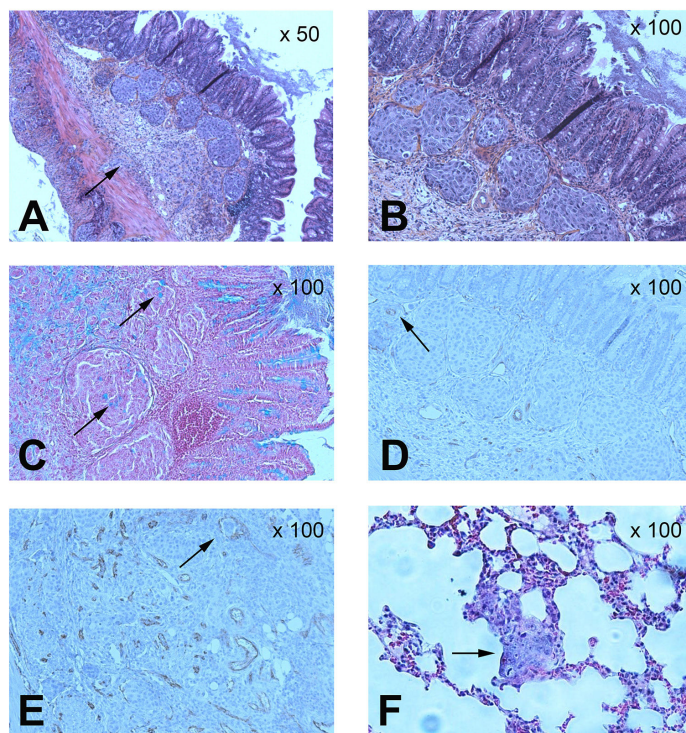
**Figure 4: Biodistribution of  $^{99m}\text{Tc}$ -anti-CEA pIgA-SH,  $^{99m}\text{Tc}$ -anti-CEA mIgA-SH and  $^{99m}\text{Tc}$ -anti-CEA IgG-SH.** Representation of radiolabeled Ig biodistribution at 4, 8 and 24 h post-injection in selected organs: plasma, spleen, caecum and Faeces. Values are expressed as the percentage of injected (intravenous; IV) dose per gram %ID/g and represent the means  $\pm$  S.D. of the %ID/g. \*:  $p$  value  $< 0.05$ ; \*\*:  $p$  value  $< 0.01$ ; \*\*\*:  $p$  value  $< 0.005$ .





**Table 4: Biodistribution of <sup>99m</sup>Tc-pIgA-SH in nude mice bearing intracaecal tumours, and in nude healthy mice at 4h and 8h, expressed as the percentage of injected dose per gram, %ID/g (values represent means ± SD of the %ID/g)**

	Nude mice bearing intracaecal tumours				Nude healthy mice	
	Anti-CEA pIgA		Irrelevant pIgA		Anti-CEA pIgA	
	4h (n=6)	8h (n=6)	4h (n=6)	8h (n=6)	4h (n=4)	8h (n=4)
Heart	2.5±0.6	2.4±1.2	2.2±0.7	1.3±0.08	3.2±0.6	2.1±0.7
Brain	0.24±0.09	0.16±0.04	0.20±0.06	0.11±0.01	0.23±0.08	0.15±0.03
Lungs	42.8±59.4	97.6±29.5	3.8±1.5	4.8±1.3	13.5±10.4	19.0±6.6
Kidneys	16.0±2.3	17.1±4.7	16.6±5.8	10.4±1.5	16.0±6.0	12.1±0.1
Spleen	17.6±12.4	14.5±7.8	9.9±4.9	20.9±8.1	8.2±2.3	7.1±1.3
Stomach	1.8±1.0	1.7±0.8	1.3±0.8	0.9±0.3	1.9±0.7	1.9±1.5
Small intestine	4.6±1.7	3.6±2.9	4.9±2.1	1.8±0.4	4.7±2.0	3.2±0.2
Caecum (empty)	15.8±5.5	18.5±8.5	16.7±6.9	11.5±6.5	7.9±2.5	19.9±1.4
Colon	6.4±2.6	3.4±1.8	7.2±4.9	3.2±0.8	10.0±0.6	4.4±0.7
Liver	50.5±20.3	46.3±21.8	50.3±7.2	41.9±9.2	34.4±1.2	40.2±3.3
Muscle	0.7±0.3	0.9±0.2	0.8±0.2	0.8±0.3	0.9±0.2	0.6±0.1
Blood cells	3.1±0.8	2.3±0.9	2.5±1.2	1.3±0.2	2.2±1.3	0.8±0.1
Plasma	9.8±2.7	8.2±5.1	7.9±2.2	4.4±0.4	9.7±2.1	5.6±0.6
Feces	34.2±24.0	13.3±9.5	34.9±22.9	9.7±6.2	35.2±12.9	35.0±10.1



**Figure 5: Pathological analysis of caecum and lungs in the tumour mouse model.** Histochemical analysis of colorectal orthotopic grafts: (A) (magnification x50) and (B) (x100): haematoxylin, eosin and safran (HES) analysis. (C) (x100): mucus cells with Alcian Blue staining. (D) (x100): vessels staining with CD31. (E) (x100): tumour cells detected in caecum tumour vessels with CD31 staining. Lungs histochemical analysis: (F) (x100): HES analysis.



## MicroSPECT/CT imaging of normal SCID mice and tumour bearing SCID or Tsg SCID-CD89 mice with <sup>99m</sup>Tc-anti-CEA pIgA-SH, <sup>99m</sup>Tc-irrelevant pIgA-SH and <sup>99m</sup>Tc-anti-CEA IgG-SH

Biodistribution of <sup>99m</sup>Tc-anti-CEA pIgA-SH, <sup>99m</sup>Tc-irrelevant pIgA-SH and <sup>99m</sup>Tc-anti-CEA IgG-SH was subsequently evaluated 4 h post-injection by microSPECT/CT in normal SCID mice and in SCID and Tsg SCID-CD89 colorectal tumour model (n=3 for each Ig and mice model).

In healthy mice, for the three Ig, an intense uptake was observed in the liver as highlighted also in the biodistribution studies. Furthermore, an intense uptake was observed in the digestive tract after injection of anti-CEA pIgA and to a lesser extent of irrelevant pIgA, whereas a low uptake of the anti-CEA IgG was revealed (Figure 6A–6C). Total digestive activity was quantified and compared to whole body, for <sup>99m</sup>Tc-anti-CEA pIgA, <sup>99m</sup>Tc-irrelevant pIgA and <sup>99m</sup>Tc-anti-CEA IgG and represented 32 ± 6%, 19 ± 8% and 8 ± 1% respectively.

In engrafted mice, a high uptake was observed in the liver for the three Ig. An intense uptake was highlighted in the digestive tract after <sup>99m</sup>Tc-anti-CEA pIgA injection (Figure 6D and 6G), compared to the low uptake of <sup>99m</sup>Tc-anti-CEA IgG (Figure 6F and 6I). Nevertheless, the whole digestive tract uptake of <sup>99m</sup>Tc-anti-CEA pIgA didn't significantly differ between healthy and engrafted mice (32 ± 6% and 30 ± 9% respectively).

In both healthy and engrafted mice, an intense uptake of <sup>99m</sup>Tc-anti-CEA IgA was also noticed in the gallbladder (n = 9, data not shown) whereas no significant uptake was ever observed in mice injected with the <sup>99m</sup>Tc-anti-CEA IgG (n = 9, data not shown). This result is consistent with the physiological hepato-biliary pathway of IgA.

When the uptake of IgA observed in engrafted SCID mice (Figure 6D and 6E) was compared to the uptake observed in engrafted Tsg SCID-CD89 mice (Figure 6G and 6H), the contribution of FCαRI does not appear to be determinant. No significant increase uptake was observed in engrafted Tsg SCID-CD89 mice.

As the engrafted mouse models (NUDE, SCID, Tsg SCID-CD89) developed metastases in both liver and lungs, colonisation was evaluated by a sensitive real-time (TaqMan) PCR technique and compared with microSPECT/CT data. The percentage of metastases infiltration in the lungs, for both engrafted SCID and Tsg SCID-CD89 mice, reached the level of 0.0003% to 0.0052% of the lung tissue (Figure 7E). At this early stage of metastases spreading, the <sup>99m</sup>Tc-anti-CEA IgA failed to detect significant uptake by microSPECT/CT imaging (mean value in pulmonary Region of Interest (ROI): 0,5 ± 0,1 in Figures 6D-6G and 7B-7D). However, when the tumour infiltrate represented higher colonisation (0.08% ± 0.01%) of the pulmonary tissue, as observed in one mouse,

a significant diffuse uptake of <sup>99m</sup>Tc-anti-CEA IgA was detected in the lung (mean value in pulmonary ROI: 1,4 in Figure 7A-7C) but no pulmonary nodule was identified in microSPECT/CT slice, while the imaging resolution was about 1 mm. In the liver, despite a high level of colonisation quantified by PCR (data not shown) and also observed macroscopically, the uptake was important but not significantly different from healthy mice.

## DISCUSSION

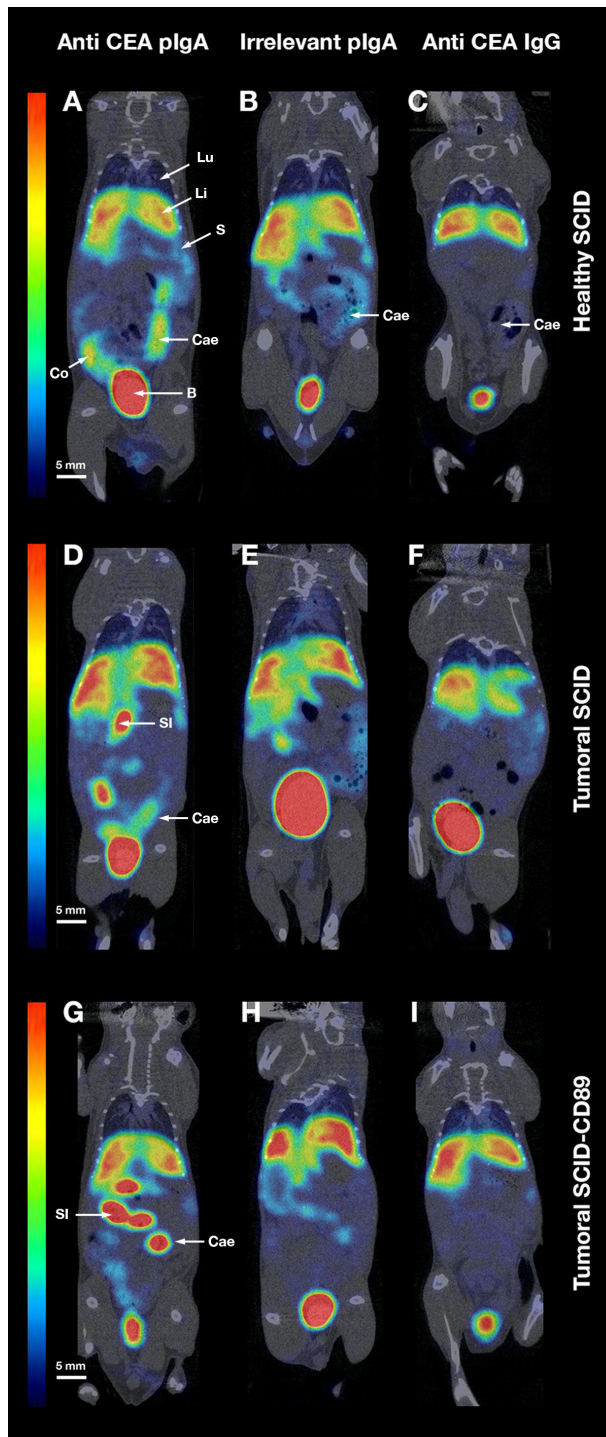
This work reported anti-CEA IgA radiolabelling with <sup>99m</sup>Tc to evaluate IgA biodistribution compared with IgG, and to estimate IgA tumour-targeting potency.

This study focused on IgA, a class of Ig that to date has been under-studied compared with IgG. Presently, IgG is the most widely used antibody in therapy and clinical trials. However, some functional limitations of IgG, such as inadequate tissue accessibility and pharmacokinetics, and impaired interactions with effector cells, have been highlighted [21]. IgA could represent a promising alternative to IgG, particularly to target mucosal tumours, considering that IgA constitutes the major Ig class at the mucosal surface.

To develop a soft <sup>99m</sup>Tc radiolabelling method applicable to IgA, which preserves the biological functions of antibody and is suitable for low-concentration samples of mAbs, we selected the method initially developed by Biechlin *et al.* [22] and subsequently optimised and evaluated by Carpenet *et al.* [20]. This method involved two steps: Ig derivatisation with a synthon, 2-IT, and Ig-SH radiolabelling by an indirect method using the <sup>99m</sup>Tc-tricarbonyl core. The <sup>99m</sup>Tc-tricarbonyl core is a complex with high kinetic stability and an excellent precursor for radiolabelling various biomolecules, especially Igs. To functionalise Ig by adding SH- groups under milder conditions, and to obtain high radiolabelling yields, the 2-IT synthon was used. Our results confirmed the significant additional value of the derivatisation step for IgA, as previously observed for IgG [19], allowing an RCP > 95%. High RCP enables radiolabelled Ig administration without an additional purification step. The stability of radiolabelled IgA was checked over 24 h in PBS and murine plasma. Studies in PBS showed that IgA radiolabelling was stable. In murine plasma, the RCP decreased progressively over 24 h, suggesting slight degradation comparable with that seen in IgG stability studies. Our stability results were also consistent with published data from Ig studies using tricarbonyl core (loss of 10–15%) [23, 24] or direct labelling after disulphide bond reduction [25].

Western blot analyses indicated that the radiolabelling process did not alter the IgA global molecular structure. Affinity studies confirmed that radiolabelled IgA showed a well-preserved binding affinity (Kd ≈ 25 nM). Concordant Kd were obtained with the two





**Figure 6: MicroSPECT/CT data.** MicroSPECT/CT imaging of normal SCID mice (A-B-C) and tumour bearing SCID (D-E-F) or Tsg SCID-CD89 mice (G-H-I) with  $^{99m}\text{Tc}$ -anti-CEA pIgA-SH (A-D-G),  $^{99m}\text{Tc}$ -irrelevant pIgA-SH (B-E-H) and  $^{99m}\text{Tc}$ -anti-CEA IgG-SH (C-F-I). Tumoural mice were grafted by direct orthotopic cell microinjection of human colon carcinoma. Imaging studies were performed 6 weeks after the cell-microinjection procedure. Each mouse received 50-60 MBq of  $^{99m}\text{Tc}$ -Ig. SPECT/CT images were acquired 4 h post-injection for 30 min. Representative coronal slice (80  $\mu\text{m}$  thickness) of microSPECT/CT acquisition in the different mice models/conditions:

- Healthy SCID after injection of  $^{99m}\text{Tc}$ -anti-CEA pIgA-SH (A),  $^{99m}\text{Tc}$ -irrelevant pIgA-SH (B),  $^{99m}\text{Tc}$ -anti-CEA IgG-SH (C)
- Tumoural SCID after injection of  $^{99m}\text{Tc}$ -anti-CEA pIgA-SH (D),  $^{99m}\text{Tc}$ -irrelevant pIgA-SH (E),  $^{99m}\text{Tc}$ -anti-CEA IgG-SH (F)
- Tumoural Tsg SCID-CD89 after injection of  $^{99m}\text{Tc}$ -anti-CEA pIgA-SH (G),  $^{99m}\text{Tc}$ -irrelevant pIgA-SH (H),  $^{99m}\text{Tc}$ -anti-CEA IgG-SH (I)

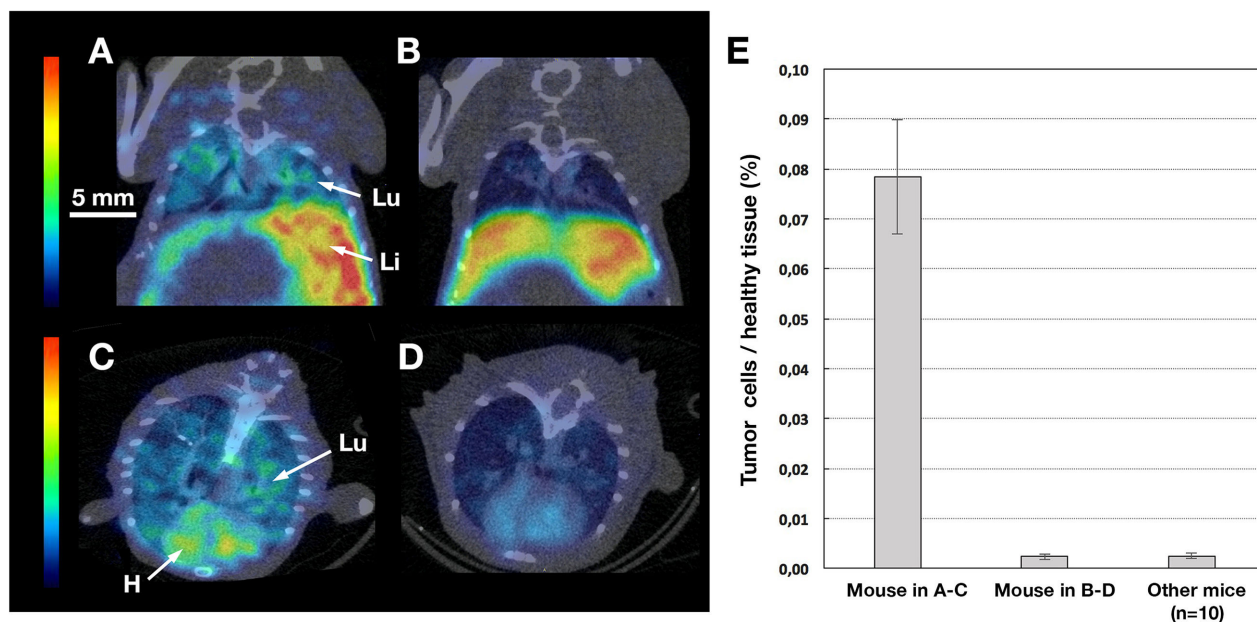
Principal organs were indicated with a white arrow and letters: B = Bladder, Cae = Caecum, Co = Colon, Li = Liver, Lu = Lung, S = Spleen, SI = Small Intestine.



affinity determination methods. All of these *in vitro* studies suggest that radiolabelling preserves IgA functionality.

Biodistribution studies in healthy mice have compared radiolabelled pIgA, mIgA and IgG, all carrying the same variable domains directed against the CEA antigen. The mIgA presented an intermediate fixation profile between the profiles shown by pIgA and IgG, except in the spleen. Although the liver is the main organ involved in IgA removal, participation of the spleen in clearance was already observed also in some cirrhosis patients and was almost invariably concurrent with normal parameters [26]. Compared with IgG, a greater and more rapid fixation (at 4 h) was observed in the caecum and faeces for pIgA and, to a lesser extent, mIgA. This study confirms that pIgA has a strong and fast tropism for the caecum at the *lamina propria*, as previously described in the literature. Rapid secretion in the gut lumen and high concentration in the faeces (4 h) was observed; however, for IgG, radioactivity was maximal in the faeces at 24 h. Significant differences between pIgA/mIgA and IgG were also observed in the plasma and blood-rich tissues. The plasma retention of mIgA and pIgA decreased rapidly, which is consistent with their well-known short sera half-life [27]. This appears to be balanced by their rapid rise in concentration, as observed in faeces, in accordance with the redistribution in the mucosal membrane. In mouse models, the short serum half-life of IgA antibodies

seems to be partially caused by rapid clearance by the asialoglycoprotein receptor (ASGPR) recognising terminal galactose residues of IgA [28, 29], even if this clearance pathway is more addressed to IgA2 than IgA1. In contrast, IgG is characterised by a relatively long biological half-life in the bloodstream. Unlike IgA, IgG binds to the neonatal receptor FcRn, expressed on vascular endothelial cells, macrophages and monocytes. This interaction could be one explanation for the longer serum half-life of IgG [30]. Recently, by molecular fusion of an albumin-binding domain (ABD), the half-life and *in vivo* serum exposure of ABD-modified IgA was found to increase significantly due to indirect targeting of the FcRn pathway [31]. Concerning the liver, a high uptake of the three Ig isotypes studied herein was observed, explained in part by the expression of the ASGPR on the hepatocyte surface, but also by engagement of IgA in the hepatobiliary system involved in IgA secretion in the bile duct [32]. pIgA is transported from blood to bile without degradation [33], through hepatocytes by the pIgR. SIgA is released in the bile. Most of the pIgA that are transported from blood to bile in the liver are synthesized by plasma cells found in the *lamina propria* of the intestinal mucosa and enters the bloodstream at the thoracic duct. Circulating pIgA is then bound at the sinusoidal front of hepatocytes and transported across the cell, processes mediated by the pIgR. According to Hoppe's work [34], 36% of <sup>125</sup>I-pIgA



**Figure 7: MicroSPECT/CT and qPCR pulmonary data.** MicroSPECT/CT imaging of tumour bearing SCID with <sup>99m</sup>Tc-anti-CEA pIgA-SH (A-B-C-D) and qPCR of pulmonary sample (E). Tumoural mice were grafted by direct orthotopic cell microinjection of human colon carcinoma. Imaging studies were performed 6 weeks after the cell-microinjection procedure. Each mouse received 50-60 MBq of <sup>99m</sup>Tc-anti-CEA pIgA-SH. SPECT/CT images were acquired 4 h post-injection for 30 min. 80 μm thickness slice in coronal (A-B) and axial plans (C-D) of a representative view of tumour bearing SCID mice. Principal organs were indicated with a white arrow and letters: H = Heart, Li = Liver, Lu = Lung. The mouse in A-C was in a higher metastasis evolution (peritoneal carcinosis) than the mouse in B-D, confirmed by qPCR analysis on pulmonary samples (E). The <sup>99m</sup>Tc-anti-CEA pIgA-SH uptake is diffuse and significantly higher in the lung of the mouse in A-C than in B-D mouse.





(mouse myeloma produced) administrated by IV in rat, are transported as intact protein in bile over 3h post-injection; the pIgA reaches its highest concentration only 30 to 60 min after injection, and 80% of the total of  $^{125}\text{I}$ -pIgA are finally secreted in bile by 90 min. MicroSPECT/CT imaging confirmed and highlighted hepatobiliary system as a high uptake was observed in the gallbladder.

To evaluate IgA targeting potency, a CRC tumour model was set up in which human cancer cells were grafted in the mucosal environment. Pathological microscopic analysis clearly revealed a structural glandular architecture of the grafted tumour and the presence of large vacuoles in the WiDr cell line, consistent with muco-secretions in the *lamina propria* layer. Cancer cells invaded the normal caecum, under the muscle layer through the *lamina propria*, to produce protruding polyps in the lumen. Depending on the delay after direct orthotopic cell microinjection, different stages of CRC have been observed from localised tumours to metastases in the lungs. Immunohistochemical analysis in tumours revealed that tumour cells were present within tumour vessels, suggesting cellular dissemination by the vascular system. All of these factors led us to consider colorectal orthotopic grafts as being useful models of human CRC, because they share the same characteristics with human tumours. In humans, in advanced colorectal cancer, metastases are spread in the liver and lung in particular. The engrafted mouse models (NUDE, SCID, Tsg SCID-CD89) mimic metastases colonisation in both liver and lung.

Surprisingly, in *post mortem* biodistribution studies, no significant difference in caecum uptake was observed after the injection of radiolabelled anti-CEA IgA or irrelevant IgA, or in healthy nude mice. Tumours implanted in the external layers of the caecum progress by infiltrating the whole caecum, without an evident limit between the tumoural tissue and the healthy tissue. Caecal tumours were partly necrotic and weakly vascularised in some regions. Furthermore, pIgA, regardless of their antigenicity, have a strong tropism for caecal mucosa. So, the tumour-to-background ratio was not optimal, and radioactive determination of the whole caecum probably hampered tumour-specific signal detection. It is well-known that, in the late stage of tumour differentiation, the majority of CEA molecules may not be reached by IV-injected mAbs because CEA is mostly expressed on the apical side of carcinoma cells hidden by the pseudo-lumen structures of the malignant glands [35]. However, human colon carcinoma cells maintained pIgR expression on the basolateral side on the external part of the pseudo-lumen structures. By transepithelial translocation, only dIgA, and not IgG, can reach the pseudo-lumen to target CEA at this late stage of adenocarcinoma differentiation. Conversely, lung metastases are characterised by an immature stage of differentiation, without any glandular organisation and certainly without CEA-specific polarisation. Thus, it is interesting and important that anti-CEA pIgA uptake

was observed in the lungs, at a significantly higher level than that seen after injection of radiolabelled irrelevant IgA, or in healthy nude mice. This result is consistent with the presence of lung metastases, as demonstrated by pathological analysis. Even if no significant uptake was seen in the tumoural caecum with anti-CEA pIgA, this antibody allowed the targeting of lung metastases derived from the initial colorectal tumour. Due to intrinsic mucosal tropism and biomarker affinity, pIgA could very effectively reach its target in the lungs.

The *in vivo* imaging by microSPECT/CT confirm that the anti-CEA pIgA is a suitable tool for the detection of mucosal tumours, particularly in lungs, but *in vivo* imaging is less sensitive (1 mm resolution) and uptake was detectable only for an intermediate tissue colonisation ( $0.08\% \pm 0.01\%$ ). In the liver, partly because of the high natural uptake of IgA (by the hepato-biliary pathway), the anti-CEA pIgA seems not be well adapted in detecting metastases, hide by the strong uptake background, despite a massive metastases colonisation quantified by qPCR (up to 0,9%, data not show).

The Fc $\alpha$ RI is a critical receptor to mediate cytotoxicity effect, but its involvement in pIgA biodistribution is still unclear. To investigate this pathway, microSPECT/CT studies were conducted in transgenic mice Tsg SCID-CD89. No significant increase uptake was observed in Tsg SCID-CD89 mice compared to SCID mice, neither in digestive tract, neither in lungs. This result suggests that the distribution of the IgA molecule could occur without active transport linked to the Fc $\alpha$ RI receptor by infiltration of myeloid cells at the mucosal site, particularly in digestive tract. In pulmonary mucosal, further investigations are necessary to clarify this preliminary report.

In conclusion, monomeric and polymeric IgA were efficiently and indirectly radiolabelled with  $^{99\text{m}}\text{Tc}$ , using limited amounts of antibody. *In vitro* studies showed that IgA radiolabelling was stable and did not alter IgA functionality. Biodistribution studies in normal BALB/c mice, of  $^{99\text{m}}\text{Tc}$ -anti-CEA mIgA-SH and  $^{99\text{m}}\text{Tc}$ -anti-CEA pIgA-SH, confirmed their shorter serum half-life and rapid and strong mucosal tropism of pIgA and, to a lesser extent, mIgA. High lung uptake of  $^{99\text{m}}\text{Tc}$ -anti-CEA pIgA-SH in the mouse tumour model suggested efficient targeting potency of pIgA, even though a significant and specific uptake in caecum was not observed. As SIgA is resistant to intestinal degradation, it would be interesting to explore SIgA biodistribution after oral administration. SIgA reverse transcytosis, mediated by epithelial M cells, remains understudied, but in humans several receptors seem to be involved (i.e. Dectin-1, DC-SIGN [36], CD71 [37]). Finally, this work represents a first step toward IgA development to envisage, in the near future, diagnostic imaging tools and therapeutic IgA-based strategies targeting tumours in mucosal epitheliums.



## MATERIALS AND METHODS

All of the chemicals and reagents were obtained from Sigma-Aldrich (Saint-Quentin Fallavier, France) unless otherwise specified.

### Ig production

All of the Ig types were produced by B Cell Design Society (Limoges, France). Monomeric and polymeric anti-CEA (relevant) or anti-peanut (irrelevant) human chimeric IgA1 were synthesised using HAMIGA™ technology [18]. For *in vitro* studies, a mIgA was used; however, for *in vivo* studies, the enriched fraction of the mIgA or pIgA form was tested separately (purities of 95% and 85%, respectively). Anti-CEA human recombinant IgG1 was synthesised after cloning the variable regions of the heavy and light chains of anti-CEA human chimeric IgA1. It was then produced in human embryonic kidney cells (HEK 293-6E; NRC, Quebec, Canada). All of the antibodies were purified by affinity chromatography using a Tricorn Column 5/100 with protein A-Sepharose at a flow rate of 1.0 mL/min (GE Healthcare, Waukesha, WI, USA) and were eluted with glycine (0.1 M pH 2.7) equilibrated in Tris/base (1.0 M). Subsequently, IgA and IgG were dialysed against PBS by centrifugation (1,000 × *g*, 15 min) using Amicon 30 kDa (Millipore, Saint-Quentin, France).

For imaging experiment, anti-CEA IgG was purchased (1105 clone, Fisher Scientific, Elancourt, France).

The protein concentrations were determined before and after radiolabelling using bicinchoninic Micro BC Assay® kits (Fisher Scientific, Elancourt, France), using bovine serum albumin (BSA) as a standard with quantification limits of 2.5 and 100 µg/mL.

### Cell culture

Two human colorectal cell lines were used: WiDr, a primary adenocarcinoma of the rectosigmoid [38], and DLD1, a colorectal glandular carcinoma [39]. WiDr, which are CEA-expressing cells, and DLD1, which are CEA-negative cells [40], were purchased from ATCC (Manassas, VA, USA). The two cell lines were routinely grown in RPMI and minimum essential medium, respectively, supplemented with 10% foetal calf serum, 1% sodium pyruvate, and 1% penicillin (100 U/mL)-streptomycin (100 µg/mL). For WiDr, the medium was also supplemented with 1% glutamine and 1% nonessential amino acids.

### Ig radiolabelling with $[^{99m}\text{Tc}(\text{CO})_3(\text{H}_2\text{O})_3]^+$

The radiolabelling method used has been described previously [20]. Briefly, the first step was thiol-

derivatisation of Ig with 2-IT. Next, 0.5 to 2.2 nmol IgA and IgG (300 µL in PBS) were incubated with 2-IT (3.8 µM, 25°C, 120 min). The solutions were purified by size exclusion chromatography. The number of thiol groups was determined by a micromethod using Ellman's reagent (5,5'-dithiobis-2-nitrobenzoic acid, DTNB). The second step was synthesis of the tricarbonyl precursor  $[^{99m}\text{Tc}(\text{CO})_3(\text{H}_2\text{O})_3]^+$ . Next, 0.8–1 mL of freshly eluted  $[\text{Na}^{99m}\text{TcO}_4]$  (CisBio, Saclay, France) in fixed activities (2,220–3,700 MBq) was added to the IsoLink® kit (Covidien, Petten, The Netherlands) and incubated for 25 min at 100°C. RCP analysis was performed by thin-layer chromatography (TLC) using two systems to separate the  $[^{99m}\text{Tc}(\text{CO})_3(\text{H}_2\text{O})_3]^+$  from free  $[^{99m}\text{Tc}]$ -pertechnetate, reduced  $^{99m}\text{Tc}$  and hydrolysed  $[^{99m}\text{Tc}(\text{OH})_n(\text{H}_2\text{O})_y]$ : (1) Baker Flex Aluminium oxide strips (JT Baker Inc., Phillipsburg, NJ, USA) - methanol/hydrochloric acid (95/5 v/v); (2) Instant thin layer chromatography-silica gel (ITLC-SG, Varian, Les Ulis, France) - methanol. The  $^{99m}\text{Tc}$ -IsoLink® labelling yields were superior to 98%. The third and last step was the radiolabelling of native or derivatised Ig with  $[^{99m}\text{Tc}(\text{CO})_3(\text{H}_2\text{O})_3]^+$ . A total of 0.5–2.2 nmol of non-derivatised IgA or thiol derivatised IgA (IgA-SH), or 1.5 nmol IgG-SH in 300 µL of PBS, was incubated for 120 min (25°C) with 150 µL (148–185 MBq) of a  $^{99m}\text{Tc}$ -tricarbonyl solution, previously neutralised to pH 7.0 (0.5 M HCl). RCP was determined by TLC with ITLC-SG/NaCl 0.9%.

### *In vitro* stability of $^{99m}\text{Tc}$ -IgA-SH

The stability of radiolabelled IgA-SH was checked for 24 h, in PBS (4°C and 25°C) and murine plasma (1/10, 37°C). Aliquots of 5 µL were analysed by TLC (ITLC-SG/NaCl 0.9%) at various time points (1, 2, 4, 16 and 24 h).

### Structural integrity of $^{99m}\text{Tc}$ -IgA-SH

SDS-PAGE was performed under non-reducing conditions with the native or labelled Ig. The proteins (1.5–10 µg per lane) and molecular weight standards (Bio-Rad, Hercules, CA, USA) were loaded and resolved with a Mini-Protean TGX Precast Gel using a Bio-Rad apparatus. Proteins were immediately transferred to a PVDF membrane. The membrane was then washed, blocked with a 3% milk phosphate buffer and incubated (90 min, 25°C) in diluted primary antibody solution (1/1,000, goat antihuman IgA-horseradish peroxidase; Southern Biotech, Birmingham, AL, USA). The membrane signals were revealed using the diaminobenzidine (DAB) substrate.

### Affinity of $^{99m}\text{Tc}$ -IgA-SH

The binding affinity was evaluated by two complementary methods: in plate with coated CEA antigen and directly on cells expressing CEA antigen.



## Binding affinity by RIA

To evaluate anti-CEA IgA binding to human CEA, removable well plates were coated with 1 µg/well of CEA antigen (AbCys SA, Courtaboeuf, France) and incubated (overnight, 4°C) in PBS at pH 7.4. Next, they were blocked with PBS-gelatine 2% (1 h, 37°C). Relevant (anti-CEA) or irrelevant (anti-peanut) radiolabelled IgA (100 µL, 0.5–16 µg/mL) was added in series to antigen-coated wells. The first series was diluted in PBS-gelatine 0.2% and the second was diluted in PBS-gelatine 0.2% containing unlabelled antibody at 500 µg/mL (non-specific binding). Plates were incubated (2 h, 37°C) and washed three times with PBS at pH 7.4. The radioactivity contained in the wells was determined by gamma counting (Cobra 5003; Canberra Packard, Frankfurt, Germany).

## Binding affinity on cells

Saturation binding studies were assessed according to the method of Lindmo *et al.* [41, 42]. The <sup>99m</sup>Tc-IgA-SH binding affinity experiments were performed on WiDr (anti-CEA and anti-peanut IgA) and DLD1 (anti-CEA IgA). Samples of cells (1 million cells/250 µL) were first pre-incubated with an anti-mouse CD16/CD32 IgG2Bk (1/250, 25°C, 30 min; BD Biosciences Pharmingen, Maharashtra, India) to block the Fc receptor. They were then pre-incubated with PBS-BSA 3% or with unlabelled antibody (final concentration of 100 µg/mL in PBS-BSA 3%) to saturate the binding sites for nonspecific binding determination (25°C, 90 min). Next, duplicate samples were incubated with <sup>99m</sup>Tc-IgA-SH increasing concentrations (125 µL, 0.1 to 10–13 µg/mL). After incubation (2 h, 25°C) with continuous rotation, the cells were filtered using Manifold® (Millipore) and washed with PBS. Filter radioactivity was evaluated with a gamma-counter.

## Affinity data analysis

For wells and cell-binding affinity studies, specific binding was evaluated by subtracting nonspecific binding (determined after incubation with unlabelled antibody) from total binding. Data analysis was performed using a Scatchard plot of the binding of <sup>99m</sup>Tc-anti-CEA IgA-SH, and the K<sub>d</sub>, number of antibody binding sites/well or cell, and B<sub>max</sub> of <sup>99m</sup>Tc-anti-CEA IgA-SH were determined.

## Biodistribution of <sup>99m</sup>Tc-anti-CEA IgA-SH and <sup>99m</sup>Tc-anti-CEA IgG-SH in normal mice

All *in vivo* experiments were performed in accordance with animal ethical rules, and all efforts were made to minimise suffering. The protocol was approved by the Comité Régional d'Ethique sur l'Expérimentation Animale du Limousin (CREEAL). Biodistribution

experiments were carried out in 7-week-old male BALB/c mice (Charles River Laboratories, Chatillon-sur-Chalaronne, L'Arbresle Cedex). <sup>99m</sup>Tc-IgA-SH monomeric, <sup>99m</sup>Tc-IgA-SH polymeric or <sup>99m</sup>Tc-IgG-SH (40 MBq, 170 µg for each antibody) was injected intravenously (IV; tail vein). Animals were euthanised, by anaesthesia and cervical dislocation, at different time points after administration (4 h, 8 h, 18 h, 24 h, 48 h post-injection). Selected tissues were excised, rinsed, and weighed, and their radioactivity levels were measured with a gamma-counter. The uptake of radioactivity in these organs was expressed as %ID/g after correcting for radioactive decay at each time point. Blood cells, plasma, and faeces were also collected and measured. Faeces refers to faecal matter collected in the small intestine and colon during dissection. Liver and gallbladder were not separated before counting.

## Biodistribution and tumour uptake of <sup>99m</sup>Tc-anti-CEA pIgA-SH in nude mice bearing human colon carcinoma

Direct orthotopic cell microinjection was achieved according to the method of Cespedes *et al.* [43].

Seven-week-old female nude mice (athymic nu/nu; Envigo, Gannat, France) were anaesthetised with ketamine (80 mg/kg) (Imalgène, Merial, Lyon, France) and xylazine (9.6 µg/kg) (Rompun 2%; Bayer, Lyon, France) to exteriorise their caecum by laparotomy. Next, 2.10<sup>5</sup> WiDr cells, suspended in 10 µL of PBS in a sterile micropipette (30 G needle, 30 µL syringe), were slowly injected between the mucosa and *muscularis externa* layers of the caecal wall, under a binocular lens with an approximate 30° angle. After injection, the caecum was extensively washed with sterile PBS before reintroduction into the abdominal cavity. The laparotomy was closed by surgical suture. Animals were weighed each day. If animals demonstrated clinical alteration or weight loss, they were euthanised by anaesthesia and cervical dislocation.

## Biodistribution and tumour uptake

Six weeks after the cell microinjection procedure, the animals were divided into two groups. The first group (*n* = 6) received 170 µg of <sup>99m</sup>Tc-anti-CEA pIgA-SH (IV) and the second (*n* = 6) received 170 µg of irrelevant <sup>99m</sup>Tc-pIgA-SH (corresponding to 35–37 MBq). Nude mouse controls received the same <sup>99m</sup>Tc-anti-CEA pIgA-SH. The animals were euthanised by anaesthesia and cervical dislocation at 4 h and 8 h after administration. Finally, the procedure described in biodistribution studies of normal mice was applied. Furthermore, the caecum was longitudinally opened, washed with PBS and counted separately from caecal faeces to evaluate luminal IgA presence. The caeca and lungs were fixed with buffered formalin during radioactive decay (48 h).



## Histological analysis of human colorectal orthotopic grafts

Mouse organs were transferred to 4% formol and included in paraffin after automated cycling of a dehydration system (ASP200S; Leica, Heidelberg, Germany). Next, 4- $\mu$ m sections were prepared using a microtome. For histological analysis, slides were stained with haematoxylin, eosin and safran (HES analysis) or with Alcian blue (secretion mucus analysis). For vascularisation analysis, CD31 staining was performed using the VENTANA robot from the pathology department of Limoges University Hospital. In each case, a pathologist was solicited to interpret the staining.

## MicroSPECT/CT imaging of normal SCID mice and tumour bearing SCID or Tsg SCID-CD89 mice with $^{99m}\text{Tc}$ -anti-CEA pIgA-SH, $^{99m}\text{Tc}$ -irrelevant pIgA-SH and $^{99m}\text{Tc}$ -anti-CEA IgG-SH

SCID (Envigo, Gannat, France) or Tsg SCID-CD89 mice (J. Leusen, University Medical Center Utrecht, The Netherlands) were grafted as described previously for nude mice.  $^{99m}\text{Tc}$ -anti-CEA pIgA-SH,  $^{99m}\text{Tc}$ -irrelevant pIgA-SH or  $^{99m}\text{Tc}$ -anti-CEA IgG-SH (60 MBq) was injected IV. Anaesthesia induction was achieved using 3% isoflurane and animals were maintained under general anaesthesia with 1,8% isoflurane for the duration of imaging. Air/Oxygen (50%/50%, 1,4 l/minutes) was maintained throughout anaesthesia induction and maintenance. Acquisitions were performed 4 h post-injection with a MicroSPECT/CT (U-SPECT4/CT, MILabs, Utrecht, The Netherlands). Images were acquired during 30 minutes and energy windows were set over the 140 keV peaks ( $\pm 20\%$ ). The SPECT resolution with  $^{99m}\text{Tc}$  is less than 1 mm. Images were analysed and uptake were quantified in Region of Interest (ROI) with PMOD Software (PMOD Technologies, Zürich, Switzerland).

## qPCR on lung samples

The presence of human tumour cells and metastases within engrafted mouse organs was quantified, by a TaqMan-chemistry based real-time PCR using a primer/probe-combination described previously [44]. Briefly, two days after microSPECT/CT imaging, animals were euthanized, by anaesthesia and cervical dislocation, and lung tissue was excised, two samples (in right lobe, left lobe) of 0.02 g were collected. Sample DNA was extracted using a nucleoSpin Tissue kit (Macherey-Nagel, Hoerd, France) and quantify on nanodrop quantifier (Ozyme, Montigny, France). A short fragment (467bp) of the  $\alpha$ -satellite region of the human chromosome 17 was amplified using a probe (Cr17-probe) 5'-labelled with the reporter fluorescent dye FAM (6-carboxy-fluoresceine) and linked to a non-fluorescent quencher dye TAMRA. The reactions were performed in a Applied

Biosystems Step One Plus Real Time PCR, and analysed by the Step One software (v2.3). The PCR were run in 96-well microtiter plates with a final volume of 20  $\mu$ L, containing 12.5 ng of genomic DNA and 10 nM of each primer. 45 cycles were performed. Each sample was tested twice in parallel. Serial dilutions of human carcinoma cells (WiDR cell line) in mouse lung tissue (0.02 g biopsy) served to define the TaqMan-calibration curves. In parallel, a human-mouse control PCR was set up in order to normalize DNA quantity per sample. A primer/probe combination was designed in the  $\beta$ -actin encoding region, to be absolutely identical both in human and in mouse sequence ( $\beta$ -actin Forward primer: (5' \_tctgcgcaagttaggtttg\_3';  $\beta$ -actin reverse primer: (5' \_gatcattgctcctctgagc\_3'; FAM $\beta$ -actin\_probeTAMRA: (5' \_tcactactctgctgctgctgac\_3') and amplified a short (211bp) fragment.

## Statistical analysis

Statistical analyses were performed by applying the nonparametric Kruskal-Wallis test for antibody comparison. The tests were conducted using StatView software (ver. 5.0 ; SAS Institute, Cary, NC, USA). A *p* value < 0.05 was considered to indicate statistical significance.

## Author contributions

Conceived and designed the experiments: I. Quelven, A. Cuvillier, H. Carpenet, J. Monteil.

Performed the experiments: H. Carpenet, A. Cuvillier, I. Quelven, O. Martin, A. Perraud.

Analyzed the data: H. Carpenet, I. Quelven, A. Cuvillier, A. Perraud.

Contributed reagents/materials/analysis tools: H. Carpenet, A. Cuvillier, G. Champier, I. Quelven.

Wrote the paper: H. Carpenet, I. Quelven, A. Cuvillier, J. Monteil, MO. Jauberteau.

## ACKNOWLEDGMENTS

Some of the work presented here was supported, in part, by the Comité d'Orientation de la Recherche sur le Cancer (CORC) en Limousin. The authors are grateful to Mallinckrodt Medical for providing IsoLink<sup>®</sup> for the research. We would also like to thank F. Dalmay (INSERM UMR-S 1094 NET) for assistance with the statistical analyses, and the Pathology Department of Limoges University Hospital for assistance with the histological analysis. We thank C. Déloménie, M. Bosselut and A. Francès (B Cell Design) for IgA production and tumoural mice model, S. Lenoir (UMR CNRS 7276) for help with Real-time QPCR experiment set up and S. Durand-Panteix (UMR CNRS 7276) for suggestions and comments on the manuscript.





## CONFLICTS OF INTEREST

The authors declare that they have no competing interests. B Cell Design Society provided support in the form of a salary for authors, A. Cuvillier and G. Champier, Ig production and molecular reagents and mice supply, but did not have any additional role in the study.

## FUNDING

Some of the work presented here was supported, in part, by the Comité d'Orientation de la Recherche sur le Cancer (CORC) en Limousin. Mallinckrodt Medical provided graciously IsoLink® for the research. B Cell Design Society provided all immunoglobulins, all molecular reagents for Western Blot, real-time QPCR, immunocompromised mice and transgenic mouse model (Tsg SCID-CD89).

## REFERENCES

- Bradley CJ, Lansdorp-Vogelaar I, Yabroff KR, Dahman B, Mariotto A, Feuer EJ, Brown ML. Productivity savings from colorectal cancer prevention and control strategies. *Am J Prev Med.* 2011; 41:e5–14.
- Mach JP, Chatal JF, Lumbroso JD, Buchegger F, Forni M, Ritschard J, Berche C, Douillard JY, Carrel S, Herlyn M, Steplewski Z, Koprowski H. Tumor localization in patients by radiolabeled monoclonal antibodies against colon carcinoma. *Cancer Res.* 1983; 43:5593–600.
- Pinkas L, Robins PD, Forstrom LA, Mahoney DW, Mullan BP. Clinical experience with radiolabelled monoclonal antibodies in the detection of colorectal and ovarian carcinoma recurrence and review of the literature. *Nucl Med Commun.* 1999; 20:689–96.
- Karube Y, Katsuno K, Takata J, Matsunaga K, Haruno M, Kuroki M, Arakawa F, Matsuoka Y, Kanda H. Radioimmunosintigraphy using Technetium-99m-labeled parental mouse and mouse-human chimeric antibodies to carcinoembryonic antigen in athymic nude mice bearing tumor. *Nucl Med Biol.* 1996; 23:753–9.
- Erb DA, Nabi HA. Clinical and technical considerations for imaging colorectal cancers with technetium-99m-labeled antiCEA Fab' fragment. *J Nucl Med Technol.* 2000; 28:12–8; quiz 21.
- Kraeber-Bodéré F, Rousseau C, Oudoux A, Bodet-Milin C, Davodeau F, Barbet J, Faivre-Chauvet A, Chérel M. L'immuno-TEP, une nouvelle approche d'imagerie moléculaire. *Médecine Nucléaire.* 2010; 34:295–8.
- Olafsen T, Wu AM. Antibody vectors for imaging. *Semin Nucl Med.* 2010; 40:167–81.
- Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. *J Exp Med.* 1965; 122:467–81.
- Pakdel A, Naghibalhossaini F, Mokaram P, Mansooreh Jaberipour M, Hosseini A. Regulation of carcinoembryonic antigen release from colorectal cancer cells. *Clin Biochem.* 2011; 44:S68–9.
- Yao Y, Yang Z, Li ZF, Gu J. Immunoscintigraphy of local recurrent rectal cancer with 99mTc-labeled anti-CEA monoclonal antibody CL58. *World J Gastroenterol.* 2007; 13:1841–6.
- Serafini AN, Klein JL, Wolff BG, Baum R, Chetanneau A, Pecking A, Fischman AJ, Hoover HC Jr, Wynant GE, Subramanian R, Goroff DK, Hanna MG Jr. Radioimmunosintigraphy of recurrent, metastatic, or occult colorectal cancer with technetium 99m-labeled totally human monoclonal antibody 88BV59: results of pivotal, phase III multicenter. *J Clin Oncol.* 1998; 16:1777–87.
- Woof JM, Mestecky J. Mucosal immunoglobulins. *Immunol Rev.* 2005; 206:64–82.
- Monteiro RC, Van De Winkel JG. IgA Fc receptors. *Annu Rev Immunol.* 2003; 21:177–204.
- Launay P, Grossetête B, Arcos-Fajardo M, Gaudin E, Torres SP, Beaudoin L, Patey-Mariaud de Serre N, Lehuen A, Monteiro RC. Fcα receptor (CD89) mediates the development of immunoglobulin A (IgA) nephropathy (Berger's disease). Evidence for pathogenic soluble receptor-IgA complexes in patients and CD89 transgenic mice. *J Exp Med.* 2000; 191:1999–2009.
- Bakema JE, van Egmond M. The human immunoglobulin A Fc receptor FcαRI: a multifaceted regulator of mucosal immunity. *Mucosal Immunol.* 2011; 4:612–24.
- Lohse S, Derer S, Beyer T, Klausz K, Peipp M, Leusen JH, van de Winkel JG, Dechant M, Valerius T. Recombinant dimeric IgA antibodies against the epidermal growth factor receptor mediate effective tumor cell killing. *J Immunol.* 2011; 186:3770–8.
- Fagarasan S, Kinoshita K, Muramatsu M, Ikuta K, Honjo T. *In situ* class switching and differentiation to IgA-producing cells in the gut lamina propria. *Nature.* 2001; 413:639–43.
- Duchez S, Amin R, Cogné N, Delpy L, Sirac C, Pascal V, Corthésy B, Cogné M. Premature replacement of mu with alpha immunoglobulin chains impairs lymphopoiesis and mucosal homing but promotes plasma cell maturation. *Proc Natl Acad Sci U S A.* 2010; 107:3064–9.
- Camacho X, García MF, Calzada V, Fernández M, Chabalgoity JA, Moreno M, Barbosa de Aguiar R, Alonso O, Gambini JP, Chammas R, Cabral P. [(99m)Tc(CO)(3)]-radiolabeled bevacizumab: *in vitro* and *in vivo* evaluation in a melanoma model. *Oncology.* 2013; 84:200–9.
- Carpenet H, Cuvillier A, Monteil J, Quelven I. Anti-CD20 immunoglobulin G radiolabeling with a <sup>99m</sup>Tc-tricarbonyl core: *in vitro* and *in vivo* evaluations. *PLoS One.* 2015; 10:e0139835.
- Beckman RA, Weiner LM, Davis HM. Antibody constructs in cancer therapy: protein engineering strategies to improve exposure in solid tumors. *Cancer.* 2007; 109:170–9.



22. Biechlin ML, d'Hardemare A du M, Fraysse M, Gilly FN, Bonmartin A. Improvement in radiolabelling proteins with the  $^{99m}\text{Tc}$ -tricarbonyl-core  $[\text{}^{99m}\text{Tc}(\text{CO})_3]^+$ , by thiol-derivatization with iminothiolane: application to  $\gamma$ -globulins and annexin V. *J Label Comp Radiopharmaceut*. 2005; 48:873–85.
23. Stopar TG, Mlinaric-Rascan I, Fettich J, Hojker S, Mather SJ.  $(99\text{m})\text{Tc}$ -rituximab radiolabelled by photo-activation: a new non-Hodgkin's lymphoma imaging agent. *Eur J Nucl Med Mol Imaging*. 2006; 33:53–9.
24. Dias CR, Jeger S, Osso JA Jr, Müller C, De Pasquale C, Hohn A, Waibel R, Schibli R. Radiolabeling of rituximab with  $(188)\text{Re}$  and  $(99\text{m})\text{Tc}$  using the tricarbonyl technology. *Nucl Med Biol*. 2011; 38:19–28.
25. Malviya G, Anzola KL, Podestà E, Laganà B, Del Mastro C, Dierckx RA, Scopinaro F, Signore A.  $(99\text{m})\text{Tc}$ -labeled rituximab for imaging B lymphocyte infiltration in inflammatory autoimmune disease patients. *Mol Imaging Biol*. 2012; 14:637–46.
26. Roccatello D, Picciotto G, Torchio M, Ropolo R, Ferro M, Franceschini R, Quattrocchio G, Cacace G, Coppo R, Sena LM. Removal systems of immunoglobulin A and immunoglobulin A containing complexes in IgA nephropathy and cirrhosis patients. The role of asialoglycoprotein receptors. *Lab Invest*. 1993; 69:714–23.
27. Monteiro RC. Role of IgA and IgA fc receptors in inflammation. *J Clin Immunol*. 2010; 30:1–9.
28. Stockert RJ, Kressner MS, Collins JC, Sternlieb I, Morell AG. IgA interaction with the asialoglycoprotein receptor. *Proc Natl Acad Sci U S A*. 1982; 79:6229–31.
29. Rifai A, Fadden K, Morrison SL, Chintalacharuvu KR. The N-glycans determine the differential blood clearance and hepatic uptake of human immunoglobulin (Ig)A1 and IgA2 isotypes. *J Exp Med*. 2000; 191:2171–82.
30. Roopenian DC, Akilesh S. FcRn: the neonatal Fc receptor comes of age. *Nat Rev Immunol*. 2007; 7:715–25.
31. Meyer S, Nederend M, Jansen JH, Reiding KR, Jacobino SR, Meeldijk J, Bovenschen N, Wuhler M, Valerius T, Ubink R, Boross P, Rouwendal G, Leusen JH. Improved *in vivo* anti-tumor effects of IgA-Her2 antibodies through half-life extension and serum exposure enhancement by FcRn targeting. *MAbs*. 2016; 8:87–98.
32. Harmatz PR, Kleinman RE, Bunnell BW, McClenathan DT, Walker WA, Bloch KJ. The effect of bile duct obstruction on the clearance of circulating IgA immune complexes. *Hepatology*. 1984; 4:96–100.
33. Orlans E, Peppard J, Reynolds J, Hall J. Rapid active transport of immunoglobulin A from blood to bile. *J Exp Med*. 1978; 147:588–92.
34. Hoppe CA, Connolly TP, Hubbard AL. Transcellular transport of polymeric IgA in the rat hepatocyte: biochemical and morphological characterization of the transport pathway. *J Cell Biol*. 1985; 101:2113–23.
35. Buchegger F, Pfister C, Fournier K, Prevel F, Schreyer M, Carrel S, Mach JP. Ablation of human colon carcinoma in nude mice by  $^{131}\text{I}$ -labeled monoclonal anti-carcinoembryonic antigen antibody F(ab')<sub>2</sub> fragments. *J Clin Invest*. 1989; 83:1449–56.
36. Rochereau N, Drocourt D, Perouzel E, Pavot V, Redelinghuys P, Brown GD, Tiraby G, Roblin X, Verrier B, Genin C, Corthésy B, Paul S. Dectin-1 is essential for reverse transcytosis of glycosylated SIgA-antigen complexes by intestinal M cells. *PLoS Biol*. 2013; 11:e1001658.
37. Lebreton C, Ménard S, Abed J, Moura IC, Coppo R, Dugave C, Monteiro RC, Fricot A, Traore MG, Griffin M, Cellier C, Malamut G, Cerf-Bensussan N, Heyman M. Interactions among secretory immunoglobulin A, CD71, and transglutaminase-2 affect permeability of intestinal epithelial cells to gliadin peptides. *Gastroenterology*. 2012; 143:698–707.e4.
38. Flatmark K, Mælandsmo GM, Martinsen M, Rasmussen H, Fodstad Ø. Twelve colorectal cancer cell lines exhibit highly variable growth and metastatic capacities in an orthotopic model in nude mice. *Eur J Cancer*. 2004; 40:1593–8.
39. Maréchal R, De Schutter J, Nagy N, Demetter P, Lemmers A, Devière J, Salmon I, Tejpar S, Van Laethem JL. Putative contribution of CD56 positive cells in cetuximab treatment efficacy in first-line metastatic colorectal cancer patients. *BMC Cancer*. 2010; 10:340.
40. Guadagni F, Witt PL, Robbins PF, Schlom J, Greiner JW. Regulation of carcinoembryonic antigen expression in different human colorectal tumor cells by interferon-gamma. *Cancer Res*. 1990; 50:6248–55.
41. Lindmo T, Boven E, Cuttitta F, Fedorko J, Bunn PA Jr. Determination of the immunoreactive fraction of radiolabeled monoclonal antibodies by linear extrapolation to binding at infinite antigen excess. *J Immunol Methods*. 1984; 72:77–89.
42. Lindmo T, Bunn PA Jr. Determination of the true immunoreactive fraction of monoclonal antibodies after radiolabeling. *Meth Enzymol*. 1986; 121:678–91.
43. Céspedes MV, Espina C, García-Cabezas MA, Trias M, Boluda A, Gómez del Pulgar MT, Sancho FJ, Nistal M, Lacal JC, Mangués R. Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. *Am J Pathol*. 2007; 170:1077–85.
44. Becker M, Nitsche A, Neumann C, Aumann J, Junghahn I, Fichtner I. Sensitive PCR method for the detection and real-time quantification of human cells in xenotransplantation systems. *Br J Cancer*. 2002; 87:1328–35.



**Article 3 : Nuclear proximity of Mtr4 to RNA exosome restricts DNA mutational asymetry.**

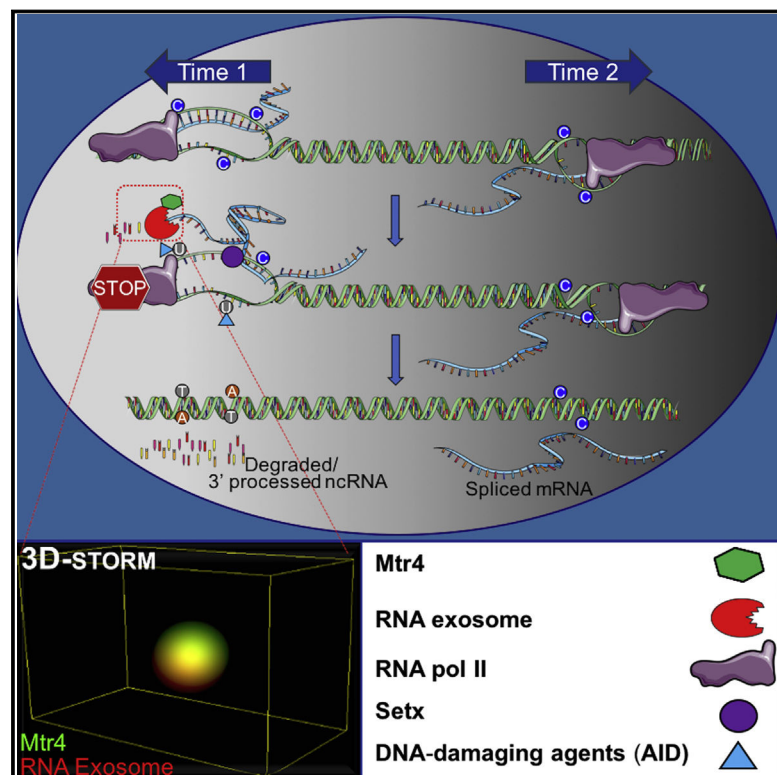
Junghyun Lim, Pankaj Kumar Giri, David Kazadi, Brice Laffleur, Wanwei Zhang, Veronika Grinstein, Evangelos Pefanis, Lewis M. Brown, Erik Ladewig, Ophélie Martin, Yuling Chen, Raul Rabadan, François Boyer, Gerson Rothschild, Michel Cogné, Eric Pinaud, Haiteng Deng, and Uttiya Basu.

Cell. 2017 April 20; 169(3): 523–537.e15. doi:10.1016/j.cell.2017.03.043.



# Nuclear Proximity of Mtr4 to RNA Exosome Restricts DNA Mutational Asymmetry

## Graphical Abstract



## Authors

Junghyun Lim, Pankaj Kumar Giri,  
David Kazadi, ..., Eric Pinaud,  
Haiteng Deng, Uttiya Basu

## Correspondence

pg2501@cumc.columbia.edu (P.K.G.),  
ub2121@cumc.columbia.edu (U.B.)

## In Brief

Limitation of local RNA-DNA structures  
restricts DNA mutational asymmetry.

## Highlights

- Mtr4 and Setx cooperatively unwind RNA exosome-sensitive ncRNA from DNA/RNA hybrids
- Physical proximity of Mtr4 and AID is a function of nuclear RNA exosome activity
- Mtr4 and Setx restrict transcription coupled asymmetric DNA mutagenesis by AID







# Nuclear Proximity of Mtr4 to RNA Exosome Restricts DNA Mutational Asymmetry

Junghyun Lim,<sup>1,6</sup> Pankaj Kumar Giri,<sup>1,6,\*</sup> David Kazadi,<sup>1</sup> Brice Laffleur,<sup>1</sup> Wanwei Zhang,<sup>1</sup> Veronika Grinstein,<sup>1</sup> Evangelos Pefanis,<sup>1</sup> Lewis M. Brown,<sup>2</sup> Erik Ladewig,<sup>3</sup> Ophélie Martin,<sup>4</sup> Yuling Chen,<sup>5</sup> Raul Rabadan,<sup>3</sup> François Boyer,<sup>4</sup> Gerson Rothschild,<sup>1</sup> Michel Cogné,<sup>4</sup> Eric Pinaud,<sup>4</sup> Haiteng Deng,<sup>5</sup> and Uttiya Basu<sup>1,7,\*</sup>

<sup>1</sup>Department of Microbiology and Immunology, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA

<sup>2</sup>Department of Biological Sciences, Quantitative Proteomics Center, Columbia University, New York, NY 10027, USA

<sup>3</sup>Departments of Systems Biology and Biomedical Informatics, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA

<sup>4</sup>Université de Limoges, Centre National de la Recherche Scientifique, CHU Limoges, CRIBL, UMR 7276, 87000 Limoges, France

<sup>5</sup>School of Life Sciences, Tsinghua University, Beijing 100084, China

<sup>6</sup>These authors contributed equally

<sup>7</sup>Lead Contact

\*Correspondence: [pg2501@cumc.columbia.edu](mailto:pg2501@cumc.columbia.edu) (P.K.G.), [ub2121@cumc.columbia.edu](mailto:ub2121@cumc.columbia.edu) (U.B.)

<http://dx.doi.org/10.1016/j.cell.2017.03.043>

## SUMMARY

The distribution of sense and antisense strand DNA mutations on transcribed duplex DNA contributes to the development of immune and neural systems along with the progression of cancer. Because developmentally matured B cells undergo biologically programmed strand-specific DNA mutagenesis at focal DNA/RNA hybrid structures, they make a convenient system to investigate strand-specific mutagenesis mechanisms. We demonstrate that the sense and antisense strand DNA mutagenesis at the immunoglobulin heavy chain locus and some other regions of the B cell genome depends upon localized RNA processing protein complex formation in the nucleus. Both the physical proximity and coupled activities of RNA helicase Mtr4 (and senataxin) with the noncoding RNA processing function of RNA exosome determine the strand-specific distribution of DNA mutations. Our study suggests that strand-specific DNA mutagenesis-associated mechanisms will play major roles in other undiscovered aspects of organismic development.

## INTRODUCTION

Patterns of DNA mutagenesis in somatic cells are implicated in many biological events, including the development of immunity, neurogenesis, and the onset of cancer (Alexandrov et al., 2013a, 2013b; Basu et al., 2011; Haradhvala et al., 2016; Keim et al., 2013; Lodato et al., 2015; Madabhushi et al., 2015; Wei et al., 2016). Regulated mutagenesis on both strands of DNA directs the formation of programmed DNA double strand breaks (DSBs) during immunoglobulin (Ig) locus recombination events. In many cancers, and in several neurodegenerative diseases, an enhanced bias of mutations to the non-template (or sense) strand of DNA has been observed (Haradhvala et al., 2016). Can-

cer genomes may generate asymmetric strand mutagenesis via transcription-coupled and/or replication-coupled mechanisms. Because both transcription and replication are essential requirements for cellular function, robust preventive measures are required to forestall strand asymmetric mutations (Haradhvala et al., 2016). The formation of secondary DNA structures such as R-loops or G-quadruplexes that attract various DNA mutator proteins, and likewise pathophysiological insults that have affinity for single-strand DNA (ssDNA) substrates, have been considered to skew transcription-associated strand asymmetric DNA mutagenesis toward the sense strand of coding genes (Santos-Pereira and Aguilera, 2015). Stalled RNA polymerase II occurring at sites of strand asymmetric DNA mutagenesis and associated secondary DNA structures, if not rapidly resolved, may cause the replication complex to collide with the transcription complex and create genomic instability and, occasionally, chromosomal translocations (Kim and Jinks-Robertson, 2012).

Activation-induced deaminase (AID)-mediated DNA hypermutation occurs in various regions of the immunoglobulin loci, catalyzing proper antibody diversification via class switch recombination (CSR) and somatic hypermutation (SHM) and occurring with relatively small asymmetry on both the sense and the antisense strands of DNA (Alt et al., 2013; Basu et al., 2011; Keim et al., 2013; Pefanis and Basu, 2015; Pefanis et al., 2014). In the variable regions of the immunoglobulin gene loci, AID mutates both strands of DNA to engineer high-affinity antibodies (Pefanis et al., 2014; Wang et al., 2014). Via a related but different mechanism in switch regions, AID mutates both strands of DNA to generate DNA double strand breaks that permit class switch recombination (Basu et al., 2011). Thus, evolutionary mechanisms that promote mutagenesis to both antisense and sense strands of DNA may be coupled with AID's function in B cells. These reasons allow us to follow AID's activity as an ideal model system to understand regulatory events associated with antisense DNA strand mutagenesis.

In this study, we report that the noncoding RNA processing activity of RNA exosome, a 3'-5' RNA exonuclease, controls sense and antisense strand DNA mutagenesis, and prevents an increased asymmetric mutagenesis burden using a mechanism



that relies on the unexpected DNA/RNA hybrid unwinding activities of the Mtr4 and senataxin RNA helicases. We unravel the role of non-coding RNA helicases in regulating transcription-associated DNA template (antisense) strand mutagenesis in mammalian genomes. Our data demonstrate that RNA helicases, RNA exosome, and mutator are exquisitely positioned relative to each other in the nucleus of B cells to control DNA strand-specific mutagenesis. Thus, we propose a new mechanism of DNA mutagenesis regulation which, if not functional, places the burden of strand asymmetric mutation-mediated genomic instability within a particular locus.

## RESULTS

### A Model System to Investigate Nuclear RNA Exosome Complex and Associated Factors In Vivo

In vitro studies demonstrate that the RNA exosome by itself has potent RNA degradation activity but lacks the ability to access DNA-associated nascent RNA (Januszynk and Lima, 2014). Because it is proposed that RNA exosome's role involves rapid degradation/processing of noncoding RNAs generated from various transcription complexes, it is likely that there are associated protein factors that provide rapid access to non-coding RNA (ncRNA) substrates that are otherwise present inside DNA heteroduplexes. The RNA exosome complex consists of 11 subunits, the core complex formed from nine subunits (Exosc1–Exosc9) and two additional subunits (Exosc10 and Dis3) providing 3'-5' RNA exonuclease activity (Januszynk and Lima, 2014). Without the Exosc3 subunit, the RNA exosome fails to function in cells. To purify the RNA exosome complex from B cells, we generated a mouse model in which we introduced a tandem affinity purification tag (TAP-tag) consisting of a FLAG-epitope and a biotinylation site. Details of the *Exosc3*<sup>TAP-tag/TAP-tag</sup> mouse model may be found in the STAR Methods and Figures S1A and S1B. Knowing that activity of the *Exosc3* gene is important for catalysis of class switch recombination in B cells (Pefanis et al., 2014), CSR efficiency in B cells from *Exosc3*<sup>TAPtag/TAPtag</sup> mice was evaluated and found to be comparable with that seen in *Exosc3*<sup>WT/WT</sup> B cells, firmly establishing that the epitope tagged *Exosc3* is an integral component of a functional RNA exosome moiety in B cells (Figure S1C). B cells isolated from the *Exosc3*<sup>TAPtag/TAPtag</sup> mouse model were activated for CSR and used for immunoprecipitation experiments with high stringency conditions using two steps, first with anti-Flag antibody and second with a streptavidin column used for eluting biotinylated protein complexes (approach outlined in Figure S1D). Purification of the complex was first verified by direct western blotting of the immunoprecipitate with anti-*Exosc3* antibody, streptavidin (to detect biotinylated *Exosc3*), and anti-Flag antibody (Figure S1E). The double-tagged purified RNA exosome complex (and a parallel purification preparation obtained from *Exosc3*<sup>WT/WT</sup> B cells expressing BirA enzyme, an engineered lysine residue biotinylation enzyme) were analyzed by mass spectrometry for composition (details in the STAR Methods). The purified *Exosc3* TAP-tagged RNA exosome complex consisted of all 11 subunits of the RNA exosome complex (Figures 1A–1C; details of protein peptides in Tables S1A and S1B).

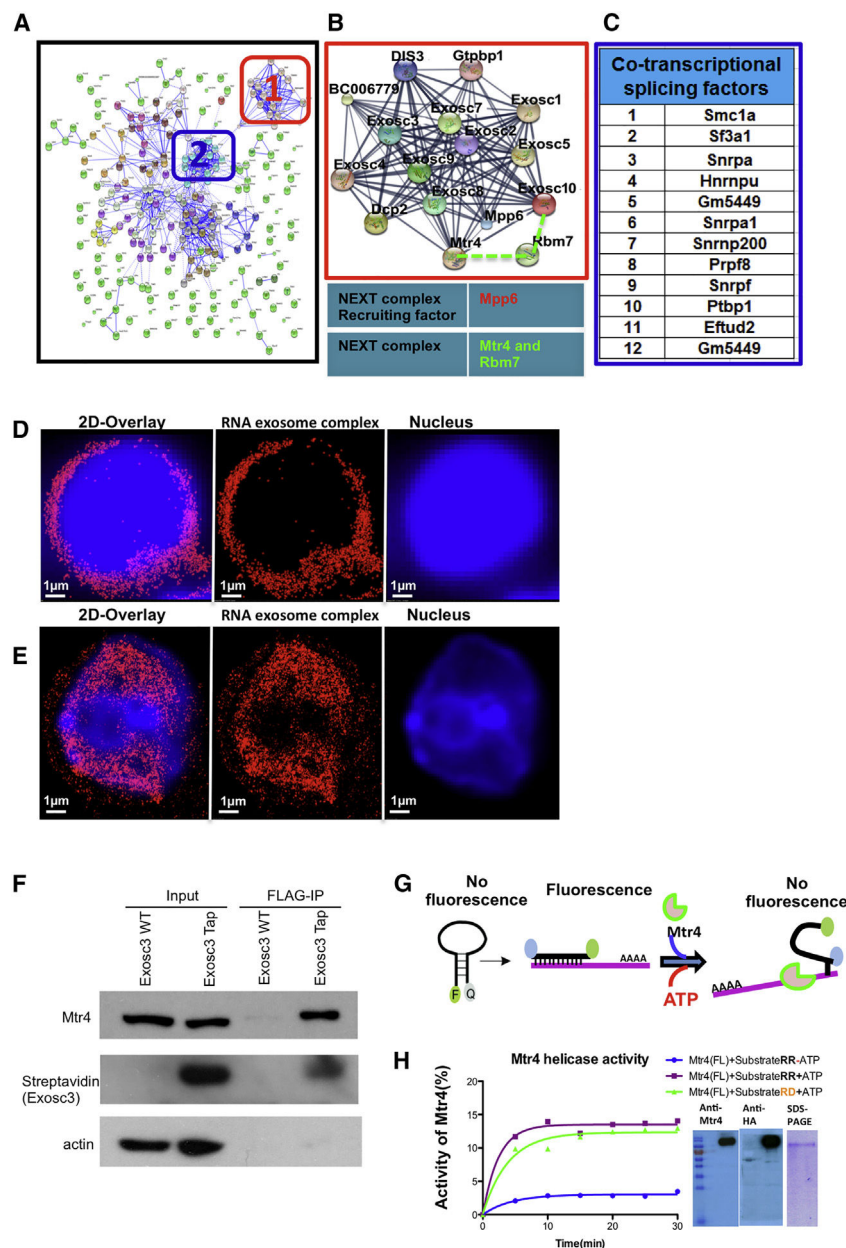
To identify cofactors that are of immediate interest, pathway analysis (ingenuity pathway analysis [IPA], parameters described in the STAR Methods) was performed, grouping cofactors based on their known functions. We identified two major clusters of proteins from the RNA exosome tandem affinity purification (Figure 1A). The two clusters were (1) the core RNA exosome complex and associated proteins with details shown in Figure 1B, and (2) non-core splicing pathway-associated proteins that have function in transcription coupled mRNA splicing pathways, with details in Figure 1C. It is evident that the core RNA exosome physically interacts with components of the nuclear exosome targeting (NEXT) complex, Mtr4, and RBM7 (Lubas et al., 2011) and with the NEXT complex recruitment factor, Mpp6 (Figure 1B).

In activated B cells, our preferred experimental system, the TAP-tagged RNA exosome complex could be evaluated by microscopy, making the system tractable to imaging experiments. Strikingly, following activation with a cocktail of stimuli, the RNA exosome complex was found to translocate into the nucleus of B cells, thereby providing a unique opportunity to investigate the function of the nuclear RNA exosome complex. As shown by three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy (3D-STORM) images, unstimulated B cells have very low levels of TAP-tagged nuclear RNA exosome (Figure 1D), but a significant amount of the RNA exosome complex translocates to the nucleus following 72 hr of stimuli cocktail treatment (Figure 1E). Thus, using the *Exosc3*<sup>TAP/TAP</sup> B cells, we developed a unique system wherein the nuclear RNA exosome complex could be biochemically and functionally evaluated.

### Mtr4 Possesses DNA/RNA Hybrid Helicase Activity

In the core RNA exosome complex (complex 1; Figure 1B), we found interaction with Mtr4. We validated the RNA exosome-Mtr4 interaction with co-immunoprecipitation experiments in B cells (Figure 1F) and also via co-expression with FLAG-tagged RNA exosome complex in 293T cells (Figure S1F). Mammalian Mtr4 possesses RNA helicase activity to unwind RNA hybrids (Johnson and Jackson, 2013), although its DNA/RNA hybrid unwinding activity has not been reported. However, its ortholog in prokaryotes has been reported to unwind RNA secondary structures and DNA/RNA hybrids (Uson et al., 2015). We expressed HA-tagged MTR4 in HEK293T cells. As shown in Figure 1H, the purified HA-MTR4 can be detected by western blotting and runs as a single purified band in a Coomassie Brilliant Blue-stained SDS-PAGE. To interrogate the DNA/RNA hybrid helicase activity, we developed an assay in which an RNA strand flanked with a fluorophore (F = FAM) and a quencher (Q = Iowa black) was duplexed with a DNA or RNA strand to form a DNA/RNA or RNA/RNA hybrid substrate. This DNA/RNA or RNA/RNA heteroduplex was analyzed for HA-Mtr4-mediated unwinding in vitro, as schematized in Figure 1G. Increased unwinding leads to quenching of FAM due to its interaction with Q (Iowa black), permitting evaluation of Mtr4 helicase activity. As shown in Figure 1H, we find that in these conditions of assay, Mtr4 has substantial RNA/DNA helicase activity (in green), and this activity is comparable to RNA/RNA helicase activity (in purple).





**Figure 1. TAP-Tagged RNA Exosome Complex in Antigen-Activated B Cells Is Localized in the Nucleus and Associates with RNA/DNA Helicase Activity of Mtr4**

(A) Interaction network of RNA exosome complex-associated proteins. In-silico network analysis of proteins identified by mass-spectrometry was performed using the STRING database combined with experimental data.

(B) Higher magnification view of the boxed region (1) shows RNA exosome complex associating with NEXT complex components Mtr4 and RBM7 along with Mpp6 (recruiter of the NEXT complex), respectively.

(C) Identification of the most stringent interacting partners shown in (A) cluster 2; these proteins are related with cotranscriptional splicing pathways.

(D) Reconstructed single color 3D-STORM image from a dataset of 50,000 frames of fixed splenic B cells before CSR stimulation, in which Exosc3 was labeled with AlexaFluor647 and the nucleus with DAPI. Scale bars, 1  $\mu$ m.

(E) Single color 3D STORM image of B cells after CSR stimulation, in which Exosc3 was labeled with AlexaFluor647 and the nucleus with DAPI. Scale bars, 1  $\mu$ m.

(F) Flag immunoprecipitation reactions (IPs) were performed on *Exosc3*<sup>TAP</sup> B cells to demonstrate Exosc3 and Mtr4 interaction.

(G) Schematic representation of molecular beacon (6-FAM and Iowa black)-based helicase assay for MTR4 protein.

(H) Two substrates (RNA-RNA and RNA-DNA) were designed with 5'-6FAM and 3' Iowa black linked to 17-mer RNA molecule, which annealed with 22-mer RNA and DNA complementary strand to evaluate the helicase activity of Mtr4 for both substrates. *hMtr4* full-length protein was expressed in HEK293T cells and purified by affinity tag. Expression of the protein was confirmed by western blot using anti Mtr4 antibodies. Percentage activity of *hMtr4*(FL) was calculated by pseudo-first order rate constant describing fluorescence decay upon ATP addition [ $k$  (min<sup>-1</sup>)]. All reactions were performed in triplicate, and error bars represent SD among independent reactions. All of the 3D STORM imaging was performed in three different B cells (from independent experiments) and repeated three or more times. 3D STORM super resolution image magnification is  $\times 100$ .

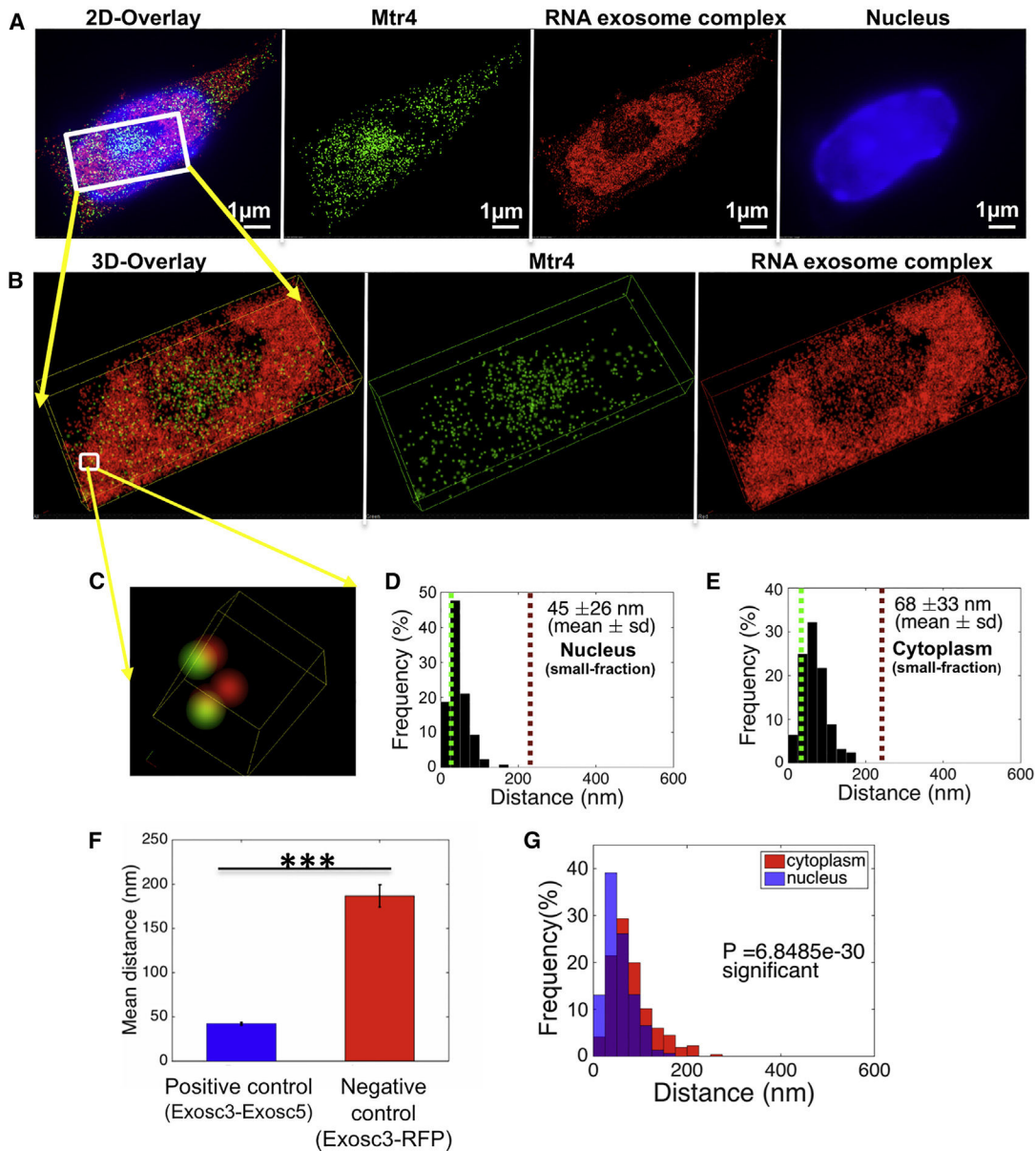
See also [Figure S1](#), [Table S1](#), and [Movies S1](#) and [S2](#).

### Increased Proximity of RNA Exosome to Mtr4 in the Nucleus of Activated B Cells

Co-immunoprecipitation of RNA exosome with Mtr4 points toward a role of Mtr4 in RNA exosome function. However, Mtr4-RNA exosome interactions may be restricted in the nucleus or occur via different sub-complexes in the nucleus versus the cytoplasm, and such variations in the mode of interaction cannot be evaluated by gross co-immunoprecipitation assays. We envision that fine regulation of Mtr4 and RNA exosome interaction can be unraveled and better understood by studying their proximity in the nucleus of activated B cells. We used 3D-STORM super resolution microscopy to evaluate RNA exosome/Mtr4 proximity in primary B cells (detailed in the [STAR Methods](#)). In the

conditions of our assays, we set closest proximity of two proteins to be 36 nm (based on the interaction of Exosc3 and Exosc5 subunits of RNA exosome; [Figures 2F](#) and [S2A–S2C](#)) and distance between two non-interacting proteins (Exosc3 with RFP-HA; [Figures 2F](#), [S2D](#), and [S2E](#)) as 205 nm. We collected two-color 3D STORM images and performed nearest neighbor analysis of Mtr4 and Exosc3 in the nucleus and cytoplasm of activated B cells. As shown in [Figures 2A–2D](#), in the nuclei of B cells there is close proximity of Mtr4 with the RNA exosome complex. It is possible that the increased frequency in proximity of RNA exosome with Mtr4 is influenced by the translocation of the RNA exosome complex to the nucleus of a B cell after stimulation ([Figure 1E](#)) (details of assay in the [STAR Methods](#)). Mtr4 and RNA





**Figure 2. Analysis of Spatial Distribution of RNA Exosome Complex and RNA Helicase Mtr4 in the Nucleus of Mouse B Cells**

B cells were harvested from Exosc3 TAP-tagged mice and treated with stimulation cocktail for 72 hr.

(A) Reconstructed two-color STORM image from a dataset of 50,000 frames with Atto488-labeled Mtr4, AlexaFluor647-labeled RNA exosome complex, and DAPI-labeled nucleus. Scale bar, 1  $\mu$ m.

(B) Three-dimensional views of the boxed region.

(C) Higher magnification three-dimensional (3D) views of the boxed region in (B).

(D and E) Histogram plot of distribution of pair interaction of Mtr4 and RNA exosome complex calculated in (D) the nuclear sub-compartment and (E) the cytoplasm by using custom written algorithm “Nearest Neighbors Search” in MATLAB (2014b, MathWorks) software.

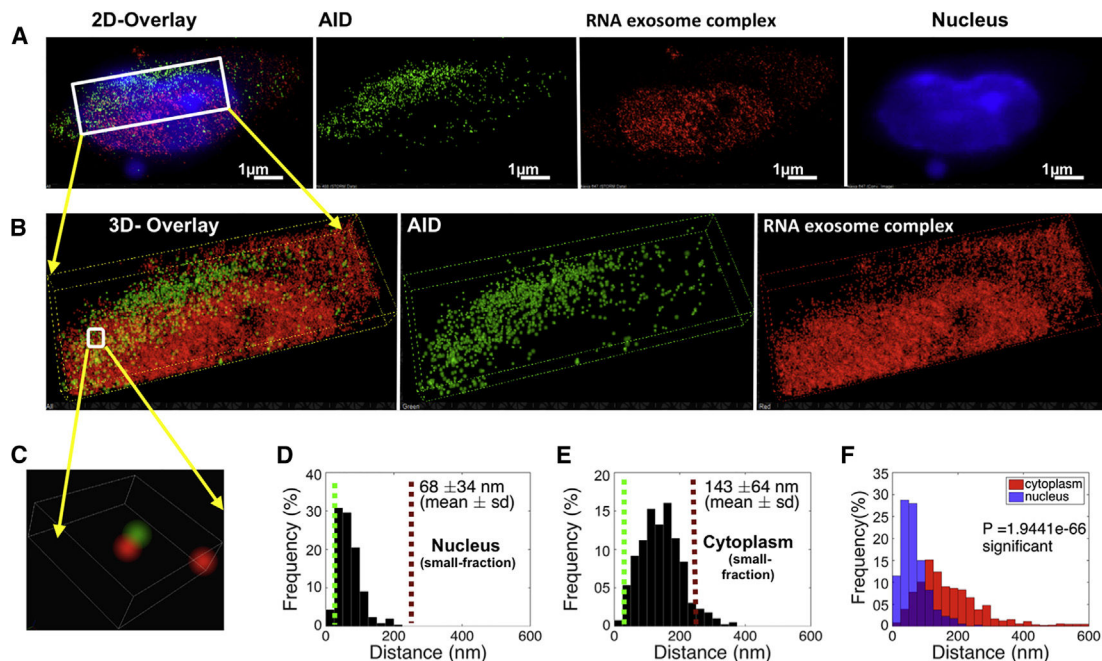
(F) Comparison of the distribution of paired interaction of Exosc3 and Exosc5 in the B cell for positive control and RFP-HA tag and RNA exosome complex in HEK293 cells for negative control.

(G) Comparison of the distribution of paired interactions in the nucleus and cytoplasm for RNA exosome complex and Mtr4 was calculated using a Student’s t test in MATLAB (2014b, MathWorks) software and p values are indicated in the graph. All of the 3D STORM imaging was performed in three separate B cells (obtained from independent experiments) and repeated three or more times. 3D STORM super resolution image magnification is  $\times 100$ . Error bars indicate SD (\*\*p value < 0.001).

See also Figure S2 and Table S1.







**Figure 3. Analysis of Spatial Distribution of AID and RNA Exosome in the Nucleus of Mouse B Cells**

B cells were harvested from Exosc3 TAP-tagged mice following 72 hr of stimulation cocktail treatment.

(A) Reconstructed two color 3D STORM image from a dataset of 50,000 frames with Atto488-labeled AID, AlexaFluor647-labeled Exosc3, and DAPI-labeled nucleus. Scale bar, 1  $\mu\text{m}$ .

(B) Three-dimensional views of the boxed region.

(C) Higher magnification three-dimensional (3D) views of the boxed region in (B).

(D and E) Histogram of the distribution of interactions of AID and the RNA exosome complex was calculated in the (D) B cell nucleus and (E) in the cytoplasm, by MATLAB (2014b, MathWorks) software.

(F) Comparison of the distribution of paired interaction of AID and RNA exosome in the nucleus and cytoplasm were calculated by using a Student's t test in MATLAB (2014b, MathWorks) software and p values are noted in the graph. All of the 3D STORM imaging was performed in three different B cells (from independent experiments) and repeated three or more times. 3D STORM super resolution image magnification is  $\times 100$ .

See also Figure S3 and Table S1D.

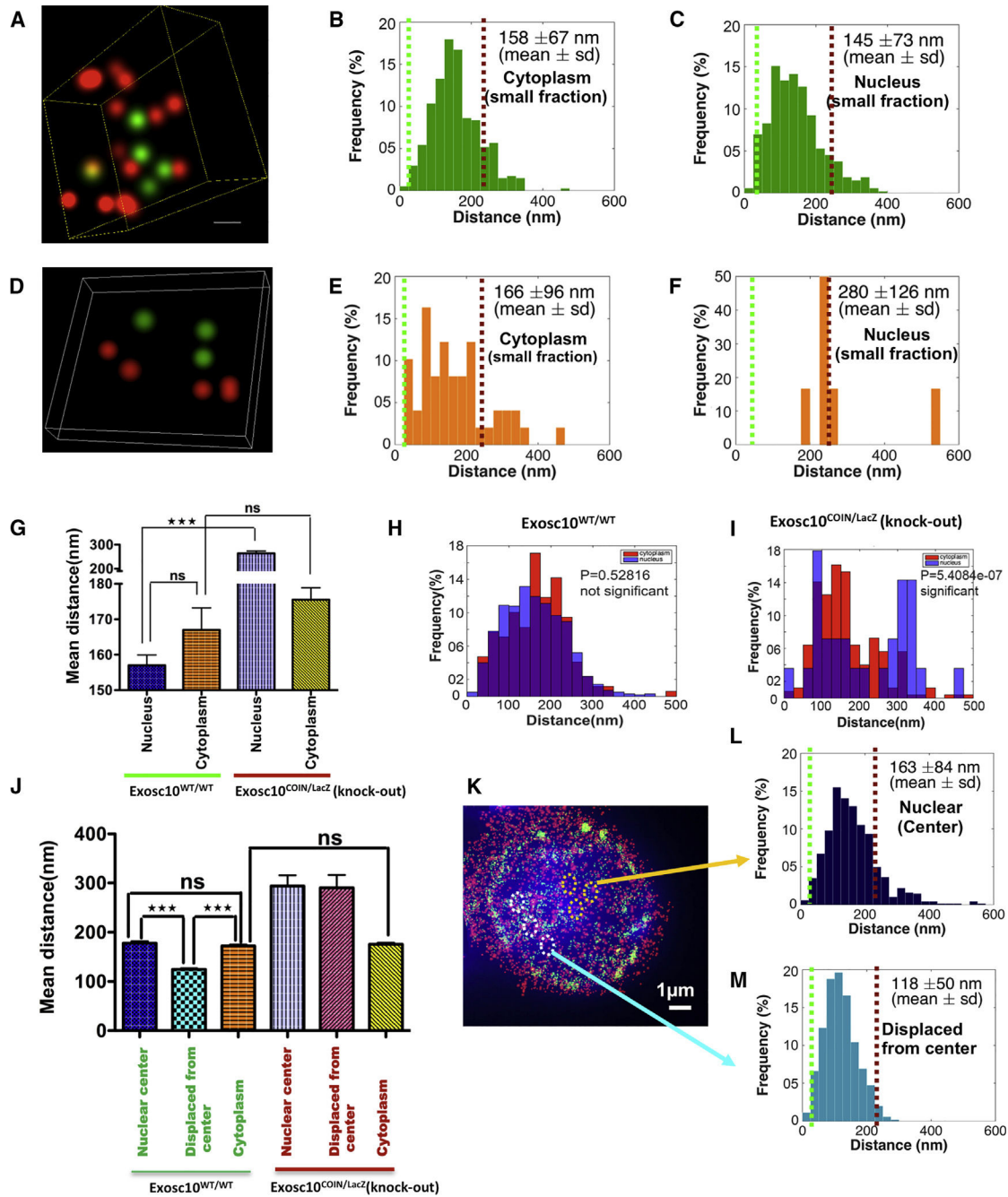
exosome are located  $45 \pm 26$  nm from each other in the nucleus (Figure 2D) versus a distance of  $68 \pm 33$  nm in the cytoplasm (Figure 2E). As shown in Figures 2F and S2A–S2E, the mean proximity of two interacting proteins is 36 nm and of two non-interactors is 205 nm. Thus, RNA exosome and Mtr4 may form a complex both in the nucleus and cytoplasm, but based on proximity analysis, the conformation and constitution of the complex is likely to be different and may employ alternative mechanisms for ncRNA processing in both compartments.

#### AID in Close Proximity to the RNA Exosome Complex in the Nucleus of Activated B Cells

We postulated that functional requirements for Mtr4 and RNA exosome proximity could be fruitfully investigated in the context of AID-mediated DNA mutagenesis and coupled ncRNA degradation mechanisms. We evaluated the proximity of AID and Exosc3<sup>TAP</sup> in stimulated B cells. AID localizes predominantly in the cytoplasm and in various regions of the nucleoplasm of activated B cells. When analyzed at single molecule resolution, nuclear AID was in close proximity to the RNA exosome complex (Figures 3A–3C). Nearest neighbor analyses show that AID and RNA exosome are separated by  $68 \pm 34$  nm in the nucleus (Fig-

ure 3D) but  $143 \pm 64$  nm in the cytoplasm (Figure 3E). An independent set of experiments is shown in Figures S3A–S3F, where the nuclear proximity of RNA exosome with AID was found to be  $64 \pm 36$  nm (Figure S3D) versus a cytoplasmic proximity of  $147 \pm 66$  nm (Figure S3E). An overlap analysis of the nuclear and cytoplasmic proximities between AID and RNA exosome demonstrates that the two factors have the tightest proximity in the nucleus of an activated B cell (Figures 3F and S3F; controls staining shown in Figure S3G). We do note that more than one AID molecule has the potential to be present in a nuclear complex, bridged by other factors like protein (Keim et al., 2013), DNA (Goodman et al., 2007), or ncRNAs (Pefanis et al., 2014), given that the distance of two AID molecules in the nucleus is close (95 nm) compared to that seen in the cytoplasm (135 nm) (Figure S3H). It should be noted that the distribution of nuclear AID in its physiological environment (B cells) (Figure S3I) differs from that seen in conditions of transient overexpression in the non-physiological environment of HEK293T cells (Figure S3J). The separation of AID/RNA exosome in the nucleus of 293T cells is  $94 \pm 47$  nm and in the cytoplasm is  $114 \pm 57$  nm (Figures S3K–S3M). Taken together, AID and RNA exosome exist in closest proximity in the nucleus of B cells, and this B cell-specific





**Figure 4. Analysis of Spatial Distribution of AID and Mtr4 in the Nucleus of B Cells Isolated from an *Exosc10*<sup>COIN/LacZ</sup> Mouse**

B cells were harvested from the spleen of an *Exosc10*<sup>COIN/LacZ</sup> mouse and fixed and prepared for 3D-STORM after 72 hr of treatment with (1) stimulation cocktail and (2) 4OHT + stimulation cocktail. Reconstructed two color 3D STORM (super-resolution) image from a dataset of 50,000 frames with Atto488-labeled AID, AlexaFluor647-labeled Mtr4, and DAPI-labeled nucleus.

(A and D) Three-dimensional views of the boxed region show spatial distribution of AID and Mtr4 molecules inside the nuclei of B cells isolated from (A) wild-type cells and (D) *Exosc10* knockout cells.

(B–F) Histogram of the distribution of interactions of AID and Mtr4 calculated in the B cell nucleus of (B) wild-type and (E) *Exosc10* knockout cells and in the corresponding cytoplasm (C) and (F) by using MATLAB (2014b, MathWorks) software.

(G) Comparison of the distribution of paired interaction of AID and Mtr4 in the nucleus versus cytoplasm by one-way ANOVA (Tukey-Kramer test) method in MATLAB (2014b, MathWorks) software. Comparison of the distribution of paired interaction of AID and Mtr4 were calculated in the cytoplasm versus nucleus of (H) wild-type and (I) *Exosc10* knockout B cells using a Student's t test in MATLAB (2014b, MathWorks) software and p values are noted in the graph.

(J) Comparison of the distribution of paired interaction of AID and Mtr4 in the nuclear center versus displaced from center versus cytoplasm by one way ANOVA (Tukey-Kramer test) method in MATLAB (2014b, MathWorks) software.

(legend continued on next page)



AID/RNA exosome proximity is not observed in a non-B cell environment.

### AID and Mtr4 Proximity Is Determined by RNA Exosome Activity in the Nucleus of B Cells

We wished to determine whether RNA exosome's RNA processing activity co-localizes Mtr4 and AID in the context of DNA mutagenesis events. It is entirely possible that RNA exosome's RNase activity is not related to Mtr4 and/or AID. To evaluate the role of functional RNA exosome in promoting the proximity of AID and Mtr4 and eventual complex formation, we performed 3D-STORM analyses in RNA exosome activity proficient and deficient B cells. We used a B cell that has a conditional allele for the *Exosc10* gene, a nuclear RNase subunit whose function is important for nuclear RNA exosome function but not for cytoplasmic RNA exosome function (Johnson and Jackson, 2013). As shown in Figures 4A–4C, the distance between AID and Mtr4 in nuclear RNA exosome activity proficient B cells is 145 nm in the nucleus and 158 nm in the cytoplasm. However, following inactivation of the nuclear RNA exosome activity (*Exosc10<sup>COIN/LacZ</sup>*), the distance of AID and Mtr4 is increased in the nucleus (280 nm) but retained in the cytoplasm (166 nm). As shown in Figure 4G, from 3D-STORM analyses of three individual B cells, the proximity of AID and Mtr4 is perturbed in *Exosc10*-deficient B cells in the nucleus but not in the cytoplasm. Statistical analyses show that the change in AID-Mtr4 proximity in the nucleus versus the cytoplasm is only significant in the *Exosc10*-deficient B cells (Figure 4I) but not in the *Exosc10*-proficient B cells (Figure 4H). Moreover, in activity-proficient B cells (*Exosc10<sup>WT/WT</sup>*), the proximity of Mtr4 and AID in the nuclear center is 163 nm (Figures 4J–4L) and in other parts of the nucleoplasm (labeled as nuclear center-displaced) is 118 nm (Figures 4J, 4K, and 4M); the loss of *Exosc10* activity leads to an increase in the proximity of Mtr4 and AID, both in the nuclear center and in the other parts of the nucleoplasm (Figure 4J; see Figures S4A and S4B for representative examples of nuclear center and nuclear center-displaced regions in Figure 4J). Finally, in *Exosc10<sup>COIN/LacZ</sup>* B cells, cellular expression of both AID and Mtr4 is robust (Figure S4E). Taken together, these observations unravel the role of nuclear RNA exosome activity in promoting the proximity of Mtr4 and AID in the nucleus, presumably driven by functional requirements to degrade newly transcribed noncoding RNAs expressed from various loci. The small but consistent difference in the proximity of AID-Mtr4 in the nuclear center versus at regions displaced from the center in the nucleoplasm implies subtle changes in the content or interaction-context of the RNA exosome complex in different subnuclear compartments. Microscopy images and movies of Mtr4 and AID localization in RNA exosome nuclear activity proficient (Figures S4A and S4C) and deficient B cells (Figures S4B and S4D) are presented.

### RNA Exosome, Mtr4, and Senataxin Cooperate to Promote DNA Antisense Strand Mutagenesis and DNA Double Strand Break-Dependent Class Switch Recombination

We generated knockout mutants of AID (*AID<sup>mut</sup>*), Mtr4 (*Mtr4<sup>mut</sup>*), and senataxin, a transcription complex-associated RNA/DNA helicase (*Setx<sup>mut</sup>*) (Hamperl and Cimprich, 2014). *Setx* is a well-known RNA/DNA helicase whose function is ubiquitous; this prompted us to evaluate Mtr4's role in DNA mutagenesis events in conjunction with *Setx*. The loss-of-function mutants were generated by introducing frameshift mutations in the coding regions of the respective genes in CH12F3 B cell lines. Frameshift mutations were introduced in exon 2 of *AID*, exon 1 of *Setx*, and exon 2 of *Mtr4* (top, middle, and bottom panels, respectively, in Figure S5A). The expression of the AID protein (Figure S5B) and cellular proliferation (Figure S6A) of the *Setx<sup>mut</sup>* and *Mtr4<sup>mut</sup>* cell lines were comparable with control “parental” CH12F3 cells. Following stimulation of parental CH12F3 cells with stimuli cocktail, CSR was observed in ~20%–30% of the cellular population. As expected, loss of AID expression in *AID<sup>mut</sup>* cells led to complete abolition of CSR (Figure 5A). Loss of Mtr4 led to a decrease in CSR efficiency whereas loss of *Setx* had a mild effect on CSR efficiency (Figures 5A and S5F). The CSR defect in the *Mtr4<sup>mut</sup>* CH12F3 cells is partially rescued by the transient transfection of an Mtr4-expressing vector in the cells (Figure S5D). We generated double mutants of cells that lack both Mtr4 and *Setx* (*Mtr4<sup>mut</sup>/Setx<sup>mut</sup>*) and found that these cells are remarkably defective in CSR (Figures 5B and 5C). The lack of Mtr4 protein expression in the *Mtr4<sup>mut</sup>/Setx<sup>mut</sup>* was confirmed by western blotting (Figure S6C) and the presence of *Setx* inactivating mutations was detected by sequencing analysis (Figure S6B). Numerous repeats were performed for each experiment, with eight independent clones of the *Mtr4<sup>mut</sup>/Setx<sup>mut</sup>* employed in Figure 5C to generate a comprehensive analysis of the roles of these proteins in catalyzing CSR. In addition, we also demonstrated that at different time points after CSR stimulation, IgM to IgA CSR was affected in the *Mtr4<sup>mut</sup>/Setx<sup>mut</sup>* CH12F3 cells (Figure S5C). Taken together, we conclude that the RNA helicase activity of Mtr4 and *Setx* is important for class switch recombination events in the *IgH* locus of B cells.

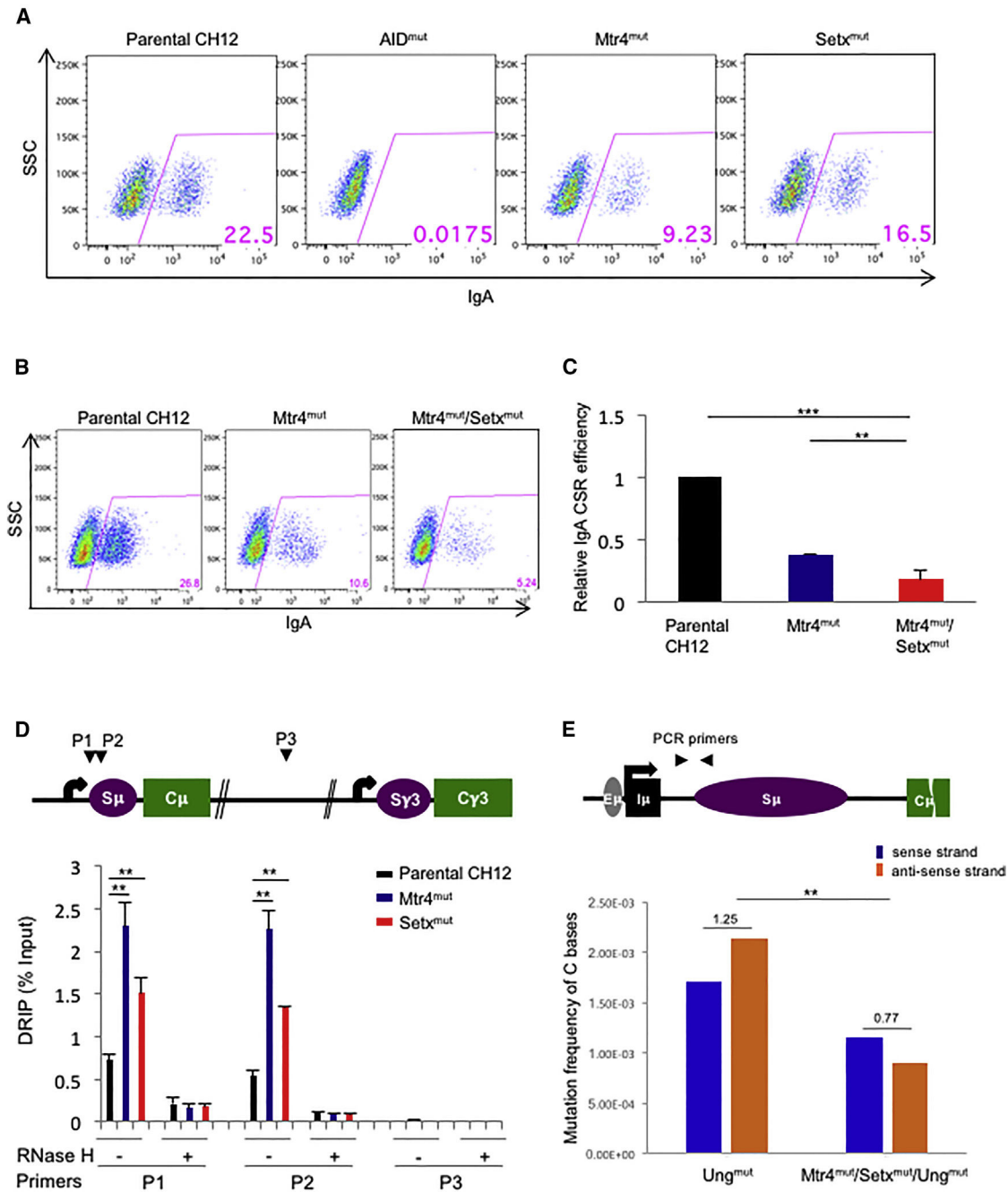
RNA exosome loss leads to increased ssDNA structures known as R-loops at regions of DNA double strand break formation in the B cell genome (Pefanis and Basu, 2015; Pefanis et al., 2014, 2015). In the 5' regions of the *IgH* locus switch sequences, we employed the DNA/RNA hybrid immunoprecipitation (DRIP) technique and found enhanced R-loop accumulation following Mtr4 or *Setx* deletion (Figure 5D). Concomitantly, following deletion of the *Ung* gene from these cells to improve mutational detection (details in Figures S6D–S6F), we found an increased ratio of somatic mutation frequency on the sense (nontemplate) DNA strand and a decrease of mutation frequency on the

(K) Reconstructed two-color 3D STORM (super-resolution) image with Alexa488-labeled AID, AlexaFluor647-labeled Mtr4, and DAPI-labeled nucleus of wild-type *Exosc10* cells. Scale bar, 1  $\mu$ m.

(L and M) Histogram of the distribution of interactions of AID and Mtr4 calculated in (L) nucleus center and (M) displaced from center of wild-type cells, by using MATLAB (2014b, MathWorks) software. All of the 3D STORM imaging was performed in three different B cells (from independent experiments) and repeated three or more times. 3D STORM super resolution image magnification is  $\times 100$ . Error bars indicate SD (\*\*p value <0.01; \*\*\*p value <0.001).

See also Figure S4 and Table S1D.





**Figure 5. Class Switch Recombination Efficiency and DNA-RNA Hybrid Accumulation and Strand-Specific Mutation Distribution at *IgH* 5' S $\mu$  Region in *Mtr4*- and *Setx*-Deficient B Cells**

(A and B) Fluorescence-activated cell sorting (FACS) assessment of CSR to IgA of *AID*, *Mtr4*, and *Setx* (single) mutants (A) and *Mtr4*/*Setx* double mutant CH12F3 clone (B). Cells were treated for 56 hr with CSR-stimulating cytokines.

(C) IgA CSR efficiency of *Mtr4* and *Mtr4*/*Setx* mutants was normalized based on CSR levels of parental CH12F3. These results were analyzed from three independently repeated experiments.

(D) Schematic map localizing the *IgH* locus and binding sites of the primer pairs used for DNA-RNA hybrid immunoprecipitation (DRIP) assays. DNA-RNA heteroduplexes were precipitated with the S9.6 antibody and normalized on input. RNase H-treated DNA-RNA heteroduplexes were used as a negative control for these experiments. Results were obtained from four independent experiments.

(E) Schematic map localizing the *IgH* locus and binding sites of the PCR primer pair for mutation analysis. Mutation frequency of AID-induced somatic hypermutation on sense (non-transcribed) strand and anti-sense (transcribed) strand of *IgH* 5' S $\mu$  region in *Mtr4*/*Setx*/*Ung* mutant cells compared to *Ung* mutant

(legend continued on next page)





antisense (template) DNA strand in Mtr4- and Setx-deficient B cells, and this change of strand mutation ratio was significantly different to what was seen in control *Ung<sup>mut</sup>* cells (Figure 5E). The change in ratio of mutation frequency on sense versus antisense DNA strands between the *Ung<sup>mut</sup>* and *Mtr4<sup>mut</sup>/Setx<sup>mut</sup>/Ung<sup>mut</sup>* was found to be significant, when compared between the hypermutated clones (Figure 5E). Similar effects on the ratio of DNA strand mutagenesis were also observed in *Mtr4<sup>mut</sup>* CH12F3 cells (Figure S5E), but with reduced overall mutation frequency due to Ung expression in these cells. The details of the mutations analyzed in Figures 5E and S5E are shown in Table S1C. Collectively, these analyses illustrate that removal of noncoding RNAs associated with the antisense DNA strand by RNA exosome requires the activity of Mtr4 and Setx, without which there is an increased level of ssDNA on the DNA sense strand (non-transcribed/non-template DNA). As would be expected in the case of bidirectional transcription or convergent transcription at switch sequences (Meng et al., 2014; Pefanis et al., 2014; Wang et al., 2016), the possibility exists that ncRNAs associated with the antisense DNA strand are processed by the RNA exosome complex similarly to germline transcripts and provide another mechanism of antisense DNA strand AID access by R-loop formation. However, such an interpretation needs to be approached carefully, as ncRNAs that are associated with the antisense DNA strand are significantly lower in expression level than germline transcripts associated with the sense DNA strand, based on RNA sequencing (RNA-seq) data obtained both in the 5'Smu region as well as in the intronic I $\mu$  and E $\mu$  regions (Figures S6G and S6H). Both possible mechanisms are outlined in Figure 7E and various mechanistic possibilities discussed later.

### ***c-Myc* Locus Strand-Specific Mutagenesis Depends upon the Activity of Mtr4 and Setx in Unwinding Antisense RNAs**

B cell lymphomas like DLBCL (diffuse large B cell lymphoma), MM (multiple myeloma), and CLL (chronic lymphocytic leukemia) are caused by inappropriate AID mutagenic activity at various locations in the B cell genome (Basso and Dalla-Favera, 2015). One example of mutagenesis in B cells is in the first intron of the *c-Myc* locus where a strongly expressed antisense RNA that is a substrate of nuclear RNA exosome activity is found (Figure 6A). This antisense RNA can be detected easily by RNA-seq in *Exosc3*-deficient cells (*Exosc3<sup>COIN/COIN</sup>*; Figure 6A, bottom panel) but not normal B cells (*Exosc3<sup>WT/WT</sup>*; Figure 6A, middle panel). We assessed the ssDNA formation efficiency following Mtr4 and Setx inactivation at this *c-Myc* intronic region and observed enhanced ssDNA structure formation (Figure 6B). Strikingly, a comparison of mutations in the antisense versus sense DNA strand of the *c-Myc* gene corresponding to this intronic sequence demonstrates an Mtr4- and Setx-dependent sense/antisense DNA strand mutation mechanism (Figure 6C).

We used *Ung<sup>mut</sup>* (base excision repair-deficient) and *Mtr4<sup>mut</sup>/Setx<sup>mut</sup>/Ung<sup>mut</sup>* triple mutant CH12F3 cells to study *c-Myc* intron-associated mutations. In the *c-Myc* intron, there is an increase in the ratio of mutation frequency on the anti-sense DNA strand in comparison to that seen on the sense DNA strand following deletion of Mtr4 and Setx, and this change in strand mutation ratio is significantly different to what is seen in the control *Ung<sup>mut</sup>* cells. In addition, there is a slight decrease in total mutation frequency in the *c-Myc* locus in the *Mtr4<sup>mut</sup>/Setx<sup>mut</sup>/Ung<sup>mut</sup>* cells (Figure 6C; details of mutational analysis in Table S1C). Thus, we speculate that the degradation of the exosome-sensitive antisense RNA is important for the strand-specific distribution of mutations in the *c-Myc* intronic region.

### **RNA Exosome Substrate ncRNA Transcription-Associated Antisense Strand Mutations at Other Regions of the B Cell Genome as Observed in *Mtr4* and *Setx* Mutant B Cells**

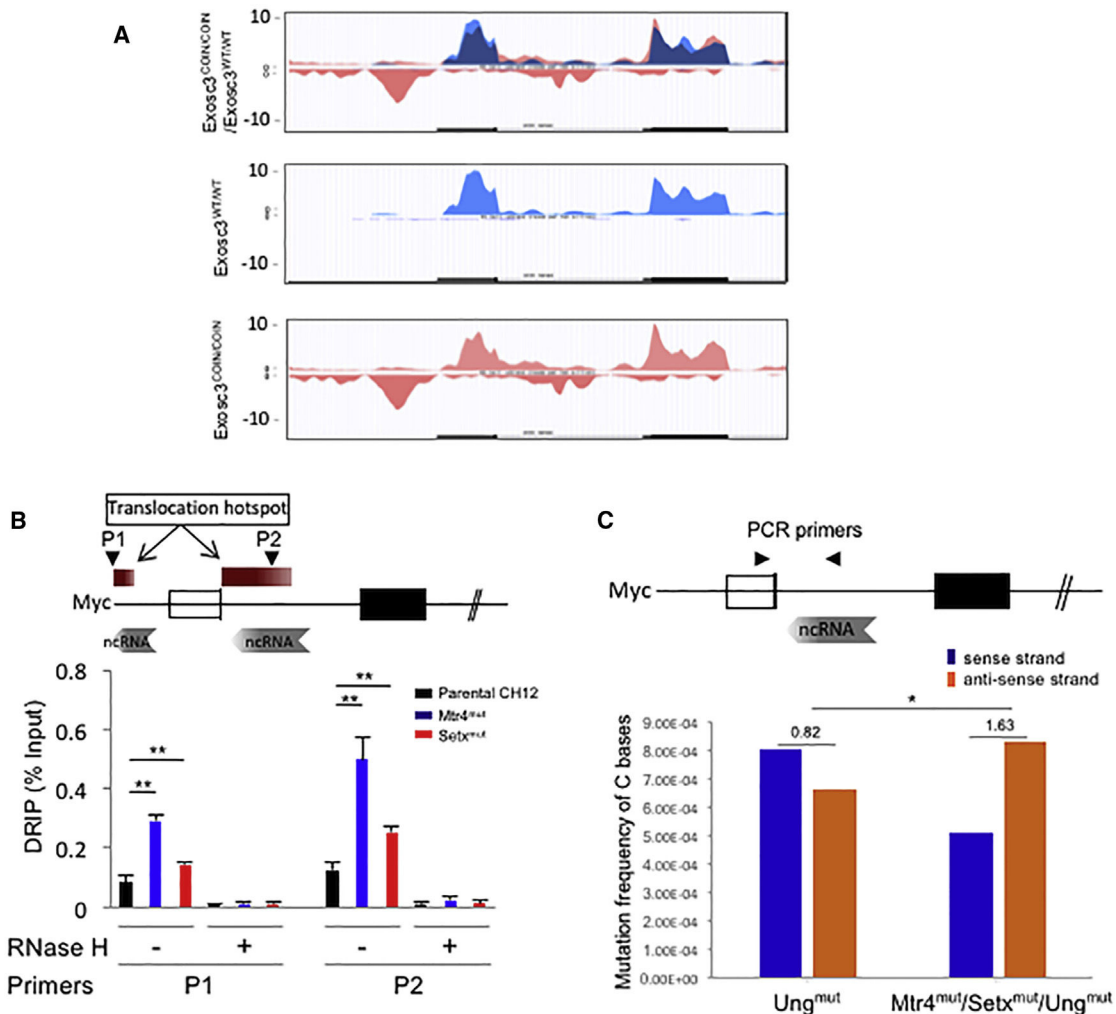
RNA exosome-sensitive RNAs have been found to overlap the sense RNA (i.e., in the same direction as the mRNA inside genic sequences) or span multiple exons in the antisense orientation inside genic sequences. We wondered if the directionality of the RNA exosome-sensitive ncRNA transcription is relevant in the determination of strand-specific mutagenesis events. In addition, the number of mutations obtained by Sanger sequencing at S $\mu$  or *c-Myc* intron was not large, and alternate assays were required to study the role of Mtr4 and Setx in controlling strand-specific mutagenesis with a cohort of a large number of mutations (see Table S1C). To address these queries, we selected two different exosome substrate ncRNAs, one in the *Pim1* locus and the other in the *CD83* locus, and investigated their role in strand mutagenesis using next generation sequencing (NGS). The exosome-sensitive RNA in the *Pim1* locus is in the antisense direction to the *Pim1* mRNA and spans multiple exons (Figure 7A; ~900 bp in length) whereas the major exosome-sensitive RNA in the *CD83* locus is in the same direction as the *CD83* mRNA (Figure 7B). Deep sequencing of either regions unravels a small mutation asymmetry on the sense and antisense strands (with respect to the directionality of the RNA exosome substrate ncRNA) in the control *Ung<sup>mut</sup>* cells, perhaps due to the loss of efficient base excision repair or due to inherent replication coupled asymmetric DNA mutagenesis mechanisms that are present in these cells. However, loss of Mtr4 and Setx leads to an increased mutation ratio on the antisense DNA strand relative to the sense DNA strand in the *Pim1* locus (Figure 7A), and this change in strand mutation ratio is significantly different to what is seen in the control *Ung<sup>mut</sup>* cells. In contrast, in the *CD83* locus where the exosome-sensitive RNA is in the same direction as the *CD83* mRNA, mutations are detectable on the DNA sense strand but not on the antisense strand (Figure 7B). In the *Pim1* or the *CD83* loci, the overall decrease in the mutation frequency following Mtr4 (and Setx) deficiency may be due to

CH12F3 cell controls. Ratio of mutation frequency between sense and anti-sense strands is indicated above bar graph. Details of mutations identified by Sanger sequencing in Table S1C.

Error bars indicate SD (\*p value <0.05, \*\*p value <0.01, \*\*\*p value <0.001).

See also Figures S5 and S6.





**Figure 6. DNA-RNA Hybrid and Strand-Specific Mutation Distribution at the *cMyc* Locus in the Absence of *Mtr4* and *Setx***

(A) RNA expression profile of the *cMyc* locus in WT (*Exosc3*<sup>WT/WT</sup>; middle panel), *Exosc3* k/o (*Exosc3*<sup>COIN/COIN</sup>; bottom panel) cells, and merged (top panel). RNA exosome substrate antisense RNA shown in pink (bottom panel). Obtained from ESC transcriptome.

(B) Schematic map localizing the *cMyc* locus and binding sites of the primer pairs used for DRIP assays. DRIP assay to determine ssDNA structure in the *cMyc* promoter and intron 1 in *Mtr4*<sup>mut</sup> and *Setx*<sup>mut</sup> CH12F3 cells. This experiment was performed as described for Figure 5D.

(C) Schematic map localizing the *cMyc* locus and binding sites of the PCR primer pair for mutation analysis. Mutation frequency of AID-induced somatic hypermutation on sense (non-transcribed) strand and anti-sense (transcribed) strand with respect to *cMyc* anti-sense RNA transcription in (A) *Mtr4*/*Setx*/Ung mutant CH12F3 cells compared to Ung mutant CH12F3 cell control. Ratio of mutation frequency between sense and anti-sense strand is indicated above bar graph. Details of mutations identified by Sanger sequencing in Table S1C.

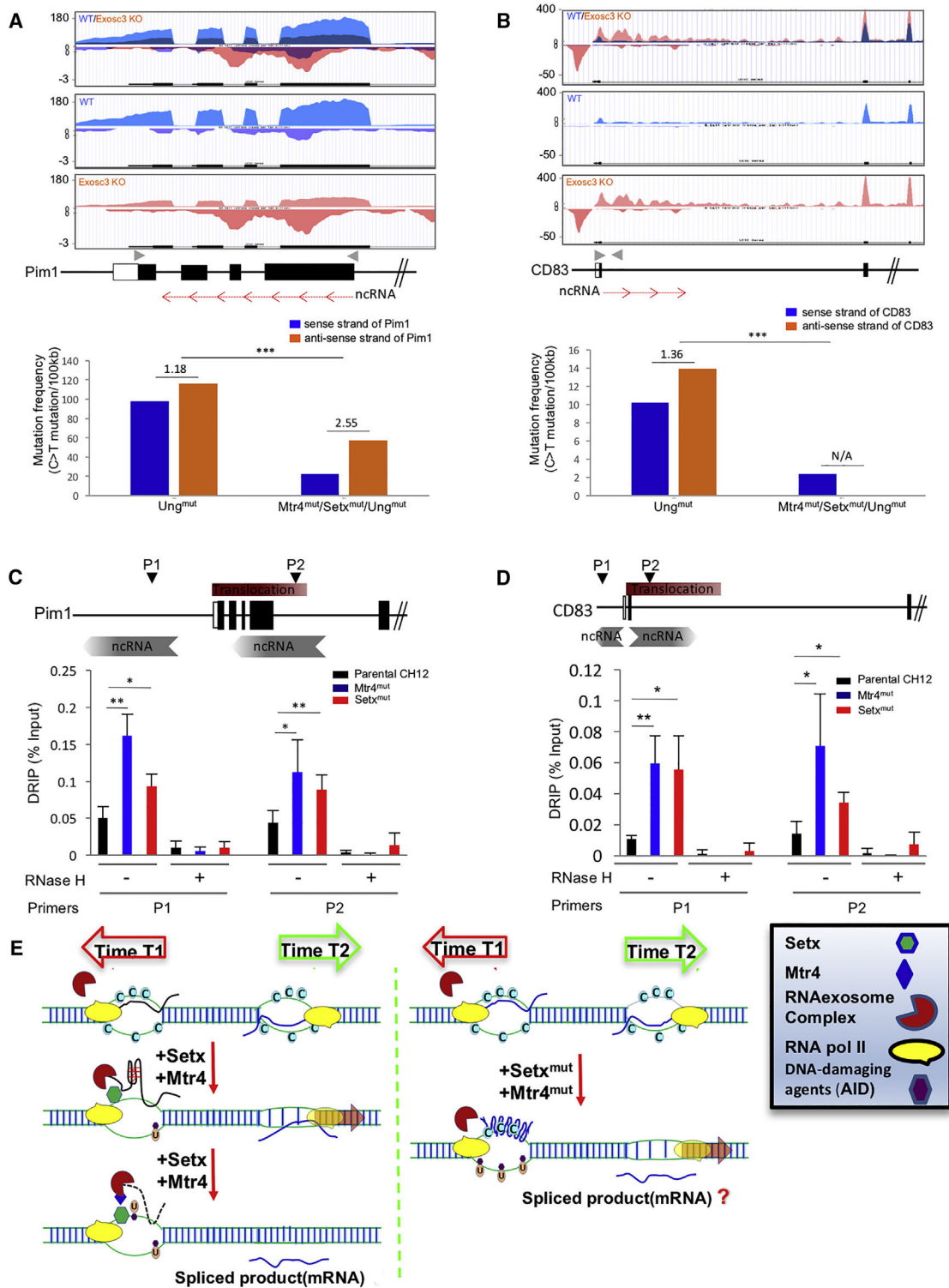
Translocation hotspots are from human and/or mouse B cells. Error bars indicate SD (\*p value <0.05, \*\*p value <0.01).

See also Figure S6.

inefficient RNA exosome-mediated AID targeting, as was shown earlier (Pefanis et al., 2014; Wang et al., 2014). When assayed for RNA/DNA hybrid accumulation, both the relevant region of *Pim1* (Figure 7C) and *CD83* (Figure 7D) accumulate DNA/RNA hybrids in *Mtr4*<sup>mut</sup> or *Setx*<sup>mut</sup> CH12F3 cells. Thus, it is likely that the RNA exosome substrate RNA unwinding in both the *Pim1* and *CD83* loci requires the activity of *Mtr4* and *Setx*. One query we wanted to address involves a region normally found unmutated in Ung<sup>mut</sup> B cells: will *Mtr4* and *Setx* mutations lead to the appearance (or abolition) of specific DNA strand mutations following stabilization of exosome-sensitive RNA? To test this possibility, we

performed deep sequencing at a region upstream of the *CD79b* locus (Figure S7A) and the *CD83* locus (Figure S7B). In the case of *CD79b*, mutations appeared only in the Ung<sup>mut</sup>/*Mtr4*<sup>mut</sup>/*Setx*<sup>mut</sup> cells on the antisense strand (Figure S7A). On the other hand, detectable mutations on the sense strand at the *CD83* upstream region in the Ung<sup>mut</sup> cells were lost following stabilization of exosome-sensitive RNA (Figure S7B). Taken together, these observations indicate that the activity of RNA exosome to degrade/process nascent ncRNAs with the help of RNA/DNA helicase cofactors *Mtr4* and *Setx* in the nucleus prevents transcription-coupled strand asymmetric mutagenesis.





**Figure 7. DNA-RNA Hybrid and Strand-Specific Mutation Distribution RNA Exosome Substrate ncRNA-Expressing Region of *Pim1* and *CD83* Genes in the Absence of *Mtr4* and *Setx***

(A and B) RNA exosome substrate ncRNA expression profiles of *Pim1* (A) and *CD83* (B) loci in WT and *Exosc3<sup>COIN/COIN</sup>* mouse B cells (top). Gray arrowheads indicate primer sites for next generation sequencing (NGS) and red arrows indicate non-coding RNA and its transcription direction on the schematic map of *Pim1* and *CD83* loci (middle). Mutation frequency of C bases on sense and anti-sense strands (with respect to *Pim1* or *CD83* gene transcription in top panel) and ratio of

(legend continued on next page)



## DISCUSSION

DNA strand asymmetric mutagenesis of mammalian genomes may be modulated by various unknown mechanisms. Here, we explore the previously unappreciated role of DNA/RNA unwinding followed by 3' end processing/degradation as a major pathway. We report that the noncoding RNA degradation complex, RNA exosome, utilizes the activities of RNA helicases Mtr4 and Setx to prevent ssDNA structures like R-loops and to prevent an overwhelming non-template strand DNA mutation burden. The eventual effects of the restriction of strand asymmetric mutagenesis include proper programmed DNA DSB-dependent DNA recombination events as seen for the *IgH* locus of B cells (failure of which may lead to genomic instability).

The regulation of many biological processes requires exquisite control over the formation of multi-molecular complexes in specific regions of the cell. Importantly, in these systems, a single protein can evoke different responses depending on its subcellular localization and the interacting partners present in a particular cellular niche. In this context, our study regarding RNA exosome/Mtr4 association and its functional interaction with DNA mutator AID sheds light on time and space restricted protein complex formation in the nucleus. Following stimulation of B cells, RNA exosome undergoes a dramatic migration to the nucleus of B cells (Figure 1). In the nucleus, it encounters Mtr4 RNA helicase and coordinates efficient strand-specific DNA mutagenesis by properly regulating levels of single strand DNA structure formation by post transcriptionally regulating ncRNA levels and associated R-loop formation (Figures 2 and 5–7). The *IgH* locus undergoes DNA mutagenesis and recombination in the nuclear center (Chan et al., 2013) or slightly displaced from the nuclear center of B cells (Holwerda et al., 2013; Yang et al., 2005). Displacement from the nuclear center of the *IgH* locus could be due to many factors including DNA motion that is associated with passage of time as chromosomes accomplish the long distance DNA interactions required during recombination (Lucas et al., 2014). AID is recruited into the nucleus of B cells during the mitotic phase of the cell cycle (Wang et al., 2016) and mutates DNA predominantly in the early G1 phase (Rush et al., 2004; Wang et al., 2016). We postulate that as B cells recover from M phase and enter G1 phase, chromosome 12 (containing *IgH*) associates with AID and complexes with RNA exosome, Mtr4, and other relevant factors to orchestrate CSR. It is unlikely that the nuclear lamina associates with the *IgH* locus, as it would lead to extreme repression of transcription-dependent DNA recombination (Zullo et al., 2012). Based on these possibilities, we propose that a localization of Mtr4 in close proximity

to RNA exosome and to DNA mutator AID creates an exquisite subnuclear space that orchestrates regulated, restricted, and strand-specific DNA mutagenesis in B cells.

Bidirectional transcription (i.e., staggered sense and antisense transcription occurring over a genomic locus) may occur at many regions of the mammalian genome, but the biological function and collateral damage of such transcription on genome integrity is underappreciated. Many possibilities of how bidirectional transcription may determine DNA strand-specific mutations are suggested by our study. First, in switch sequences, transcription of exosome-sensitive ncRNA occurs both in the sense and antisense orientations and plays a role in mutagenesis on both DNA strands via ssDNA structure formation, but this effect should be proportional to the relative levels of transcription on each strand of the template DNA. As is the case for *IgS $\mu$* , levels of canonical germline transcripts are many fold higher than levels of ncRNAs transcribed in the antisense direction (Figures S6G and S6H), making the regulation of germline RNA transcription the primary determining factor for switch sequence mutagenesis. Thus, in loci that express RNA exosome-sensitive ncRNAs on sense and antisense strands, it is possible that the relative levels of one transcript will determine the frequency of mutation on either DNA strand. Second, as in the *c-Myc* intronic region (Figure 6) or in the *Pim 1* locus (Figure 7A), sense mRNA transcription may play a less significant role in DNA mutagenesis via R-loop formation because mRNAs are rarely transcriptionally stalled and rapidly excluded from the DNA template during transcription by various mechanisms including co-transcriptional splicing (Bentley, 2014; Henriques et al., 2013; Schlackow et al., 2017; Wahba et al., 2011). However, the overlapping antisense *c-Myc* transcript or *Pim1* ncRNA transcript requires unwinding and degradation by the Mtr4/Setx/RNA exosome pathway as otherwise it will increase the asymmetric DNA mutagenesis burden through R-loop formation and stabilization (Figure 7E, right hand side). Thus, it is possible that at genomic coordinates that have overlapping mRNA and ncRNAs, the metabolism of ncRNAs has a dominant role in determining DNA mutagenesis. Third, RNA exosome substrate ncRNAs are not necessarily expressed in the antisense orientation to gene mRNAs. In the case of the *CD83* promoter region (Figure 7B), it can be seen that a ncRNA in the sense orientation, i.e., in the same direction as the mRNA is transcribed, is responsible for creating DNA mutations and in the absence of Mtr4 and Setx unravels an increased asymmetry. Thus, ncRNAs overlapping mRNAs may have a role in creating genomic instability, if not unwound from DNA by Mtr4/Setx and degraded by RNA exosome. Fourth, intergenic regions in the genome that only have

sense strand mutation frequency compare to anti-sense (transcribed) strand mutation frequency in *Ung* mutant and *Mtr4/Setx/Ung* mutant CH12F3 cells. Ratio of mutation frequency between sense and anti-sense strand is indicated above bar graph (bottom). (A) is obtained from ESC transcriptome.

(C and D) Schematic map localizing the *Pim1* (C) and *CD83* (D) loci and binding sites of the primer pairs used for DNA-RNA hybrid immunoprecipitation (DRIP) assays. This experiment was performed as in Figure 5D.

(E) A model elucidating the role of Mtr4 and Setx in restricting ssDNA structure formation at various regions in the genome that have antisense RNA transcription or sense/antisense-paired RNA transcription. Time 1 and time 2 represents two separate time points at which two RNAs are expressed inside a genic locus (e.g., *Pim1* or *CD83*). The RNA processing of the two RNAs are potentially through two different mechanisms (e.g., splicing and RNA degradation, respectively). It is to be determined whether the efficient processing of the exosome-sensitive ncRNA is important for proper mRNA biogenesis.

Translocation hotspots are from mouse B cells. Error bars indicate SD (\*p value <0.05, \*\*p value <0.01, \*\*\*p value <0.001). Details of mutations identified by NGS deep sequencing are presented in Table S1C.

See also Figures S6 and S7.





antisense RNA expression have very low but detectable mutagenesis levels on the sense DNA strand in Mtr4/Setx-deficient cells (Figures S7A and S7B), thus indicating the role of these RNA helicases in promoting RNA exosome-mediated removal of ncRNAs without which strand-specific mutations can ensue.

Considering our study as a whole, we propose a new role of RNA exosome, Mtr4, and Setx in controlling the pattern of programmed somatic mutagenesis and preventing enhanced and aberrant strand asymmetric mutagenesis at various other regions which, when not resolved, may create genomic instability. The widespread distribution of mutational asymmetry in cancer genomes originating from various cellular lineages has recently been described (Haradhvala et al., 2016). From a more global perspective, large cohorts of cancer genomes have been shown to demonstrate either transcription-associated asymmetric distribution of mutagenesis (T-mutations) or replication-associated mutagenesis (R-mutations) (Haradhvala et al., 2016); the possibility exists that ncRNA transcription-associated ssDNA structure (shown here) can contribute toward replication asymmetry. More specifically, lung carcinomas harbor an overload of G oxidation-driven mutations on the sense strand of various coding genes (Pfeifer and Hainaut, 2003), bladder cancer samples accumulate G/C mutation asymmetry associated with replication (Haradhvala et al., 2016), and liver hepatocellular carcinomas demonstrate sense strand mutations on A residues (Alexandrov et al., 2013a) (Figure S7C). Some cancer genomes originating from B cells also demonstrate detectable strand asymmetry in the mutagenesis pattern, for example those obtained from multiple myeloma patients (Figure S7D) (Lohr et al., 2012, 2014). These observations imply that strand asymmetric mutagenesis may play an important role in cancer initiation or progression. Therefore, control of the spreading of strand asymmetric mutagenesis should be managed by overlapping and/or distinct mechanisms and our study provides insight into one such pathway.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Mouse Models
  - Cell Culture Conditions
- **METHOD DETAILS**
  - *Exosc3*<sup>TAPtag</sup> Allele Construction
  - Cell Culture and IgH Class Switch Recombination Analysis
  - Purification of RNA Exosome Complex Binding Proteins
  - Sample Preparation for Mass Spectrometry
  - IP/Western
  - Cloning
  - Protein Expression and Purification
  - Molecular Beacon-Based Helicase Assay for Mtr4
  - 3D-STORM Super-Resolution Sample Preparation
  - Controls Used to Set Up 3D-STORM

- RT-qPCR
- DRIP
- CRISPR Cas9
- Proliferation Assay
- Somatic Mutation Experiment and Analysis for Sanger Sequencing
- Next-Generation Sequencing Experiments
- Next-Generation Sequencing Analysis
- Analysis of Somatic Mutation in Cancer
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - 3D-STORM Super-Resolution Data Statistical Analysis
- **ADDITIONAL RESOURCES**

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, two tables, and six movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2017.03.043>.

## AUTHOR CONTRIBUTIONS

J.L. generated CRISPR/Cas9 genome edited clones, performed DRIP, immunoprecipitation (I.P.), mutation analyses studies, and mass spectrometry data analyses. P.K.G. performed, analyzed, and stored 3D-STORM data, I.P., and assessed Mtr4's DNA/RNA unwinding activity. D.K. studied Setx mutants. V.G. and E.Pefanis generated *Exosc3*<sup>TAP</sup> mouse. E.L. and R.R. performed analyses for Figures S7C and S7D. Y.C. and H.D. performed mass spectrometry on *Exosc3*<sup>TAP</sup> complex. L.M.B. advised on mass spectrometry. W.Z. performed various statistical analyses in Figures 5–7 and S5–S7. O.M., F.B., B.L., M.C., and E.Pinaud, performed NGS assays and subsequent strand mutation analyses in Figure 7A, 7B, S7A, and S7B. U.B. conceived, directed, and analyzed experiments. U.B. (with J.L.) wrote the manuscript.

## ACKNOWLEDGMENTS

We thank Drs. David Schatz (Yale University) for insightful comments, Victor Lin (Herbert Irving Comprehensive Cancer Center transgenic facility) for targeting and generation of *Exosc3*<sup>TAP</sup> allele ESCs and mouse models, Dr. Tasuku Honjo (Kyoto University) for providing CH12-F3 cells, Dr. Theresa Swayne (confocal microscopy facility, CUMC), and Dr. Pierre Chambon (IGBMC and Université de Strasbourg) for ROSA:Ert2-Cre allele. This work was supported by NIAID (1R01AI099195), NIH (DP2OD008651), the Leukemia and Lymphoma Society, and the Pershing Square Sohn Cancer Research Alliance to U.B.; NIAID (F31AI098411-01A1) to D.K.; EMBO fellowship (ALTF 906-2015) to B.L.; Regeneron Pharmaceuticals to E.Pefanis; M.E.N.S.E.R. PhD fellowship of the French government to O.M.; Fondation Partenariale de l'Université de Limoges and ALURAD to F.B.; Comité d'Orientation Recherche Cancer en Limousin (FJA/NP-2014-126) and La Ligue contre le Cancer-Institut National du Cancer (INCa grant 9363; PLBIO 15-256) to E.Pinaud; and Association pour la Recherche contre le Cancer (PGA120150202338) and Agence Nationale de la Recherche (16-CE15-0019-01) to M.C.

Received: May 25, 2016  
 Revised: January 19, 2017  
 Accepted: March 27, 2017  
 Published: April 20, 2017

## REFERENCES

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Biggnell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain (2013a). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.



- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013b). Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* 3, 246–259.
- Alt, F.W., Zhang, Y., Meng, F.-L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* 152, 417–429.
- Basso, K., and Dalla-Favera, R. (2015). Germinal centres and B cell lymphomagenesis. *Nat. Rev. Immunol.* 15, 172–184.
- Basu, U., Meng, F.-L., Keim, C., Grinstead, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144, 353–363.
- Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15, 163–175.
- Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacio, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313, 1642–1645.
- Chan, E.A., Teng, G., Corbett, E., Choudhury, K.R., Bassing, C.H., Schatz, D.G., and Krangel, M.S. (2013). Peripheral subnuclear positioning suppresses Tcrb recombination and segregates Tcrb alleles from RAG2. *Proc. Natl. Acad. Sci. USA* 110, E4628–E4637.
- Dani, A., Huang, B., Bergan, J., Dulac, C., and Zhuang, X. (2010). Superresolution imaging of chemical synapses in the brain. *Neuron* 68, 843–856.
- de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* 100, 7480–7485.
- Driegen, S., Ferreira, R., van Zon, A., Strouboulis, J., Jaegle, M., Grosveld, F., Philipsen, S., and Meijer, D. (2005). A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res.* 14, 477–482.
- Goodman, M.F., Scharff, M.D., and Romesberg, F.E. (2007). AID-initiated purposeful mutations in immunoglobulin genes. *Adv. Immunol.* 94, 127–155.
- Hamperl, S., and Cimprich, K.A. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst.)* 19, 84–94.
- Haradhvala, N.J., Polak, P., Stojanov, P., Covington, K.R., Shinbrot, E., Hess, J.M., Rheinbay, E., Kim, J., Maruvka, Y.E., Braunstein, L.Z., et al. (2016). Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell* 164, 538–549.
- Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol. Cell* 52, 517–528.
- Holwerda, S.J., van de Werken, H.J., Ribeiro de Almeida, C., Bergen, I.M., de Bruijn, M.J., Verstegen, M.J., Simonis, M., Splinter, E., Wijchers, P.J., Hendriks, R.W., and de Laat, W. (2013). Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. *Nucleic Acids Res.* 41, 6905–6916.
- Januszky, K., and Lima, C.D. (2014). The eukaryotic RNA exosome. *Curr. Opin. Struct. Biol.* 24, 132–140.
- Johnson, S.J., and Jackson, R.N. (2013). Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol.* 10, 33–43.
- Keim, C., Kazadi, D., Rothschild, G., and Basu, U. (2013). Regulation of AID, the B-cell genome mutator. *Genes Dev.* 27, 1–17.
- Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nat. Rev. Genet.* 13, 204–214.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Lodato, M.A., Woodworth, M.B., Lee, S., Evrony, G.D., Mehta, B.K., Karger, A., Lee, S., Chittenden, T.W., D’Gama, A.M., Cai, X., et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350, 94–98.
- Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L., et al. (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc. Natl. Acad. Sci. USA* 109, 3879–3884.
- Lohr, J.G., Stojanov, P., Carter, S.L., Cruz-Gordillo, P., Lawrence, M.S., Auclair, D., Sougnez, C., Knoechel, B., Gould, J., Saksena, G., et al.; Multiple Myeloma Research Consortium (2014). Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer Cell* 25, 91–101.
- Lubas, M., Christensen, M.S., Kristiansen, M.S., Domanski, M., Falkenby, L.G., Lykke-Andersen, S., Andersen, J.S., Dziembowski, A., and Jensen, T.H. (2011). Interaction profiling identifies the human nuclear exosome targeting complex. *Mol. Cell* 43, 624–637.
- Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell* 158, 339–352.
- Madabhushi, R., Gao, F., Pfenning, A.R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T.X., Yamakawa, H., Pao, P.C., et al. (2015). Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* 161, 1592–1605.
- Meng, F.-L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberger, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538–1548.
- Nakamura, M., Kondo, S., Sugai, M., Nazarea, M., Imamura, S., and Honjo, T. (1996). High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int. Immunol.* 8, 193–201.
- Pefanis, E., and Basu, U. (2015). RNA exosome regulates AID DNA mutator activity in the B cell genome. *Adv. Immunol.* 127, 257–308.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514, 389–393.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P., et al. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 161, 774–789.
- Pfeifer, G.P., and Hainaut, P. (2003). On the origin of G→T transversions in lung cancer. *Mutat. Res.* 526, 39–43.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
- Rush, J.S., Fugmann, S.D., and Schatz, D.G. (2004). Staggered AID-dependent DNA double strand breaks are the predominant DNA lesions targeted to S $\mu$  in Ig class switch recombination. *Int. Immunol.* 16, 549–557.
- Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* 16, 583–597.
- Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive patterns of transcription and RNA processing for human lincRNAs. *Mol. Cell* 65, 25–38.
- Uson, M.L., Ordóñez, H., and Shuman, S. (2015). Mycobacterium smegmatis HelY is an RNA-activated ATPase/dATPase and 3'-to-5' helicase that unwinds 3'-tailed RNA duplexes and RNA:DNA hybrids. *J. Bacteriol.* 197, 3057–3065.
- Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* 44, 978–988.
- Wang, X., Fan, M., Kalis, S., Wei, L., and Scharff, M.D. (2014). A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. *Nat. Commun.* 5, 4137.



Wang, Q., Kieffer-Kwon, K.R., Oliveira, T.Y., Mayer, C.T., Yao, K., Pai, J., Cao, Z., Dose, M., Casellas, R., Jankovic, M., et al. (2016). The cell cycle restricts activation-induced cytidine deaminase activity to early G1. *J. Exp. Med.* *214*, 49–58.

Wei, P.C., Chang, A.N., Kao, J., Du, Z., Meyers, R.M., Alt, F.W., and Schwer, B. (2016). Long neural genes harbor recurrent DNA break clusters in neural stem/progenitor cells. *Cell* *164*, 644–655.

Yang, Q., Riblet, R., and Schildkraut, C.L. (2005). Sites that direct nuclear compartmentalization are near the 5' end of the mouse immunoglobulin heavy-chain locus. *Mol. Cell. Biol.* *25*, 6021–6030.

Zullo, J.M., Demarco, I.A., Piqué-Regi, R., Gaffney, D.J., Epstein, C.B., Spooner, C.J., Luperchio, T.R., Bernstein, B.E., Pritchard, J.K., Reddy, K.L., and Singh, H. (2012). DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell* *149*, 1474–1487.



## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Anti-AID	(Basu et al., 2011)	N/A
Goat polyclonal anti AID	Novusbio	NB100-93454
Anti- $\beta$ actin	Sigma-Aldrich	A1978
Anti-DNA/RNA hybrid (S9.6)	Kerafast	ENH001
Anti-Exosc3	Genway	GWB-FF795C
Anti Exosc3	Santa Cruz Biotechnology	Sc-98776
Anti-SKIV2L2 (Mtr4)	Abcam	ab70551
B220-PE	Biolegend	103208
IgG1-FITC	BD Bioscience	553443
IgA-FITC	BD Bioscience	559354
Anti-Flag M2 magnetic beads for IP	Sigma-Aldrich	M8823
Anti-Flag M2-HRP for WB	Sigma-Aldrich	A8592
Chicken polyclonal anti Flag-tag	Abcam	Ab1170
Alexa488 (Against-Goat)	Jackson Immuno Research	805-547-008
Alexa Fluor 647 (Anti-Rabbit)	Jackson Immuno Research	711-605-152
Goat antibody, ATTO 488 conjugated	Rockland Immunochemical	610-152-041
Nuclease-Free Water	QIAGEN	129115
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
LPS	Sigma-Aldrich	L4130
IL4	Peptotech	214-14
TGF $\beta$	R&D	240-B-010
cOmpete	Roche	11836170001
Tris base	Fisher	BP152-1
Streptavidin-HRP	Thermo Scientific	PI-21130
VPD450	BD Bioscience	562158
Q5 HF-Taq	NEB	M0491S
NaCl	Sigma	S5150-1L
HA-peptide	Fisher	26184
Flag-peptide	Sigma	F3290-4MG
M-280 Streptavidin Dynabeads	Life Technologies	11205D
Protein G Dynabeads	Life Technologies	10003D
Protein A/G agarose	Pierce	20423
Anti-HA agarose	Fisher	26181
M2-agarose anti-FLAG resin	Sigma	A22220-1ml
Glycerol	Sigma	G5516-1L
Dithiothreitol (DTT)	Gold Biotech	DTT25
MOPS	Sigma	M1442-500ml
Polyethylenimine (PEI)	Fisher	NC9197339
Cystamine	Sigma	30070-10G
Protease Inhibitor Cocktail tab.	Sigma	11-836-170-001
Dulbecco's Modified Eagle Medium (DMEM)	Life Technology	10565018
Paraformaldehyde	Electron Microscopy Sciences	15710
Glutaraldehyde	Fisher	50-262-10

(Continued on next page)





### Cell Culture and IgH Class Switch Recombination Analysis

Splenic B cells of TAPtag mice were isolated using CD43 magnetic beads (Miltenyi Biotec) negative selection and cultured in RPMI1640 containing 15% FBS with 20  $\mu$ g/ml LPS (Sigma-Aldrich) and/or 20 ng/ml IL4 (R&D) for 48~72 hr. CH12F3 cells were cultured in RPMI1640 containing 10% FBS and stimulated with 20  $\mu$ g/ml LPS (Sigma-Aldrich), 20 ng/ml IL4 (R&D) and 1 ng/ml TGF $\beta$  (R&D). Stimulated splenic B cells were stained with B220-PE (Biolegend) and IgG1-FITC (BD Bioscience) and stimulated mutant CH12F3 cells were stained with IgA-FITC (BD Bioscience). Cells were analyzed by using LSRII and FlowJo (BD Bioscience).

### Purification of RNA Exosome Complex Binding Proteins

The biotin and FLAG affinity tags were genetically incorporated in the endogenous locus of the Exosc3 subunit of the mouse RNA exosome (Figure S1). B cells were isolated from two TAP-tagged Exosc3, as well as, two control mouse spleens and grown in culture for 72hours, yielding approximately  $100 \times 10^6$  cells per condition. Cells were collected, washed 2X with cold PBS, pelleted and lysed with 300ul per 25-30 million cells of Nuclear Extract Buffer A (20mM HEPES, 10mM KCl, 1mM EDTA, 0.1mM Na<sub>3</sub>VO<sub>4</sub>, 0.2%(v/v) NP-40, 10%(v/v) Glycerol, supplemented with 1mM DTT, 1mM PMSF, Protease Inhibitor Cocktail1:1000). Following a 10 min incubation on ice, cells were slowly homogenized using a glass douncer (100-300 times), to achieve at least 80% cell lysis, and subsequently pelleted at 4300 g for 15 min at 4°C. The protein lysate was collected, and the cell pellet was further lysed with equal volume (1:1 volume of Buffer A) of Nuclear Extract Buffer B (20mM HEPES, 10mM KCl, 1mM EDTA, 0.1mM Na<sub>3</sub>VO<sub>4</sub>, 250mM NaCl, 20%(v/v) Glycerol, supplemented with 1mM DTT, 1mM PMSF, Protease Inhibitor Cocktail1:1000), homogenized slowly using a glass douncer (30 times), incubated with rotation for 30 min at 4°C and pelleted at 20000 g for 20 min at 4°C. The second fraction of the protein lysate was collected. In case of purification of core RNA exosome complex, 100 mM KCl was used in Buffers A and B. To increase the yield, this procedure was repeated using the Buffer A isolated protein fraction and the remaining cell pellet. The combined lysate fractions were precleared for 2 hr with 80 $\mu$ l of mouse IgG beads per sample. Immunoprecipitation was carried out using a total of 120 $\mu$ l of pre-equilibrated mouse FLAG-M2 agarose per sample. Extracts were incubated overnight at 4°C, with rotation. Subsequently, agarose was pelleted and washed 3X using Buffer A/B (1:1) – 15 min riation at 4°C, and pelleted at 300 g for 4 min. Proteins were eluted 2X - 3 hr elution with rotation each time, using 320 $\mu$ l of FLAG peptide (100ug/ul) per elution. Supernatant containing the RNA exosome complex and its interacting proteins was collected and further purified using 200 $\mu$ l of pre-equilibrated Streptavidin agarose - overnight incubation at 4°C with rotation. Subsequently, agarose was pelleted, washed 3X with Buffer A/B, and bound proteins were eluted in 400 $\mu$ l of 0.3% SDS buffer by boiling - 10 min at 95°C.

### Sample Preparation for Mass Spectrometry

First, mass spectrometry of core RNA exosome complex was performed. Tryptic peptides from digested proteins were trapped on a Symmetry C18 Trap column and analyzed in 120 min chromatograms with a 75  $\mu$ m ID x 25 cm HSS T3, 1.8  $\mu$ m particle diameter reverse phase C18 column at a flowrate of 300 nL/min with an acetonitrile/formic acid gradient (NanoAcquity UPLC, Waters Corp.). Spectra were recorded in resolution ion mobility positive ion mode with a Synapt G2 quadrupole-time-of-flight HDMS mass spectrometer (Waters Corp.). Data analysis was performed with ProteinLynx Global Server (PLGS) (Vers. 2.5, RC9, Waters Corp.) with search against a reference database of mouse protein sequences (Uniprot). Post-processing of PLGS data by accurate mass and retention time-based data mining was done Elucidator Protein Expression Data Analysis System, Version 3.3 (3.3.0.1.SP3\_CRE52.21) (PerkinElmer/Rosetta Biosoftware). Cluster analysis in Figure 1A was performed with the Elucidator software. This section is relevant to Figure 1A.

Next, mass spectrometry of RNA exosome megacomplex with associated proteins was performed. The protein complexes recovered from TAP purification were fractionated on a 4%–20% SDS-PAGE gradient gel. A zinc staining protocol was used to stain the proteins. Unknown proteins were directly excised from the gel and destained in 2% citric acid (9). In-gel digestions were performed as described previously, and modified porcine trypsin (Promega) was employed (100–200 ng/digestion) to digest the sample overnight at 37°C (10). Formic acid was added to terminate the reaction (1  $\mu$ l of 88% formic acid for every 50  $\mu$ l of digestion sample), and the gel slices were sonicated for 20 min to recover the peptides. This aqueous sample was concentrated in a speed-Vac and further purified by a C18 reverse phase Ziptip (Whatman). The peptides bound to the column were eluted in 50% acetonitrile/0.1% trifluoroacetic acid. LC-MS/MS (Thermo/LTQ-Orbitrap Velos) was used to identify the unknown proteins. The MS/MS spectra from each LC-MS/MS run were searched against the Uniprot Mouse database using SEQUEST search engine in Proteome Discoverer 1.3 software. This section is relevant to Figure 1A.

### IP/Western

Cultured splenic B cells were lysed in IP buffer (50 mM Tris [7.4], 150 mM NaCl, 1 mM EDTA, 0.1% Triton X-100) containing proteinase inhibitor (11836170001 Roche) on the ice for 20 min. Extracted protein was precipitated with FlagM2-magnetic beads (Sigma-Aldrich) or Streptavidin agarose beads (Thermo scientific) at 4°C for 16 hr. Beads were washed 3 times with IP lysis buffer and eluted. IP products were separated on the 12% SDS gel and transferred on the PVDF membrane. IP products were analyzed using FlagM2-HRP (Sigma-Aldrich), Streptavidin-HRP (Thermo Scientific) and Exosc3 antibody (Genway).



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Normal goat serum(NGS)	Jackson Immuno Research Laboratories	005-000-121
RiboLock	Fisher	EO0381
MgCl <sub>2</sub>	Sigma	M1028
MOPS buffer	Sigma	M1442
Critical Commercial Assays		
Ion Xpress Plus gDNA Fragment Library Preparation	Life Technologies	4471269
Ion Shear Plus Reagents Kit	Life Technologies	4471248
Ion Plus Fragment Library Kit	Life Technologies	4471252
CD43 magnetic beads	Miltenyi biotec	130-049-801
Mouse B Cell Nucleofector Kit	Lonza	VPA-1010
Deposited Data		
RNA-seq data	(Pefanis et al., 2014, 2015)	SRP042355
Experimental Models: Cell Lines		
CH12F3	(Nakamura et al., 1996)	N/A
129S6/SvEvTac X C57BL/6Tac hybrid mESC	Columbia U. transgenic facility	N/A
HEK293T	ATCC	ATCC-CRL-3216
Experimental Models: Organisms/Strains		
<i>Exosc3</i> <sup>TAPtag/TAPtag</sup> mouse	This paper	N/A
<i>Exosc10</i> <sup>COIN/COIN</sup> mouse	(Pefanis et al., 2015)	N/A
<i>Rosa26</i> CreERT2 mouse	Gift from Dr. Pierre Chambon (IGBMC)	N/A
<i>Rosa26</i> <sup>BirA/BirA</sup> mouse	(Drieggen et al., 2005)	N/A
Recombinant DNA		
BAC clone	Source Bioscience	bMQ386a13
pSpCas9(BB)-2A-GFP (pX458)	Addgene	48138
pGEM-T Easy	Promega	A1360
pcDNA3.1	Fisher	V79020
Sequence-Based Reagents		
Primers for sanger sequencing (see Table S2)	This paper	N/A
Primers for NGS (see Table S2)	This paper	N/A
Primers for DRIP (see Table S2)	This paper	N/A
Primers for RT-qPCR (see Table S2)	This paper	N/A
Sequences of AID gRNA: TGAGACCTACCTCTGCTACG	This paper	N/A
Sequences for Setx gRNA: AAGCGCTATGCATCTAGCAC	This paper	N/A
Sequences for Mtr4 gRNA: AATCAGAGTCAGCCAGTGGG and AAATGTGGCCCCAGTTGCCC	This paper	N/A
Sequences for Ung gRNA: CGCAACGTGCCTGCCGGCTT and TCACGGACGCGGTCGTGTCC	This paper	N/A
Primers for sub cloning of hMtr4 in to pCDNA3.1- forward primer: GTGGTCGCTAGCATG TACCCATACGATGTT CCA GATTACGCTATG GAT TCT ACT GAT, reverse primer: 5'ACGTCAGCTAGCCTATAA ATA CAA AGA3'	This paper	N/A
Primers for sub cloning of hAID in to pCDNA3.1- forward primer: 5'GTGGTCGCTAGCATG TACCCATACGATGTT CCAGATTACGCTATG GACAGCCTCTTG3'	This paper	N/A
Primers for sub cloning of hAID in to pCDNA3.1- reverse primer: 5'ACGTCAGCTAGCTCA AAG TCCCAAAGT3'	This paper	N/A
Primers for unwinding assay for Mtr4- R16-Top strand-reporter: 5'AGCACCGUAAAGACGC3'	This paper	N/A

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primers for unwinding assay for Mtr4- R22-Bottom strand: 5'GCGUCUUUACGGUGCUUAAAAA3'	This paper	N/A
Primers for unwinding assay for Mtr4- D22-Bottom strand: 5'GCGTCTTACGGTGCTTAAAAA3'	This paper	N/A
Software and Algorithms		
BWA-MEM. arXiv:1303.3997v1 [q-bio.GN]		<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
Bowtie2	(Langmead and Salzberg, 2012)	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
Samtools		<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
MATLAB		<a href="https://www.mathworks.com">https://www.mathworks.com</a>
Rstudio		<a href="https://www.rstudio.com/">https://www.rstudio.com/</a>
R		<a href="https://www.r-project.org/">https://www.r-project.org/</a>
NIS element	NIKON	N/A
FlowJo	BD Bioscience	N/A
Other		
Bladder, Liver and Lung mutational exosome data		<a href="http://gdac.broadinstitute.org">gdac.broadinstitute.org</a>
Multiple Myeloma	(Lohr et al., 2014)	<a href="https://www.ncbi.nlm.nih.gov/pubmed/24434212">https://www.ncbi.nlm.nih.gov/pubmed/24434212</a>
Illustration image	Servier Medical Art	<a href="http://www.servier.com/Powerpoint-image-bank">http://www.servier.com/Powerpoint-image-bank</a>

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Uttiya Basu ([ub2121@cumc.columbia.edu](mailto:ub2121@cumc.columbia.edu)).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS****Mouse Models**

*Exosc3*<sup>TAPtag/TAPtag</sup>, *Exosc10*<sup>COIN/LacZ</sup>, *Rosa26* CreERT2, and *Rosa26*<sup>BirA/BirA</sup> mouse models were generated and bred according to the IACUC guidelines at Columbia University, New York.

**Cell Culture Conditions**

B cells and CH12F3 cells were cultured in RPMI 1640 containing 10% FBS, 2 mM Glutamine, 55 μM β-Mercaptoethanol, 100 U/ml penicillin, 100 μg/ml streptomycin and 25 mM HEPES. HEK293T cells were cultured in DMEM containing 10% FBS, 100 U/ml penicillin and 100 μg/ml streptomycin. All cells were grown at 37°C, under an atmosphere containing 5% CO<sub>2</sub>.

**METHOD DETAILS*****Exosc3*<sup>TAPtag</sup> Allele Construction**

Mouse BAC library clone bMQ386a13 harboring the *Exosc3* locus was modified using bacterial homologous recombination. A TAP tag comprising coding sequences for the FLAG epitope tag, TEV protease cleavage site, and biotin tag (de Boer et al., 2003) was inserted in frame immediately downstream of the *Exosc3* ATG initiator codon. A neo selection cassette flanked by FRT sites was inserted into a non-conserved region of the first intron of *Exosc3*. BAC clones modified through bacterial homologous recombination were screened by junctional PCR. The 3.8 kb TAPtag-neo modification was fully sequence verified for the BAC clone used in mESC targeting. Upstream and downstream homology arms in the *Exosc3*<sup>TAPtag-neo</sup> BAC targeting vector were 41 kb and 84 kb, respectively. Gene targeting was performed in 129S6/SvEvTac x C57BL/6Tac hybrid mESC cells. Targeted clones were identified using loss of allele assay. *Exosc3*<sup>TAPtag-neo/+</sup> mESCs were microinjected into blastocysts to give rise to chimeric mice, which were subsequently crossed with Tg(ACTB:FLPe) female mice (Jackson Laboratory) to remove the neo selection cassette and germline transmit the *Exosc3*<sup>TAPtag</sup> allele. Following removal of the FLPe transgene, *Exosc3*<sup>TAPtag/+</sup> mice were crossed with *ROSA26*<sup>BirA/BirA</sup> mice (Driegen et al., 2005) to generate experimental cohorts. All mouse experiments were performed in adherence with protocols approved by the Columbia University Institutional Animal Care and Use Committee.



### Cloning

The full-length human AID and Mtr4 genes were PCR amplified using primers listed in key resource table and inserted into mammalian expression pCDNA3.1 vector (Invitrogen), which was previously linearized by the NheI restriction enzyme. HA-tag was introduced at the N terminus of the AID and Mtr4 proteins.

### Protein Expression and Purification

hMtr4 was isolated from HEK293T cells. Multiple 15cm plates of HEK293T cells were grown to confluence and maintained in Dulbecco's Modified Eagle Medium (DMEM medium supplemented with 4 mM L-Glutamine, 10% fetal bovine serum at 37°C with 5% CO<sub>2</sub>). HEK293T cells were transfected with hMtr4 construct with previously described Polyethylenimine (PEI) based transfection protocol. After 72-96 hr of transfection, HEK293T cells were removed from the plates and dissolved in lysis buffer. Cells were disrupted by a French press and the supernatant was collected after centrifuging at 10,000 g for 1 hr at 4°C. Affinity purification with agarose anti-HA resin was conducted based on principles originally described (Rigaut et al., 1999), except for solubilization procedure and buffers.

### Molecular Beacon-Based Helicase Assay for Mtr4

DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA) in purified form as lyophilized solids. They were then dissolved in DNase/ RNase-free water (QIAGEN, USA), and concentrations were determined from the extinction coefficients provided. The reporter RNA oligonucleotide (16-mer), labeled with 6-FAM at its 5' end and Iwova black at its 3' end was annealed to the bottom strand of a 22 nucleotide ssRNAs and ssDNAs at a 1:1 molar ratio in 10 mM Tris-HCl pH 8.5, placing in 95°C, and then allowing to cool to room temperature for ~1 hr to anneal. The sequences of the oligonucleotides used in the study are listed in key resources table. The Mtr4 helicase enzyme used for this study was a recombinant protein with N-terminal HA-tag. Cloning, expression, and purification of this Mtr4 protein has been described in early section of methods. Each reaction contained 40 mM MOPS pH 6.5, 100 mM NaCl, 4 mM MgCl<sub>2</sub>, 5% glycerol, 0.01% NP-40 substitute, 4 mM DTT, 2 μM substrate nucleic acids (RNA-RNA & RNA-DNA), 10 μM enzyme (Mtr4), 4 mM ATP and 1 U/μl of Ribolock (Thermo Fisher). Reactions were carried out in 100 μL, in triplicate, in white half-volume 96-well polystyrene plates at room temp. Fluorescence was measured as arbitrary units (a.u.) in each well every 5 min. using a Tecan infinite 500 fluorescence spectrophotometer [excitation/emission at 485/535 nm (5/10 nm slit width)]. Three replicates of data were analyzed using MATLAB software, and a first-order exponential decay model was used to determine the pseudo-first order rate constant.

### 3D-STORM Super-Resolution Sample Preparation

Splenic B cells from *Exosc3*<sup>TAPtag</sup>, *Exosc10*<sup>Wt/Wt</sup> and *Exosc10*<sup>COIN/LacZ</sup> mouse were prepared using CD43 microbead (Miltenyi Biotec) negative selection and cultured in RPMI 1640 containing 15% FBS, 2 mM Glutamine, 55 μM β-Mercaptoethanol, 100 U/ml penicillin, 100 μg/ml streptomycin and 25 mM HEPES. Isolated B cells were cultured with 20 μg ml<sup>-1</sup> LPS (Sigma) and 20 ng ml<sup>-1</sup> IL4 (R&D system) for three days. B Cells from day1 and day3 were immobilized on MatTek 1.5mm glass coverslips by cytospin and subsequently fixed with 3% paraformaldehyde (Electron Microscopy Sciences) + 0.1% glutaraldehyde (Electron Microscopy Sciences) for 5 min. at room temperature, washed three times in 0.1 M phosphate buffer (PB), permeabilized in 0.5% (v/v) Triton X-100 supplemented with 5% normal goat serum (NGS) in 0.1 M phosphate buffer (PB) for 10 min and immunostained with primary antibodies, anti-Mtr4 (Polyclonal anti Mtr4 antibody produced in mouse-cat.no. ab187884-Abcam), anti-AID (Basu et al., 2011) and anti-Flag-tag (Polyclonal Anti-flag tag antibody produced in Chicken, cat.no. ab1170, Abcam for TAP tagged RNA exosome complex) in PBS for 3h. The cells were then washed in PBS three times and stained with photo-switchable secondary antibodies Atto488 (Anti-Rabbit IgG(H+L), Host: Goat antibody, ATTO 488 conjugated, cat.no. 18772-Sigma) and Alexa fluor647 (Anti-Chicken IgG (H+L), Host: Donkey antibody, Alexa Fluor 647 conjugated, cat.no. 703-606-155- Jackson Immuno Research) compatible for STORM imaging for 2 hr. Cells were then washed twice each in PBS. Immediately before STORM imaging, PBS was removed and freshly prepared imaging medium buffer (Containing cystamine and glucose oxidase mixture for oxygen scavenging) (Dani et al., 2010). The two colors (Atto488/Alexa488 and Alexa flour 647) were imaged sequentially. Imaging buffer helped to keep dye molecules in a transient dark state. Subsequently, individual dye molecules were excited stochastically with high laser power at their excitation wavelength (488 nm for Atto488/Alexa488 or 647 nm for Alexa flour 647, respectively) to induce blinking on millisecond timescales. STORM images and the correlated high-power confocal stacks were acquired via a CFI Apo TIRF 100 × objective (1.49 NA) on a Nikon Ti-E inverted microscope equipped with a Nikon N-STORM system, an Agilent laser launch system, an Andor iXon Ultra 897 EMCCD (with a cylindrical lens for astigmatic 3D-STORM imaging) camera, and an NSTORM Quad cube (Chroma). This setup was controlled by Nikon NIS-Element AR software with N-STORM module. To obtain images, the field of view was selected based on the live EMCCD image under 488-nm illumination. 3D STORM datasets of 20,000 frames for HEK293T and 50,000 frames for B Cells were collected. Lateral drift between frames were corrected by tracking 488, 561, and 647 fluorescent beads (TetraSpeck, Life Technologies). STORM images were processed to acquire coordinates of localization points using the N-STORM module in NIS-Elements AR software. Identical settings were used for every image. Each localization is depicted in the STORM image as a Gaussian peak, the width of which is determined by the number of photons detected (Betzig et al., 2006). All of the 3D STORM imaging were performed in three different HEK293T/B Cells (from independent experiments) and repeated three or more times. Details of 3D-STORM microscopy associated statistical analyses, number of biological repeats and number of technical repeats are provided in Table S1D.





### Controls Used to Set Up 3D-STORM

To evaluate the non-specific binding of secondary antibodies with the cells, we collected 3D STORM images from (A) B cells (isolated from AID K/O mouse) labeled with primary antibody (anti AID) and secondary antibody (Atto 488) (Figure S3G) (B) B Cells (isolated from *Exosc3*<sup>TAPtag/TAPtag</sup>) labeled with secondary antibody (alexa fluor 647) alone (Figure S3G).

As a negative control for the localization of RNA exosome complex and its cofactors, we transfected HEK293T cells with RNA *exosc3* constructs along with an HA-tagged RFP protein (non-interacting protein with RNA exosome complex) and collected 3D STORM images (Figures S2D–S2E). In the conditions of our assays, we set closest proximity of two proteins by labeling B cells with *Exosc3* and *Exosc5* subunits of RNA exosome (Figures S2A–S2C).

### RT-qPCR

RNA was isolated by Trizol reagent (Invitrogen, 15596018) from B cells, and then DNaseI (QIAGEN, 79254) treated 1 µg RNA was used for cDNA synthesis using superscript IV reverse transcriptase (Life Technology, 18091050). Primers sequences used for PCR are in Table S2.

### DRIP

DRIP was performed following previously used protocol (Pefanis et al., 2014). WT, *Setx*<sup>mut</sup> and *Mtr4*<sup>mut</sup> CH12F3 cells were stimulated by LPS, IL4 and TGFβ for 24 hr, and then cells were collected and lysed. DNA of each cell type was purified by phenol/chloroform extraction followed by EtOH precipitation and subsequently digested using restriction enzymes BsoBI, HindIII, NheI, NcoI and StuI (New England Biolabs). DNA/RNA fragments treated with or without RNase H were precipitated by S9.6 antibody (Kerafast) with Protein A/G agarose (Thermo Scientific). Precipitated DNA was eluted and purified by phenol/chloroform extraction and EtOH precipitation. DRIP products were measured using qPCR analysis employing the primers in Table S2.

### CRISPR Cas9

Mutations in each target were generated following previous protocols (Pefanis et al., 2014). Guide RNAs were designed using the online tool (<http://tools.genome-engineering.org>) and the following guide RNA sequences were used: AID - TGAGACCTACCTCTGC TACG, *Setx* - AAGCGCTATGCATCTAGCAC, *Mtr4* - AATCAGAGTCAGCCAGTGGG and AAATGTGGCCCCAGTTGCC, Ung- CGC AACGTGCCTGCCGGCTT and TCACGGACGCGGTCTGTGCC. DNA oligos of target guide RNA sequences were inserted into BbsI site of pSpCas9(BB)-2A-GFP (Addgene), and construction of each target was transfected into CH12F3 cells by nucleofector (Lonza). Transfected cells were isolated into 96 well culture plates by cell sorter (BD FACSAria II) after incubation for 24 hr and incubated for 1 week. Genomic DNA of each clones were used as a template for PCR/sequencing screening to seek mutated clones by insertion or deletion. Primers sequences used for screening are in Table S2. PCR products of mutated clones were inserted into pGEM-T Easy (Promega) vector and their mutated regions were confirmed by sequencing analysis.

### Proliferation Assay

Proliferation of each mutant CH12F3 cells was tested using Violet Proliferation Dye 450 (BD Bioscience). Cells were collected and treated at 1 µM VPD at 37°C for 10 min. After washing with PBS, cells were cultured and/or tested by FACS for 4 days.

### Somatic Mutation Experiment and Analysis for Sanger Sequencing

UNG mutant and *Mtr4/Setx* mutant CH12F3 cells were stimulated by LPS (Sigma-Aldrich), IL4 (BD Bioscience) and TGFβ (R&D) for 48 hr, and genomic DNA of each clone was prepared for PCR. Primers (see Table S2) were used for amplifying *IgH* 5' S<sub>μ</sub> and *cMyc* intron 1 regions. PCR products were inserted into pGEMT-easy vector (Promega) and transformed into XL1 Blue competent cells. Transformed XL1 blue colonies were submitted for sequencing analysis. Specific DNA regions of Non-transcribed strand were sequenced. Raw sequences were manually reversed and complemented to match the reference if necessary. We used ClustalW to do multiple sequence alignment. Non-relevant ends of each sequence were removed by truncate the primer region include extra 5 bases. Sequence with few bases were removed. Since AID primarily affects C, we focus on mutation frequency of C and G to other bases. G mutation frequency of non-transcribed strand can be used to reflect the C mutation frequency of transcribed strand, as transcribed and non-transcribed strands are complementary. Clones without C or G mutation were excluded from further analysis. For clones that have C or G mutation, we calculate mutation frequency of C and G as follows.

$$\mu_C = \frac{\# \text{ of mutated C bases}}{\# \text{ of total C base}}$$

$$\mu_G = \frac{\# \text{ of mutated G bases}}{\# \text{ of total G base}}$$

where denominators are counted from reference sequence and numerators are counted from clonal sequence compared to reference sequence.

After the clone specific mutation frequencies are calculated, we use the ratio of C mutation frequency in anti-sense (transcribed) strand over C mutation frequency in sense (non-transcribed) strand to depict the strand mutation bias for case and control group. For



example, if  $(\mu_{C1}, \mu_{C2}, \dots, \mu_{Cm})$  and  $(\mu'_{C1}, \mu'_{C2}, \dots, \mu'_{Cm})$  are C mutation frequencies in anti-sense (transcribed) and sense (non-transcribed) strand respectively, the ratios are

$$\left( \frac{\mu_{C1}}{M}, \frac{\mu_{C2}}{M}, \dots, \frac{\mu_{Cm}}{M} \right)$$

where  $M = (1/m) \sum_{i=1}^m \mu'_{Ci}$  is the average C mutation frequency in non-transcribed strand. To test whether there is significant difference between strand mutation biases in control samples and in case samples, we employ the Wilcoxon rank sum test to access the difference between ratios in case and control samples.

### Next-Generation Sequencing Experiments

Genomic DNA was extracted 6 days (experiment 1) or 7 days (experiment 2) after stimulation with LPS/IL4/ TGF $\beta$ , from *Ung* mutant and *Mtr4/Setx/Ung* mutant CH12F3 cells or unstimulated WT CH12F3 cells as a control. Two independent sets of PCR were performed for each target using high fidelity Taq polymerase (M0491S, NEB) and following primers. 100ng Template genomic DNA was amplified in 25~28 cycle. PCR primers sequences used for amplification of non IgH loci are in [Table S2](#). PCR products were purified using spin column (28004, QIAGEN). Next Generation Sequencing was performed according to the user guide Ion Xpress Plus gDNA Fragment Library Preparation (Cat. no. 4471269, Life Technologies). Briefly, PCR products (1 $\mu$ g) were fragmented by enzymatic digestion (Ion Shear Plus Reagents Kit, Cat. no. 4471248) and ligated to Barcodes and Adapters (Ion Plus Fragment Library Kit, Cat. no. 4471252). After 200 bp size selection step on E-Gel precast agarose electrophoresis system, final amplification was performed. Libraries were sequenced on an Ion Proton System.

### Next-Generation Sequencing Analysis

Detailed pipeline analysis protocol will be published in a method article (F.B., O.M., and E.Pinaud, unpublished data). Briefly, fastq data files were aligned on reference sequences using BWA-MEM algorithm. ([arXiv:1303.3997v1 \[q-bio.GN\]](#)), and nucleotides were counted position-wise with *igvtools* ([Robinson et al., 2011](#)). Unstimulated WT CH12F3 samples were used to determine background error frequencies for each possible mutation event. Then, mutations in stimulated samples were called at each site with a nucleotide frequency significantly greater than local background ( $p < 0.001$ , fold increase  $> 2$ ). AID hotspots were defined as WRCY (or reverse complement RGYW) and C (or G) mutations were studied. Total mutations and mutation frequencies were calculated, as well as mutation ratios between C and G (corresponding to sense or anti-sense DNA strand, depending on genome region). Finally, Fisher's exact test was used to evaluate mutation differences between *Ung* mutated cells and *Mtr4/Setx/Ung* mutated CH12F3 cells.

### Analysis of Somatic Mutation in Cancer

We analyzed somatic coding mutations from whole exome data in six different control matched tumors. We considered those with coverage  $\geq 20$  reads in both tumor and normal,  $< 2\%$  variant allele frequency in the Normal and  $\geq 5\%$  in the tumor. For variants harbored in genes transcribed on the reverse strand we reverse complemented the variant nucleotides. Mutations were tallied and compared in pairs for differences in abundance. Significance was assessed using a Proportion Test. The following mutations were grouped together  $A > C/T > G$ ,  $A > G/T > C$ ,  $A > T/T > A$ ,  $C > G/G > C$ ,  $C > A/G > T$ ,  $C > T/G > A$ . Bladder, Liver and Lung whole exome sequencing data were obtained from Broad's GDAC at [gdac.broadinstitute.org](#). Multiple Myeloma data were obtained from [Lohr et al. \(2014\)](#).

## QUANTIFICATION AND STATISTICAL ANALYSIS

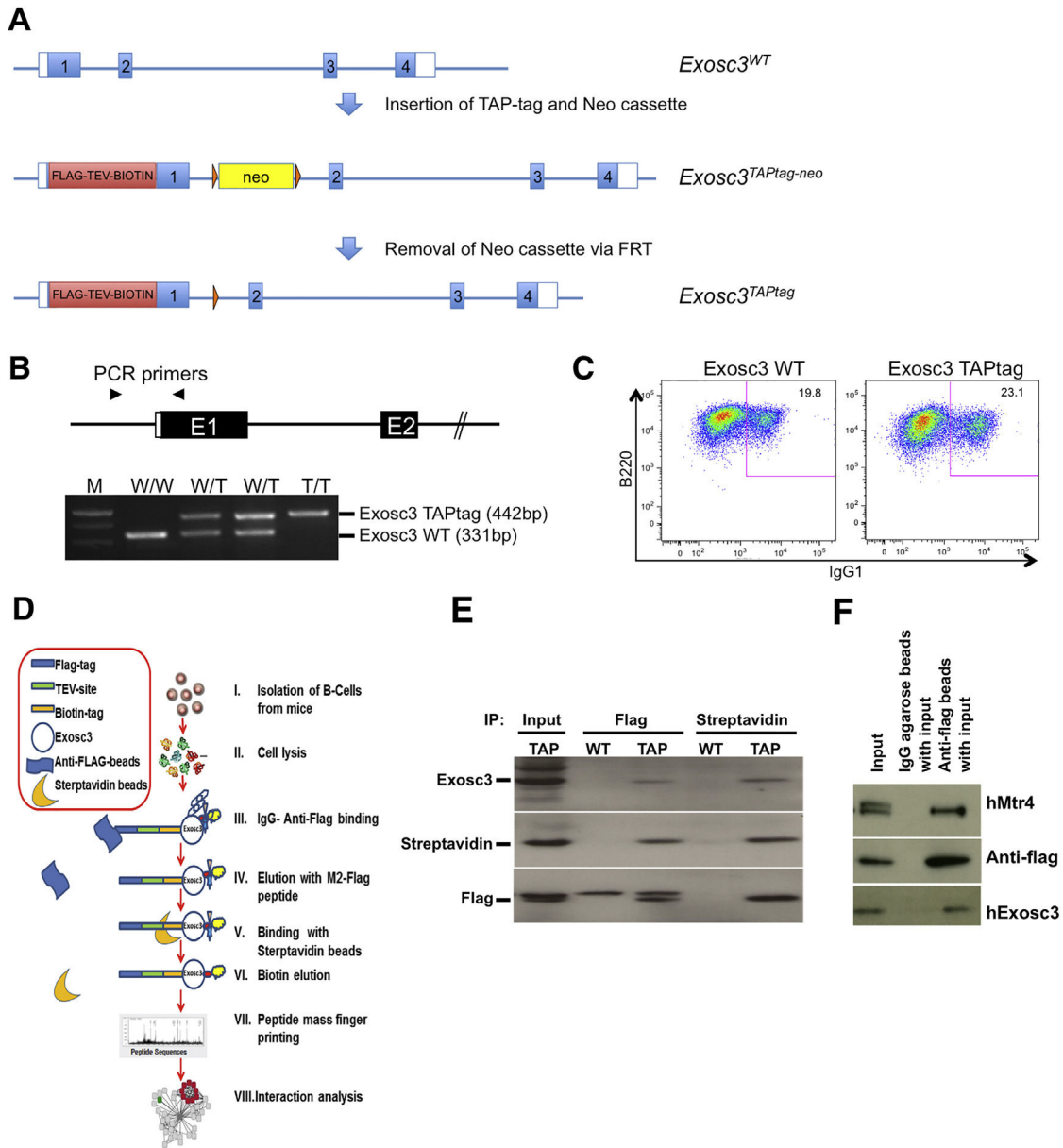
### 3D-STORM Super-Resolution Data Statistical Analysis

Proximity of RNA exosome complex with its cofactors (AID and Mtr4) were calculated by using algorithm "Nearest Neighbors Search" in the MATLAB (2014b, Math works). The 50nm range was used to maximize detection of co-localization. Distribution of the interaction between RNA exosome complex and its cofactors were represented in the form of frequency histogram plot by using MATLAB (2014b, Math works) software. We also compared the distribution of interaction between RNA exosome complex with its cofactors in the nucleus and cytoplasm by using a paired Student's t test in MATLAB (2014b, Math works) software. Comparison of the distribution of paired interaction of AID and Mtr4 in the nuclear center versus displaced from center versus cytoplasm were performed by one-way ANOVA (Tukey-Kramer test) method in MATLAB (2014b, MathWorks) software. All of the 3D STORM imaging was performed in three different B Cells (from independent experiments) and repeated three or more times. Data are presented as the means  $\pm$  SEM or SD (indicated in the legend for each figure). Statistical precision measures (mean  $\pm$  SEM/SD) and statistical significance are reported in the Figures and the Figure Legends when necessary. All error bars indicate SD (p values: \*\*  $< 0.01$ , \*\*\*  $< 0.001$ ).

## ADDITIONAL RESOURCES

The *Exosc3<sup>TAP</sup>* mice are available for other investigators following completion of the Uniform Biological Material Transfer Agreement implemented by Columbia University.

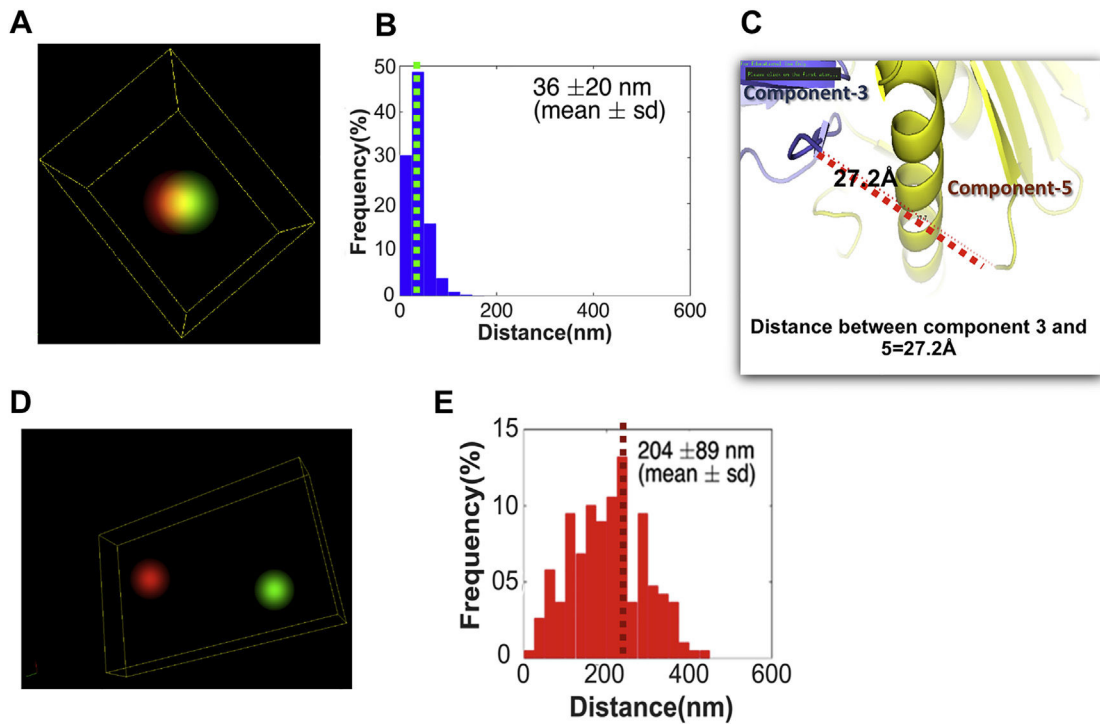




**Figure S1. Generation of *Exosc3<sup>TAP</sup>* Mouse Model and Tandem Affinity-Tagged Purification Protocol for RNA Exosome and Associated Proteins, Related to Figure 1**

(A) Schematic representation of strategy showing the TAP-tag *Exosc3* allele. The TAP-tag was inserted in the 5' region of exon1 together with a *Neo* cassette. The *Neo* gene was removed via mating to a FLPe recombinase mouse. Orange triangle represents FRT site. This diagram is not drawn to scale. (B) PCR products resulting from genotyping of *Exosc3<sup>TAPtag</sup>* mice. Map shows primers location in *Exosc3* gene locus. (C) Resting B cells from *Exosc3<sup>TAPtag</sup>* mice were treated with LPS and IL4 and assayed for IgG1 expression after 72 hr. (D) Schematic representation of the steps involved in the purification of the RNA exosome complex from B cells. (E) Flag and Streptavidin IPs were performed and the presence and functionality of tags validated by western blot prior to mass spectrometry. (F) Co-immunoprecipitation of RNA exosome complex with Mtr4 in HEK293T cells. Flag-tagged *Exosc3* and HA-tagged Mtr4-expressing plasmids were overexpressed in HEK293T cells. Lysates were co-immunoprecipitated using anti-FLAG-agarose beads, and western blots were probed with indicated antibodies.



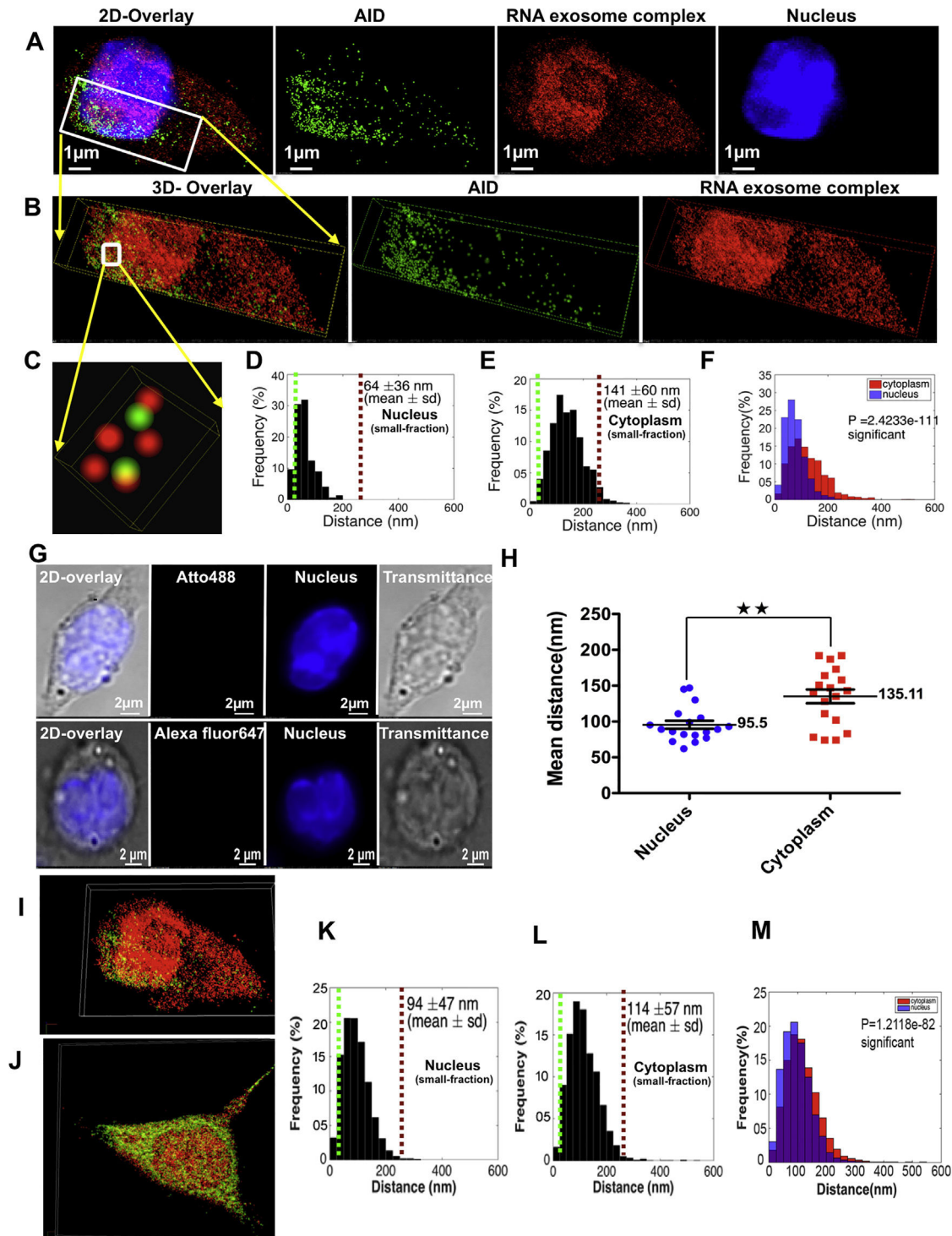


**Figure S2. 3D-STORM Imaging for Spatial Distribution of Exosc3 and Exosc5 in B Cells, as Positive Control for Closest Interaction Assessment, and Red Fluorescence Protein along with RNA Exosome Complex, as Negative Control, Non-interacting Protein Pair, Related to Figures 2–4**

B cells were harvested from *Exosc3* TAP-tagged mice and fixed after 72 hr of treatment with stimulation cocktail. HEK293T cells were transfected with Red fluorescence protein(RFP) with N terminus HA-tag and hExosc3 mammalian expression vectors and then fixed similar to B Cells post 16hrs of transfection. Reconstructed two color 3D STORM (super-resolution) image for (A) Atto488 labeled Exosc5, AlexaFluor647 labeled Exosc3 and DAPI labeled nucleus of B Cell and (D) for Atto488 labeled RFP and AlexaFluor647 labeled hExosc3 in HEK293T cell. Histogram of the distribution of interactions of Exosc3 and Exosc5 calculated in the B cell (B) and RFP and Exosome complex in HEK293T cell (E), by using custom written algorithm “Nearest Neighbors Search” in the MATLAB (2014b, Math works) software. (C) Two components Exosc3 and Exosc5 of RNA exosome complex were selected since they have closest distance in crystal structure (PDB-2NN6) and used as positive control for the interaction. All of the 3D STORM imaging were performed in three different B Cells/HEK293T cells (from independent experiments) and repeated three or more times. 3D STORM super resolution image magnification is x100.







**Figure S3. 3D-STORM Imaging for Analysis of Spatial Distribution of AID and RNA Exosome in the Nucleus of MOUSE B cells and non-B Cells, Related to Figure 3**

B cells were harvested from *Exosc3* TAP-tagged mice and fixed after 72 hr of treatment with stimuli cocktail. (A) Reconstructed two color 3D STORM (super-resolution) image for Atto488 labeled AID, AlexaFluor647 labeled *Exosc3* and DAPI labeled nucleus. (B) Three dimensional views of the boxed region. (C) Higher magnification three dimensional [3D] view of the boxed region. Histogram for the distribution of interactions of AID and RNA exosome complex were calculated in the nucleus of (D) B cell and (E) HEK293T cell and cytoplasm of (E) B cell and (L) HEK293T cell, by using MATLAB (2014b, Math works) software. Comparison of the distribution of paired interaction of AID and RNA exosome in the nucleus and cytoplasm of (F) B cell and (M) HEK293T cell were calculated by using a paired

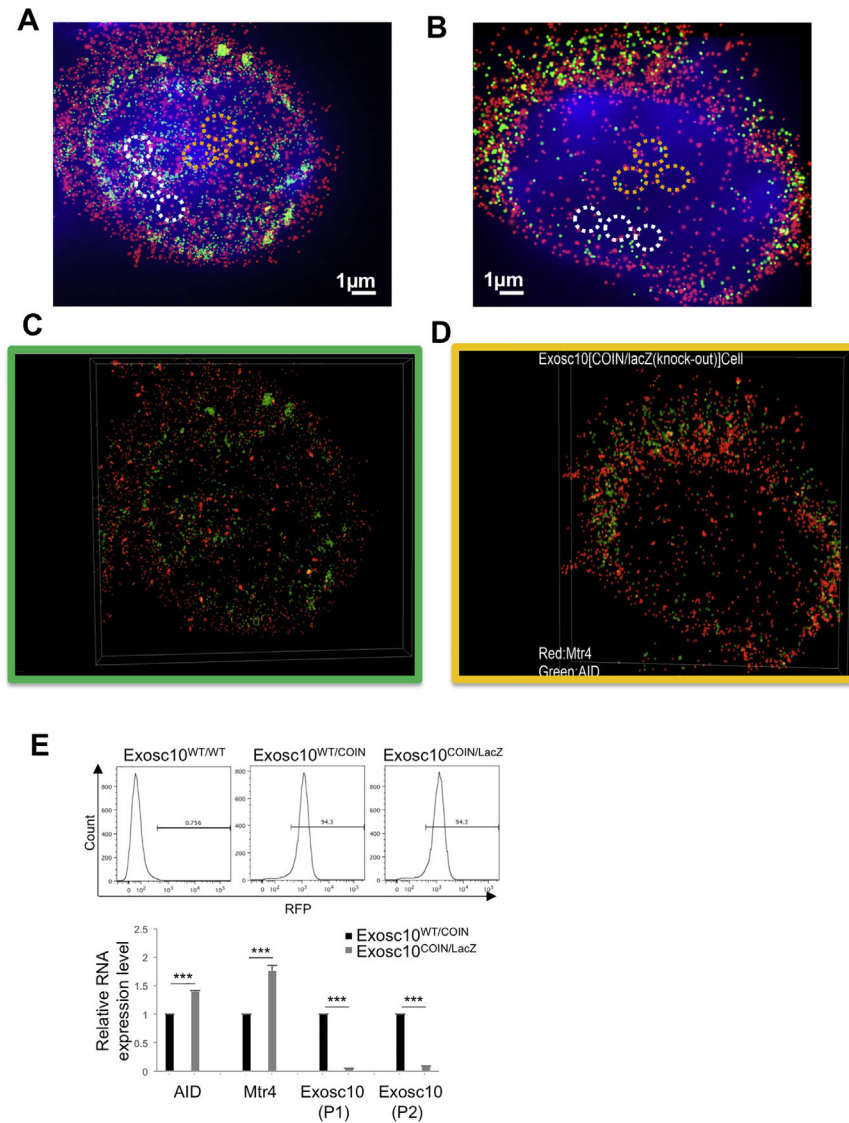
(legend continued on next page)



---

Student's t test in MATLAB (2014b, Math works) software and p values are mentioned in the graph. (G) Control experiment for non-specific binding primary and secondary antibodies- (Top panel) B cells were harvested from *AID* K/O mice and labeled with primary antibody AID, secondary antibody Atto 488& nucleus with DAPI and (Bottom panel) B cells were harvested from *Exosc3* TAP-tagged mice and labeled with secondary antibody AlexaFluor647& nucleus with DAPI. (H) Comparison of the inter molecular distance between AID molecules in the nucleus and cytoplasm of a activated B cell, calculated by using MATLAB (2014b, Math works) software. Reconstructed two color 3D STORM (super-resolution) image from (I) B cell and (J) HEK293T cell, in which AID and Exosc3 were labeled with atto488 and Alexa Fluor 647 respectively. All of the 3D STORM imaging were performed in three different B Cells/HEK293T cells (from independent experiments) and repeated three or more times. 3D STORM super resolution image magnification is x100. Scale bars, 1 $\mu$ m (A) and 2 $\mu$ m (G). (p value: \*\* < 0.01). See also [Movies S3](#) and [S4](#).

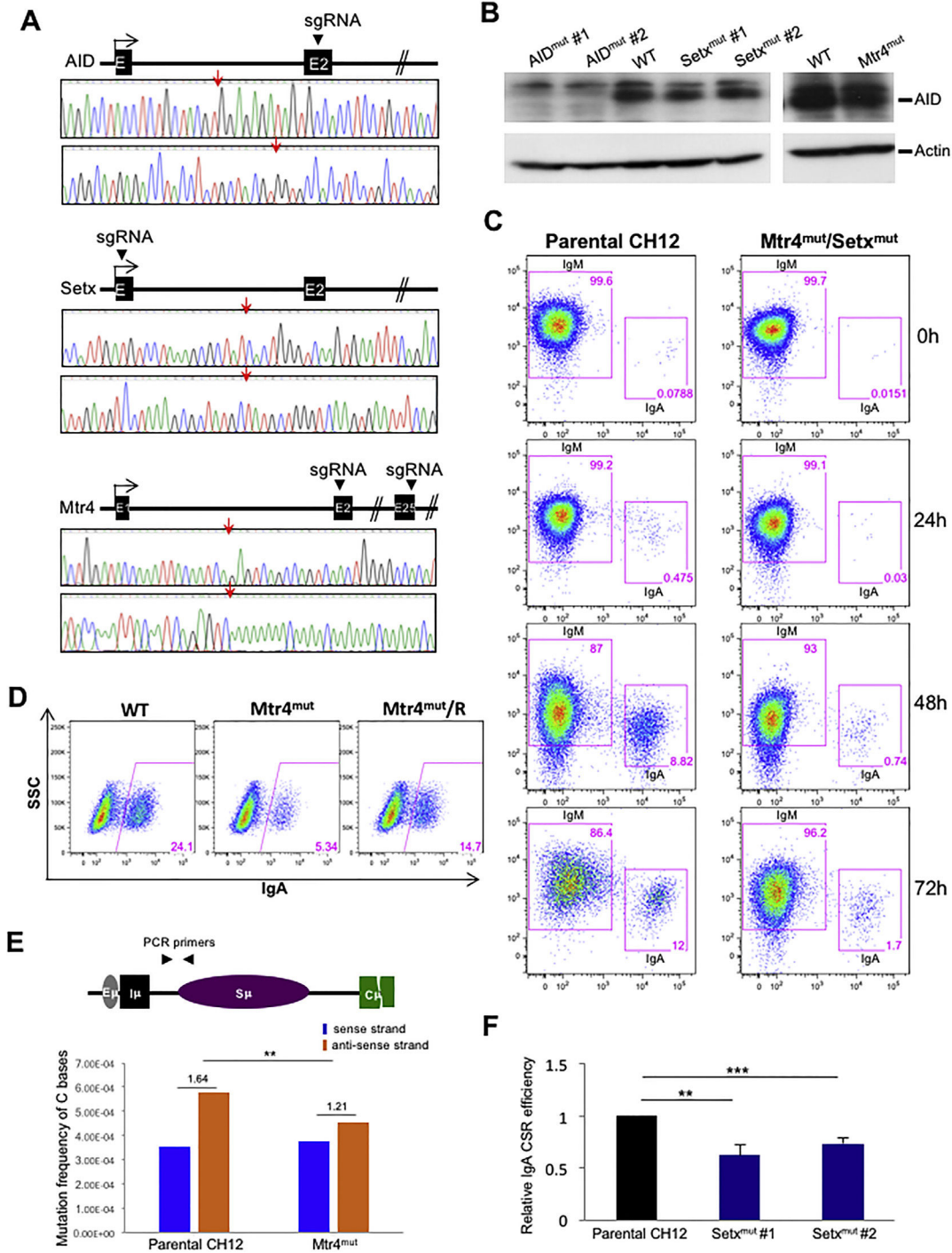




**Figure S4. 3D-STORM Images for the Analysis of Spatial Distribution of AID and Mtr4 in the Nucleus of B Cells Isolated from RNA Exosome Nuclear Activity-Deficient *Exosc10*<sup>COIN/LacZ</sup> Mouse, Related to Figure 4**

B cells were harvested from the spleen of *Exosc10*<sup>COIN/LacZ</sup> mice and fixed after 72 hr of treatment with (1) stimulation cocktail and (2) 4OHT with stimulation cocktail. Reconstructed two color 3D STORM (super-resolution) (A and C) image from a dataset of 50,000 frames with Alexa488 labeled AID, AlexaFluor647 labeled Mtr4 and DAPI labeled nucleus, show spatial distribution of AID and Mtr4 molecules inside the nuclei of B cells isolated from wild-type cells, and (B and D) image from *Exosc10* knockout cells. (E) Conditional inversion allele of *Exosc10* in 4OHT and stimuli cocktail treated *Exosc10* deficient (*Exosc10*<sup>COIN/LacZ</sup>) cells was validated by flow cytometry analysis (upper panel). RNA expression level of *Exosc10*, AID and Mtr4 genes in *Exosc10* deficient (*Exosc10*<sup>COIN/LacZ</sup>) cells was tested by RT-qPCR after 72 hr of 4OHT and stimulation cocktail treatment. RNA level (lower panel). Results were obtained from 4 experimental repeats. Error bars indicate SD (p value: \*\*\* < 0.001). All of the 3D STORM imaging were performed in three different B Cells (from independent experiments) and repeated three or more times. 3D STORM super resolution image magnification is x100. Scale bars, 1 μm(A) & (B). See also [Movies S5](#) and [S6](#).



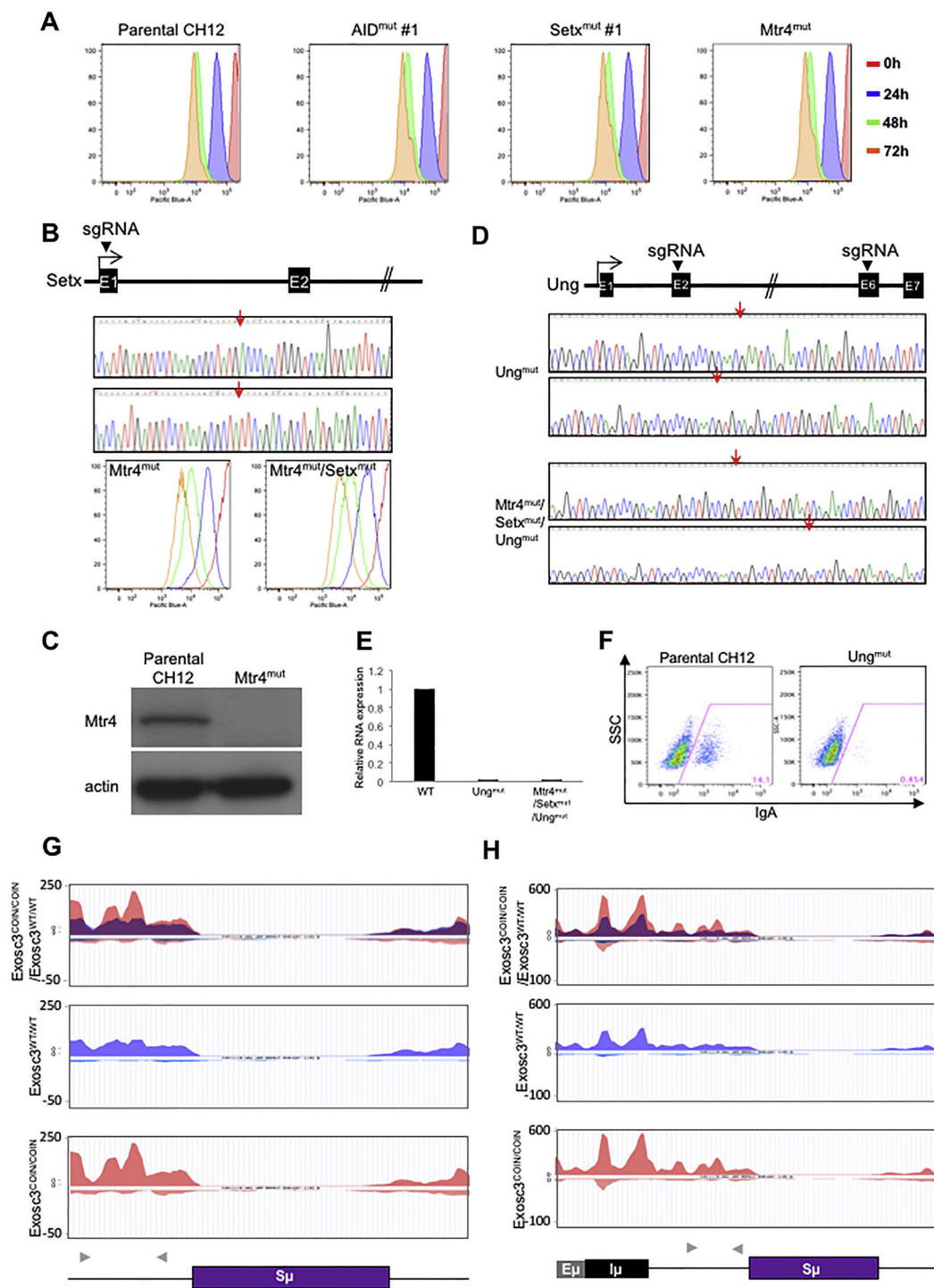


**Figure S5. Verification of *AID*, *Mtr4*, and *Setx* Mutant CH12F3 Cells and Defect in CSR Frequency, Related to Figure 5**

(A) Brief map showing gRNA target regions and sequencing results of target region in *AID* mutant, *Setx* mutant and *Mtr4* mutant. Red arrows indicate insertion or deletion junction of each allele. (B) western blot analysis shows AID expression in stimulated *AID*, *Setx* and *Mtr4* mutant CH12F3. (C) IgM to IgA CSR in *Mtr4/Setx* mutant CH12F3 cells was tested every day for 72h. (D) IgA CSR was assayed to test *Mtr4* rescue effect after transfecting *Mtr4* expressing plasmid DNA. (E) Schematic map localizing the IgH locus and binding sites of the PCR primer pair for mutation analysis. Mutation frequency of AID-induced somatic hypermutation on sense (non-transcribed) strand and anti-sense (transcribed) strand of *IgH* 5'/S $\mu$  region in *Mtr4* mutant CH12F3 cells compared to WT CH12F3 cell controls. Ratio of mutation frequency between sense and anti-sense strands is indicated above bar graph. Details of mutations identified by Sanger sequencing in Table S1C. (p values: \*\* < 0.01) (F) IgA efficiency of two different *Setx*<sup>mut</sup> clones was normalized based on CSR level of parental cells. Error bars indicate SD (p values: \*\* < 0.01, \*\*\* < 0.001).

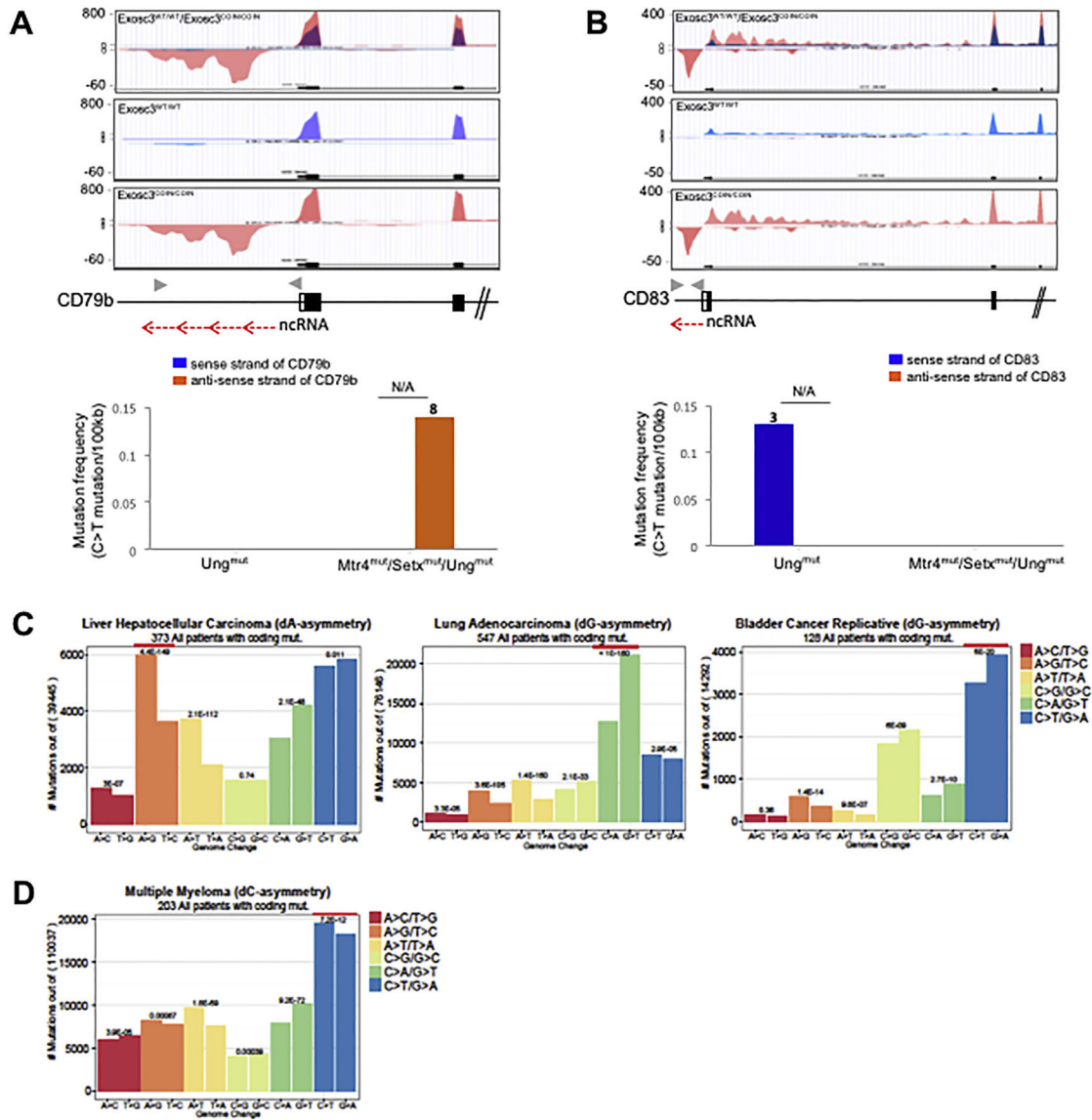






**Figure S6. Proliferation Test of Mutant CH12F3 Cells and Sense/Antisense RNA Expression in *IgH* S $\mu$  Region, Related to Figures 5–7**  
 (A) Proliferation test of mutant CH12F3 clones using VPD450. (B) Sequencing analysis of gRNA target region on the *Setx* gene locus in *Setx*/*Mtr4* double mutant clone. *Setx* gRNA was transfected into *Mtr4* mutant cells. Red arrows indicate the junction of deleted region. Proliferation test of *Setx*/*Mtr4* double mutant using VPD450 (bottom panel). (C) Western analysis of *Mtr4* expression in *Mtr4* mutant CH12F3 cells. (D) Sequencing analysis of gRNA target region at the *Ung* gene locus in *Ung* mutant CH12f3 clone and *Mtr4*/*Setx*/*Ung* mutant CH12F3 clone. (E) Validation of *Ung* gene expression using qRT-PCR in *Ung* mutant CH12F3 cells (F) Class switch recombination efficiency in *Ung* mutant CH12F3 cells (G) RNA-seq tracks from IGV browser for WT B cells (*Exosc3*<sup>WT/WT</sup>; middle panel) RNA exosome mutant B Cells (*Exosc3*<sup>COIN/COIN</sup>; bottom panel) and overlap (top panel) for sense (blue) and antisense (red) RNAs expressed in 5' and 3' *IgH* S $\mu$  regions (H) and intronic enhancer E $\mu$ , I $\mu$  and S $\mu$  region.





**Figure S7. Next Generation Sequencing Analysis in Mtr4- and Setx-Deficient CH12F3 and Genome-Wide Mutation Analysis of Cancer Cells, Related to Figure 7**

(A, B) RNA exome substrate ncRNA expression profiles of *CD79b* (C) and *CD83* (D) TSS locus in WT and *Exosc3<sup>COIN/COIN</sup>* mouse B cells (Top). Gray arrow heads indicate primer sites for next generation sequencing (NGS) and red arrows indicate non-coding RNA and its transcription direction on the schematic map (Middle). Mutation frequency of C bases on transcribed and non-transcribed strands and ratio of non-transcribed strand mutation frequency compare to transcribed strand mutation frequency in *Ung* mutant and *Mtr4/Setx/Ung* mutant CH12F3 cells (Bottom). The number above bar graph means C to T mutation count on AID mutation hotspot motif. Details of mutations identified by NGS deep sequencing in Table S1C. (C, D) Mutational counts representing somatic coding nucleotide variants across 4 different tumors in patients with coding mutations in analyzed genes. A red bar highlights specific transcriptional, replicative or mutational nucleotide changes. Indicating asymmetry in Liver, Lung and Bladder cancer, respectively (E). The largest number of mutations concurrent with Multiple Myeloma (F). p value is indicated above each bar graph.



## *Références bibliographiques*



## Références bibliographiques.

---

### A

- Afshar, R., Pierce, S., Bolland, D.J., Corcoran, A., and Oltz, E.M. (2006). Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J. Immunol. Baltim. Md 1950* *176*, 2439–2447.
- Aida, M., Hamad, N., Stanlie, A., Begum, N.A., and Honjo, T. (2013). Accumulation of the FACT complex, as well as histone H3.3, serves as a target marker for somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 7784–7789.
- Akiba, Y., Kuwabara, T., Mukozu, T., Mikami, T., and Kondo, M. (2018). Special AT-rich sequence binding protein 1 is required for maintenance of T cell receptor responsiveness and development of experimental autoimmune encephalomyelitis. *Microbiol. Immunol.* *62*, 255–268.
- Alamyar, E., P. Duroux, M.-P. Lefranc, and V. Giudicelli. 2012. IMGT<sup>(®)</sup> tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* *882*: 569–604.
- Alessandrini, A., and Desiderio, S.V. (1991). Coordination of immunoglobulin DJH transcription and D-to-JH rearrangement by promoter-enhancer approximation. *Mol. Cell. Biol.* *11*, 2096–2107.
- Allen, C.D.C., Ansel, K.M., Low, C., Lesley, R., Tamamura, H., Fujii, N., and Cyster, J.G. (2004). Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat. Immunol.* *5*, 943–952.
- Allman, D., Jain, A., Dent, A., Maile, R.R., Selvaggi, T., Kehry, M.R., and Staudt, L.M. (1996). BCL-6 expression during B-cell activation. *Blood* *87*, 5257–5268.
- Allman, D., Lindsley, R.C., DeMuth, W., Rudd, K., Shinton, S.A., and Hardy, R.R. (2001). Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. *J. Immunol. Baltim. Md 1950* *167*, 6834–6840.
- Allman, D., Srivastava, B., and Lindsley, R.C. (2004). Alternative routes to maturity: branch points and pathways for generating follicular and marginal zone B cells. *Immunol. Rev.* *197*, 147–160.
- Allman, D.M., Ferguson, S.E., and Cancro, M.P. (1992). Peripheral B cell maturation. I. Immature peripheral B cells in adults are heat-stable antigenhi and exhibit unique signaling characteristics. *J. Immunol. Baltim. Md 1950* *149*, 2533–2540.
- Almeida, K.H., and Sobol, R.W. (2007). A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair* *6*, 695–711.
- Alt, F.W., Yancopoulos, G.D., Blackwell, T.K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. (1984). Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J.* *3*, 1209–1219.
- Alt, F.W., Oltz, E.M., Young, F., Gorman, J., Taccioli, G., and Chen, J. (1992). VDJ recombination. *Immunol. Today* *13*, 306–314.
- Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., and Kohwi-Shigematsu, T. (2000a). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev.* *14*, 521–535.





Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., and Kohwi-Shigematsu, T. (2000b). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev.* *14*, 521–535.

Álvarez-Prado, Á.F., Pérez-Durán, P., Pérez-García, A., Benguria, A., Torroja, C., de Yébenes, V.G., and Ramiro, A.R. (2018). A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. *J. Exp. Med.* *215*, 761–771.

An, G., Miner, C.A., Nixon, J.C., Kincade, P.W., Bryant, J., Tucker, P.W., and Webb, C.F. (2010). Loss of Bright/ARID3a Function Promotes Developmental Plasticity. *Stem Cells Dayt. Ohio* *28*, 1560–1567.

Angelin-Duclos, C., Cattoretto, G., Lin, K.I., and Calame, K. (2000). Commitment of B lymphocytes to a plasma cell fate is associated with Blimp-1 expression in vivo. *J. Immunol. Baltim. Md* *1950* *165*, 5462–5471.

Aoufouchi, S., Faili, A., Zober, C., D’Orlando, O., Weller, S., Weill, J.-C., and Reynaud, C.-A. (2008). Proteasomal degradation restricts the nuclear lifespan of AID. *J. Exp. Med.* *205*, 1357–1368.

Aoufouchi, S., De Smet, A., Delbos, F., Gelot, C., Guerrero, I.C., Weill, J.-C., and Reynaud, C.-A. (2015). 129-Derived Mouse Strains Express an Unstable but Catalytically Active DNA Polymerase Iota Variant. *Mol. Cell. Biol.* *35*, 3059–3070.

Arakawa, H., Shimizu, T., and Takeda, S. (1996). Re-evaluation of the probabilities for productive arrangements on the kappa and lambda loci. *Int. Immunol.* *8*, 91–99.

Arakawa, H., Moldovan, G.-L., Saribasak, H., Saribasak, N.N., Jentsch, S., and Buerstedde, J.-M. (2006). A role for PCNA ubiquitination in immunoglobulin hypermutation. *PLoS Biol.* *4*, e366.

Azuma, T., Motoyama, N., Fields, L.E., and Loh, D.Y. (1993). Mutations of the chloramphenicol acetyl transferase transgene driven by the immunoglobulin promoter and intron enhancer. *Int. Immunol.* *5*, 121–130.

## **B**

Bachl, J., and Wabl, M. (1996). Enhancers of hypermutation. *Immunogenetics* *45*, 59–64.

Bahjat, M., and Guikema, J.E.J. (2017). The Complex Interplay between DNA Injury and Repair in Enzymatically Induced Mutagenesis and DNA Damage in B Lymphocytes. *Int. J. Mol. Sci.* *18*.

Balázs, M., Martin, F., Zhou, T., and Kearney, J. (2002). Blood dendritic cells interact with splenic marginal zone B cells to initiate T-independent immune responses. *Immunity* *17*, 341–352.

Baranello, L., Kouzine, F., and Levens, D. (2013). DNA topoisomerases beyond the standard role. *Transcription* *4*, 232–237.

Bassing, C.H., and Alt, F.W. (2004). The cellular response to general and programmed DNA double strand breaks. *DNA Repair* *3*, 781–796.

Basso, K., Schneider, C., Shen, Q., Holmes, A.B., Setty, M., Leslie, C., and Dalla-Favera, R. (2012). BCL6 positively regulates AID and germinal center gene expression via repression of miR-155. *J. Exp. Med.* *209*, 2455–2465.

Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J.P., Manis, J.P., and Alt, F.W. (2005). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* *438*, 508–511.



- Basu, U., Franklin, A., Schwer, B., Cheng, H.-L., Chaudhuri, J., and Alt, F.W. (2009). Regulation of Activation-Induced Cytidine Deaminase DNA Deamination Activity in B Cells by Serine-38 Phosphorylation. *Biochem. Soc. Trans.* *37*, 561–568.
- Basu, U., Meng, F.-L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wesemann, D.R., Januszyk, K., et al. (2011). The RNA Exosome Targets the AID Cytidine Deaminase to Both Strands of Transcribed Duplex DNA Substrates. *Cell* *144*, 353–363.
- Batrak, V., Blagodatski, A., and Buerstedde, J.-M. (2011). Understanding the immunoglobulin locus specificity of hypermutation. *Methods Mol. Biol. Clifton NJ* *745*, 311–326.
- Baumgarth, N. (2011). The double life of a B-1 cell: self-reactivity selects for protective effector functions. *Nat. Rev. Immunol.* *11*, 34–46.
- Baumgarth, N. (2017). A Hard(y) Look at B-1 Cell Development and Function. *J. Immunol. Baltim. Md 1950* *199*, 3387–3394.
- Bébin, A.-G., Carrion, C., Marquet, M., Cogné, N., Lecardeur, S., Cogné, M., and Pinaud, E. (2010). In Vivo Redundant Function of the 3' IgH Regulatory Element HS3b in the Mouse. *J. Immunol.* *184*, 3710–3717.
- Béguelin, W., Popovic, R., Teater, M., Jiang, Y., Bunting, K.L., Rosen, M., Shen, H., Yang, S.N., Wang, L., Ezponda, T., et al. (2013). EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* *23*, 677–692.
- Béguelin, W., Rivas, M.A., Calvo Fernández, M.T., Teater, M., Purwada, A., Redmond, D., Shen, H., Challman, M.F., Elemento, O., Singh, A., et al. (2017). EZH2 enables germinal centre formation through epigenetic silencing of CDKN1A and an Rb-E2F1 feedback loop. *Nat. Commun.* *8*.
- Begum, N.A., Stanlie, A., Nakata, M., Akiyama, H., and Honjo, T. (2012). The Histone Chaperone Spt6 Is Required for Activation-induced Cytidine Deaminase Target Determination through H3K4me3 Regulation. *J. Biol. Chem.* *287*, 32415–32429.
- Berezney, R., and Coffey, D.S. (1974). Identification of a nuclear protein matrix. *Biochem. Biophys. Res. Commun.* *60*, 1410–1417.
- Betz, A.G., Milstein, C., González-Fernández, A., Pannell, R., Larson, T., and Neuberger, M.S. (1994). Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell* *77*, 239–248.
- Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* *463*, 1042–1047.
- Bhutani, N., Decker, M.N., Brady, J.J., Bussat, R.T., Burns, D.M., Corbel, S.Y., and Blau, H.M. (2013). A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. *FASEB J.* *27*, 1107–1113.
- Birshtein, B.K. (2012). The role of CTCF binding sites in the 3' immunoglobulin heavy chain regulatory region. *Front. Genet.* *3*.
- Blagodatski, A., Batrak, V., Schmidl, S., Schoetz, U., Caldwell, R.B., Arakawa, H., and Buerstedde, J.-M. (2009). A cis-Acting Diversification Activator Both Necessary and Sufficient for AID-Mediated Hypermutation. *PLoS Genet.* *5*.



- Boboila, C., Yan, C., Wesemann, D.R., Jankovic, M., Wang, J.H., Manis, J., Nussenzweig, A., Nussenzweig, M., and Alt, F.W. (2010). Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. *J. Exp. Med.* *207*, 417–427.
- Boboila, C., Alt, F.W., and Schwer, B. (2012). Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. *Adv. Immunol.* *116*, 1–49.
- Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004a). Antisense intergenic transcription in V(D)J recombination. *Nat. Immunol.* *5*, 630–637.
- Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004b). Antisense intergenic transcription in V(D)J recombination. *Nat. Immunol.* *5*, 630–637.
- Bolland, D.J., Wood, A.L., Afshar, R., Featherstone, K., Oltz, E.M., and Corcoran, A.E. (2007a). Antisense Intergenic Transcription Precedes Igh D-to-J Recombination and Is Controlled by the Intronic Enhancer  $\epsilon_{\mu}$ . *Mol. Cell. Biol.* *27*, 5523–5533.
- Bolland, D.J., Wood, A.L., Afshar, R., Featherstone, K., Oltz, E.M., and Corcoran, A.E. (2007b). Antisense Intergenic Transcription Precedes Igh D-to-J Recombination and Is Controlled by the Intronic Enhancer  $\epsilon_{\mu}$ . *Mol. Cell. Biol.* *27*, 5523–5533.
- Borchert, G.M., Holton, N.W., and Larson, E.D. (2011). Repression of human activation induced cytidine deaminase by miR-93 and miR-155. *BMC Cancer* *11*, 347.
- Bottaro, A., Young, F., Chen, J., Serwe, M., Sablitzky, F., and Alt, F.W. (1998a). Deletion of the IgH intronic enhancer and associated matrix-attachment regions decreases, but does not abolish, class switching at the  $\mu$  locus. *Int. Immunol.* *10*, 799–806.
- Bottaro, A., Young, F., Chen, J., Serwe, M., Sablitzky, F., and Alt, F.W. (1998b). Deletion of the IgH intronic enhancer and associated matrix-attachment regions decreases, but does not abolish, class switching at the  $\mu$  locus. *Int. Immunol.* *10*, 799–806.
- Bouamar, H., Jiang, D., Wang, L., Lin, A.-P., Ortega, M., and Aguiar, R.C.T. (2015). MicroRNA 155 Control of p53 Activity Is Context Dependent and Mediated by Aicda and Socs1. *Mol. Cell. Biol.* *35*, 1329–1340.
- Boyer, F., Boutouil, H., Dalloul, I., Dalloul, Z., Cook-Moreau, J., Aldigier, J.-C., Carrion, C., Herve, B., Scaon, E., Cogné, M., et al. (2017). CSReport: A New Computational Tool Designed for Automatic Analysis of Class Switch Recombination Junctions Sequenced by High-Throughput Sequencing. *J. Immunol.* *198*, 4148–4155.
- Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 4102–4107.
- Britanova, O., Akopov, S., Lukyanov, S., Gruss, P., and Tarabykin, V. (2005). Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. *Eur. J. Neurosci.* *21*, 658–668.
- Buchanan, K.L., Smith, E.A., Dou, S., Corcoran, L.M., and Webb, C.F. (1997). Family-specific differences in transcription efficiency of Ig heavy chain promoters. *J. Immunol. Baltim. Md 1950* *159*, 1247–1254.



## C

- Cai, S., Han, H.-J., and Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat. Genet.* *34*, 42–51.
- Cai, S., Lee, C.C., and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* *38*, 1278–1288.
- Calado, D.P., Sasaki, Y., Godinho, S.A., Pellerin, A., Köchert, K., Sleckman, B.P., de Alborán, I.M., Janz, M., Rodig, S., and Rajewsky, K. (2012). MYC is essential for the formation and maintenance of germinal centers. *Nat. Immunol.* *13*, 1092–1100.
- Calame, K.L., Lin, K.-I., and Tunyaplin, C. (2003). Regulatory mechanisms that determine the development and function of plasma cells. *Annu. Rev. Immunol.* *21*, 205–230.
- Carson, S., and Wu, G.E. (1989). A linkage map of the mouse immunoglobulin lambda light chain locus. *Immunogenetics* *29*, 173–179.
- Casellas, R., Basu, U., Yewdell, W.T., Chaudhuri, J., Robbiani, D.F., and Di Noia, J.M. (2016). Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity. *Nat. Rev. Immunol.* *16*, 164–176.
- Castiblanco, D.P., Norton, D.D., Maul, R.W., and Gearhart, P.J. (2018). JH6 downstream intronic sequence is dispensable for RNA polymerase II accumulation and somatic hypermutation of the variable gene in Ramos cells. *Mol. Immunol.* *97*, 101–108.
- Cattoretti, G., Chang, C.C., Cechova, K., Zhang, J., Ye, B.H., Falini, B., Louie, D.C., Offit, K., Chaganti, R.S., and Dalla-Favera, R. (1995). BCL-6 protein is expressed in germinal-center B cells. *Blood* *86*, 45–53.
- Cattoretti, G., Büttner, M., Shaknovich, R., Kremmer, E., Alobeid, B., and Niedobitek, G. (2006). Nuclear and cytoplasmic AID in extrafollicular and germinal center B cells. *Blood* *107*, 3967–3975.
- Chakraborty, T., Chowdhury, D., Keyes, A., Jani, A., Subrahmanyam, R., Ivanova, I., and Sen, R. (2007). Repeat Organization and Epigenetic Regulation of the DH-C $\mu$  Domain of the Immunoglobulin Heavy-Chain Gene Locus. *Mol. Cell* *27*, 842–850.
- Chakraborty, T., Perlot, T., Subrahmanyam, R., Jani, A., Goff, P.H., Zhang, Y., Ivanova, I., Alt, F.W., and Sen, R. (2009a). A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. *J. Exp. Med.* *206*, 1019–1027.
- Chakraborty, T., Perlot, T., Subrahmanyam, R., Jani, A., Goff, P.H., Zhang, Y., Ivanova, I., Alt, F.W., and Sen, R. (2009b). A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. *J. Exp. Med.* *206*, 1019–1027.
- Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* *4*, 541–552.
- Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* *422*, 726–730.
- Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* *430*, 992–998.





- Chaumeil, J., and Skok, J.A. (2012). The role of CTCF in regulating V(D)J recombination. *Curr. Opin. Immunol.* **24**, 153–159.
- Chauveau, C., and Cogné, M. (1996). Palindromic structure of the IgH 3' locus control region. *Nat. Genet.* **14**, 15–16.
- Chauveau, C., Pinaud, E., and Cogne, M. (1998). Synergies between regulatory elements of the immunoglobulin heavy chain locus and its palindromic 3' locus control region. *Eur. J. Immunol.* **28**, 3048–3056.
- Chen, J., Young, F., Bottaro, A., Stewart, V., Smith, R.K., and Alt, F.W. (1993). Mutations of the intronic IgH enhancer and its flanking sequences differentially affect accessibility of the JH locus. *EMBO J.* **12**, 4635–4645.
- Cheng, H.-L., Vuong, B.Q., Basu, U., Franklin, A., Schwer, B., Astarita, J., Phan, R.T., Datta, A., Manis, J., Alt, F.W., et al. (2009). Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2717–2722.
- Chesi, M., Bergsagel, P.L., Shonukan, O.O., Martelli, M.L., Brents, L.A., Chen, T., Schröck, E., Ried, T., and Kuehl, W.M. (1998). Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. *Blood* **91**, 4457–4463.
- Chowdhury, D., and Sen, R. (2001). Stepwise activation of the immunoglobulin  $\mu$  heavy chain gene locus. *EMBO J.* **20**, 6394–6403.
- Chowdhury, D., and Sen, R. (2003). Transient IL-7/IL-7R signaling provides a mechanism for feedback inhibition of immunoglobulin heavy chain gene rearrangements. *Immunity* **18**, 229–241.
- Chung, J.B., Sater, R.A., Fields, M.L., Erikson, J., and Monroe, J.G. (2002). CD23 defines two distinct subsets of immature B cells which differ in their responses to T cell help signals. *Int. Immunol.* **14**, 157–166.
- Chung, J.B., Silverman, M., and Monroe, J.G. (2003). Transitional B cells: step by step towards immune competence. *Trends Immunol.* **24**, 343–349.
- Ciabrelli, F., and Cavalli, G. (2015). Chromatin-Driven Behavior of Topologically Associating Domains. *J. Mol. Biol.* **427**, 608–625.
- Cibulskis, K., M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander, and G. Getz. 2013. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**: 213–219.
- Cinamon, G., Zachariah, M.A., Lam, O.M., Foss, F.W., and Cyster, J.G. (2008). Follicular shuttling of marginal zone B cells facilitates antigen transport. *Nat. Immunol.* **9**, 54–62.
- Clybourn, C., Fischer, S., Auffredou, M.T., Hugues, P., Alexia, C., Bouillet, P., Raphael, M., Leca, G., Strasser, A., Tarlinton, D.M., et al. (2011). Regulation of memory B-cell survival by the BH3-only protein Puma. *Blood* **118**, 4120–4128.
- Cobaleda, C., Schebesta, A., Delogu, A., and Busslinger, M. (2007). Pax5: the guardian of B cell identity and function. *Nat. Immunol.* **8**, 463–470.
- Cockerill, P.N., and W.T. Garrard. 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell.* **44**:273–282



Cockerill, P.N., Yuen, M.H., and Garrard, W.T. (1987). The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *J. Biol. Chem.* *262*, 5394–5397.

Cogné, M. (2013). Activation-induced deaminase in B lymphocyte maturation and beyond. *Biomed. J.* *36*, 259–268.

Cogné, M., Lansford, R., Bottaro, A., Zhang, J., Gorman, J., Young, F., Cheng, H.L., and Alt, F.W. (1994). A class switch control region at the 3' end of the immunoglobulin heavy chain locus. *Cell* *77*, 737–747.

Conticello, S.G., Thomas, C.J.F., Petersen-Mahrt, S.K., and Neuberger, M.S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol. Biol. Evol.* *22*, 367–377.

Cortizas, E.M., Zahn, A., Hajjar, M.E., Patenaude, A.-M., Noia, J.M.D., and Verdun, R.E. (2013). Alternative End-Joining and Classical Nonhomologous End-Joining Pathways Repair Different Types of Double-Strand Breaks during Class-Switch Recombination. *J. Immunol.* *191*, 5751–5763.

Crotty, S., Johnston, R.J., and Schoenberger, S.P. (2010). Effectors and memories: Bcl-6 and Blimp-1 in T and B lymphocyte differentiation. *Nat. Immunol.* *11*, 114–120.

## D

Daly, J., Licence, S., Nanou, A., Morgan, G., and Mårtensson, I.-L. (2007). Transcription of productive and nonproductive VDJ-recombined alleles after IgH allelic exclusion. *EMBO J.* *26*, 4273–4282.

Dariavach, P., Williams, G.T., Campbell, K., Pettersson, S., and Neuberger, M.S. (1991). The mouse IgH 3'-enhancer. *Eur. J. Immunol.* *21*, 1499–1504.

De Silva, N.S., and Klein, U. (2015). Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* *15*, 137–148.

de Yébenes, V.G., and Ramiro, A.R. (2010). MicroRNA activity in B lymphocytes. *Methods Mol. Biol.* Clifton NJ *667*, 177–192.

Degner, S.C., Wong, T.P., Jankevicius, G., and Feeney, A.J. (2009). Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *J. Immunol. Baltim. Md 1950* *182*, 44–48.

Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., et al. (2011a). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci.* *108*, 9566–9571.

Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., et al. (2011b). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 9566–9571.

DeKoter, R.P., and Singh, H. (2000). Regulation of B Lymphocyte and Macrophage Development by Graded Expression of PU.1. *Science* *288*, 1439–1441.

DeKoter, R.P., Lee, H.-J., and Singh, H. (2002). PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors. *Immunity* *16*, 297–309.

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.-C., and Reynaud, C.-A. (2005). Contribution of DNA polymerase  $\eta$  to immunoglobulin gene hypermutation in the mouse. *J. Exp. Med.* *201*, 1191–1196.



- Delbos, F., Aoufouchi, S., Faili, A., Weill, J.-C., and Reynaud, C.-A. (2007). DNA polymerase  $\eta$  is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *J. Exp. Med.* *204*, 17–23.
- Delpy, L., Le Bert, M., Cogné, M., and Khamlichi, A.A. (2003). Germ-line transcription occurs on both the functional and the non-functional alleles of immunoglobulin constant heavy chain genes. *Eur. J. Immunol.* *33*, 2108–2113.
- Desiderio, S.V., Yancopoulos, G.D., Paskind, M., Thomas, E., Boss, M.A., Landau, N., Alt, F.W., and Baltimore, D. (1984). Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* *311*, 752–755.
- Di Noia, J., and Neuberger, M.S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* *419*, 43–48.
- Di Noia, J. M., and M. S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* *76*: 1–22.
- Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (2003). AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* *197*, 1291–1296.
- Dickinson, L.A., Joh, T., Kohwi, Y., and Kohwi-Shigematsu, T. (1992). A tissue-specific MARSAR DNA-binding protein with unusual binding site recognition. *Cell* *70*, 631–645.
- DiMenna, L.J., and Chaudhuri, J. (2016). Regulating infidelity: RNA-mediated recruitment of AID to DNA during class switch recombination. *Eur. J. Immunol.* *46*, 523–530.
- Dobrev, G. (2003). SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin gene expression. *Genes Dev.* *17*, 3048–3061.
- Dobrev, G., Chahrour, M., Dautzenberg, M., Chirivella, L., Kanzler, B., Fariñas, I., Karsenty, G., and Grosschedl, R. (2006). SATB2 Is a Multifunctional Determinant of Craniofacial Patterning and Osteoblast Differentiation. *Cell* *125*, 971–986.
- Dogan, I., Bertocci, B., Vilmont, V., Delbos, F., Mégret, J., Storck, S., Reynaud, C.-A., and Weill, J.-C. (2009). Multiple layers of B cell memory with different effector functions. *Nat. Immunol.* *10*, 1292–1299.
- Dominguez, P.M., Teater, M., Chambwe, N., Kormaksson, M., Redmond, D., Ishii, J., Vuong, B., Chaudhuri, J., Melnick, A., Vasanthakumar, A., et al. (2015). DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID. *Cell Rep.* *12*, 2086–2098.
- Dominguez-Sola, D., Victora, G.D., Ying, C.Y., Phan, R.T., Saito, M., Nussenzweig, M.C., and Dalla-Favera, R. (2012). c-MYC is required for germinal center selection and cyclic re-entry. *Nat. Immunol.* *13*, 1083–1091.
- Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.-H., Robbani, D.F., Di Virgilio, M., San-Martin, B.R., Heidkamp, G., Schwickert, T.A., et al. (2008). MicroRNA-155 Suppresses Activation-Induced Cytidine Deaminase-Mediated Myc-Igh Translocation. *Immunity* *28*, 630–638.
- Dougier, H.-L., S. Reynaud, E. Pinaud, C. Carrion, L. Delpy, and M. Cogné. 2006. Interallelic class switch recombination can reverse allelic exclusion and allow trans-complementation of an *IgH* locus switching defect. *Eur. J. Immunol.* *36*:2181–2191. doi:10.1002/eji.200535529.



Downen, J.M., Z.P. Fan, D. Hnisz, G. Ren, B.J. Abraham, L.N. Zhang, A.S. Weintraub, J. Schujers, T.I. Lee, K. Zhao, and R.A. Young. 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell*. 159:374–387. doi:10.1016/j.cell.2014.09.030.

Duke, J.L., Liu, M., Yaari, G., Khalil, A.M., Tomayko, M.M., Shlomchik, M.J., Schatz, D.G., and Kleinstein, S.H. (2013). Multiple transcription factor binding sites predict AID targeting in non-Ig genes. *J. Immunol. Baltim. Md 1950* 190, 3878–3888.

## E

Ebert, A., McManus, S., Tagoh, H., Medvedovic, J., Salvagiotto, G., Novatchkova, M., Tamir, I., Sommer, A., Jaritz, M., and Busslinger, M. (2011). The Distal VH Gene Cluster of the Igh Locus Contains Distinct Regulatory Elements with Pax5 Transcription Factor-Dependent Activity in Pro-B Cells. *Immunity* 34, 175–187.

Engel, H., Rolink, A., and Weiss, S. (1999). B cells are programmed to activate kappa and lambda for rearrangement at consecutive developmental stages. *Eur. J. Immunol.* 29, 2167–2176.

Ernst, P., and Smale, S.T. (1995). Combinatorial regulation of transcription II: The immunoglobulin mu heavy chain gene. *Immunity* 2, 427–438.

Esplin, B.L., Welner, R.S., Zhang, Q., Borghesi, L.A., and Kincade, P.W. (2009). A differentiation pathway for B1 cells in adult bone marrow. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5773–5778.

Evans, K., Ott, S., Hansen, A., Koentges, G., and Wernisch, L. (2007). A comparative study of S/MAR prediction tools. *BMC Bioinformatics* 8, 71.

## F

Featherstone, K., Wood, A.L., Bowen, A.J., and Corcoran, A.E. (2010). The mouse immunoglobulin heavy chain V-D intergenic sequence contains insulators that may regulate ordered V(D)J recombination. *J. Biol. Chem.* 285, 9327–9338.

Feldman, S., Achour, I., Wuerffel, R., Kumar, S., Gerasimova, T., Sen, R., and Kenter, A.L. (2015). Constraints contributed by chromatin looping limit recombination targeting during Ig class switch recombination. *J. Immunol. Baltim. Md 1950* 194, 2380–2389.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. *Nature* 421, 448–453.

Ferguson, A.R., Youd, M.E., and Corley, R.B. (2004). Marginal zone B cells transport and deposit IgM-containing immune complexes onto follicular dendritic cells. *Int. Immunol.* 16, 1411–1422.

Ferrier, P., Krippel, B., Blackwell, T.K., Furley, A.J., Suh, H., Winoto, A., Cook, W.D., Hood, L., Costantini, F., and Alt, F.W. (1990). Separate elements control DJ and VDJ rearrangement in a transgenic recombination substrate. *EMBO J.* 9, 117–125.

Fitzsimmons, S.P., Bernstein, R.M., Max, E.E., Skok, J.A., and Shapiro, M.A. (2007). Dynamic changes in accessibility, nuclear positioning, recombination, and transcription at the Ig kappa locus. *J. Immunol. Baltim. Md 1950* 179, 5264–5273.

Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10460–10465.





Forrester, W.C., van Genderen, C., Jenuwein, T., and Grosschedl, R. (1994). Dependence of enhancer-mediated transcription of the immunoglobulin mu gene on nuclear matrix attachment regions. *Science* 265, 1221–1225.

Fortini, P., and Dogliotti, E. (2007). Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair* 6, 398–409.

Fortini, P., Pascucci, B., Parlanti, E., Sobol, R.W., Wilson, S.H., and Dogliotti, E. (1998). Different DNA polymerases are involved in the short- and long-patch base excision repair in mammalian cells. *Biochemistry (Mosc.)* 37, 3575–3580.

Fraenkel, S., Mostoslavsky, R., Novobrantseva, T.I., Pelanda, R., Chaudhuri, J., Esposito, G., Jung, S., Alt, F.W., Rajewsky, K., Cedar, H., et al. (2007). Allelic “choice” governs somatic hypermutation in vivo at the immunoglobulin kappa-chain locus. *Nat. Immunol.* 8, 715–722.

Franklin, A., and Blanden, R.V. (2005). Hypothesis: biological role for J-C intronic matrix attachment regions in the molecular mechanism of antigen-driven somatic hypermutation. *Immunol. Cell Biol.* 83, 383–391.

Fritz, E.L., and Papavasiliou, F.N. (2010). Cytidine deaminases: AIDing DNA demethylation? *Genes Dev.* 24, 2107–2114.

Fritz, E.L., Rosenberg, B.R., Lay, K., Mihailović, A., Tuschl, T., and Papavasiliou, F.N. (2013). A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat. Immunol.* 14, 749–755.

Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* 9, 105–114.

Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., and Busslinger, M. (2004). Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.* 18, 411–422.

## G

Garot, A., Marquet, M., Saintamand, A., Bender, S., Le Noir, S., Rouaud, P., Carrion, C., Oruc, Z., Bébin, A.-G., Moreau, J., et al. (2016). Sequential activation and distinct functions for distal and proximal modules within the IgH 3' regulatory region. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1618–1623.

Garrett, F.E., Emelyanov, A.V., Sepulveda, M.A., Flanagan, P., Volpi, S., Li, F., Loukinov, D., Eckhardt, L.A., Lobanenkova, V.V., and Birshtein, B.K. (2005). Chromatin Architecture near a Potential 3' End of the Igh Locus Involves Modular Regulation of Histone Modifications during B-Cell Development and In Vivo Occupancy at CTCF Sites. *Mol. Cell. Biol.* 25, 1511–1525.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* 7, 703–713.

Gazumyan, A., Timachova, K., Yuen, G., Siden, E., Di Virgilio, M., Woo, E.M., Chait, B.T., Reina San-Martin, B., Nussenzweig, M.C., and McBride, K.M. (2011). Amino-Terminal Phosphorylation of Activation-Induced Cytidine Deaminase Suppresses c-myc/IgH Translocation. *Mol. Cell. Biol.* 31, 442–449.

Genetta, T., Ruezinsky, D., and Kadesch, T. (1994). Displacement of an E-box-binding repressor by basic helix-loop-helix proteins: implications for B-cell specificity of the immunoglobulin heavy-chain enhancer. *Mol. Cell. Biol.* 14, 6153–6163.



- Giallourakis, C.C., Franklin, A., Guo, C., Cheng, H.-L., Yoon, H.S., Gallagher, M., Perlot, T., Andzelm, M., Murphy, A.J., Macdonald, L.E., et al. (2010). Elements between the IgH variable (V) and diversity (D) clusters influence antisense transcription and lineage-specific V(D)J recombination. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 22207–22212.
- Girelli Zubani, G., Zivojnovic, M., De Smet, A., Albagli-Curiel, O., Huetz, F., Weill, J.-C., Reynaud, C.-A., and Storck, S. (2017). Pms2 and uracil-DNA glycosylases act jointly in the mismatch repair pathway to generate Ig gene mutations at A-T base pairs. *J. Exp. Med.* *214*, 1169–1180.
- Giudicelli, V., Chaume, D., and Lefranc, M.-P. (2005). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* *33*, D256–D261.
- Giusti, A.M., and Manser, T. (1993). Hypermutation is observed only in antibody H chain V region transgenes that have recombined with endogenous immunoglobulin H DNA: implications for the location of cis-acting elements required for somatic mutation. *J. Exp. Med.* *177*, 797–809.
- Gluch, A., Vidakovic, M., and Bode, J. (2008). Scaffold/matrix attachment regions (S/MARs): relevance for disease and therapy. *Handb. Exp. Pharmacol.* 67–103.
- Goebel, P., Janney, N., Valenzuela, J.R., Romanow, W.J., Murre, C., and Feeney, A.J. (2001). Localized gene-specific induction of accessibility to V(D)J recombination induced by E2A and early B cell factor in nonlymphoid cells. *J. Exp. Med.* *194*, 645–656.
- Goebel, P., Montalbano, A., Ayers, N., Kompfner, E., Dickinson, L., Webb, C.F., and Feeney, A.J. (2002). High frequency of matrix attachment regions and cut-like protein  $\chi$ /CCAAT-displacement protein and B cell regulator of IgH transcription binding sites flanking Ig V region genes. *J. Immunol. Baltim. Md 1950* *169*, 2477–2487.
- Goetze, S., Baer, A., Winkelmann, S., Nehlsen, K., Seibler, J., Maass, K., and Bode, J. (2005). Performance of Genomic Bordering Elements at Predefined Genomic Loci. *Mol. Cell. Biol.* *25*, 2260–2272.
- Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., and Bergman, Y. (2005). Epigenetic ontogeny of the Igk locus during B cell development. *Nat. Immunol.* *6*, 198–203.
- Gonda, H., Sugai, M., Nambu, Y., Katakai, T., Agata, Y., Mori, K.J., Yokota, Y., and Shimizu, A. (2003). The balance between Pax5 and Id2 activities is the key to AID gene expression. *J. Exp. Med.* *198*, 1427–1437.
- González-Fernández, A., D. Gilmore, and C. Milstein. 1994. Age-related decrease in the proportion of germinal center B cells from mouse Peyer's patches is accompanied by an accumulation of somatic mutations in their immunoglobulin genes. *Eur. J. Immunol.* *24*: 2918–2921.
- Gordon, M.S., Kanegai, C.M., Doerr, J.R., and Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). *Proc. Natl. Acad. Sci. U. S. A.* *100*, 4126–4131.
- Gourzi, P., Leonova, T., and Papavasiliou, F.N. (2006). A role for activation-induced cytidine deaminase in the host response against a transforming retrovirus. *Immunity* *24*, 779–786.
- Goyenechea, B., Klix, N., Yélamos, J., Williams, G.T., Riddell, A., Neuberger, M.S., and Milstein, C. (1997). Cells strongly expressing Ig(kappa) transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. *EMBO J.* *16*, 3987–3994.



Gramlich, H.S., Reisbig, T., and Schatz, D.G. (2012). AID-targeting and hypermutation of non-immunoglobulin genes does not correlate with proximity to immunoglobulin genes in germinal center B cells. *PLoS One* 7, e39601.

Gregor, P.D., and Morrison, S.L. (1986). Myeloma mutant with a novel 3' flanking region: loss of normal sequence and insertion of repetitive elements leads to decreased transcription but normal processing of the alpha heavy-chain gene products. *Mol. Cell. Biol.* 6, 1903–1916.

Gu, X., Booth, C.J., Liu, Z., and Strout, M.P. (2016). AID-associated DNA repair pathways regulate malignant transformation in a murine model of BCL6-driven diffuse large B-cell lymphoma. *Blood* 127, 102–112.

Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.-L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., et al. (2011a). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424–430.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011b). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy chain gene locus. *Cell* 147, 332–343.

Gyorgy, A.B., Szemes, M., de Juan Romero, C., Tarabykin, V., and Agoston, D.V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. *Eur. J. Neurosci.* 27, 865–873.

## H

Habir, K., Aeinehband, S., Wermeling, F., and Malin, S. (2017). A Role for the Transcription Factor Arid3a in Mouse B2 Lymphocyte Expansion and Peritoneal B1a Generation. *Front. Immunol.* 8, 1387.

Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat. Struct. Mol. Biol.* 21, 198–206.

Hackney, J.A., Misaghi, S., Senger, K., Garris, C., Sun, Y., Lorenzo, M.N., and Zarrin, A.A. (2009). DNA targets of AID evolutionary link between antibody somatic hypermutation and class switch recombination. *Adv. Immunol.* 101, 163–189.

Haddad, D., Puget, N., Laviolette-Malirat, N., Conte, C., and Khamlichi, A.A. (2011). Seeking sense of antisense switch transcripts. *Transcription* 2, 183–188.

Hakes, D.J., and Berezney, R. (1991). DNA binding properties of the nuclear matrix and individual nuclear matrix proteins. Evidence for salt-resistant DNA binding sites. *J. Biol. Chem.* 266, 11131–11140.

Halverson, R., Torres, R.M., and Pelanda, R. (2004). Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. *Nat. Immunol.* 5, 645–650.

Han, J.-H., Akira, S., Calame, K., Beutler, B., Selsing, E., and Imanishi-Kari, T. (2007). Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell- and Toll-like receptors. *Immunity* 27, 64–75.

Han, L., Masani, S., and Yu, K. (2011). Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11584–11589.

Hao, B., Naik, A.K., Watanabe, A., Tanaka, H., Chen, L., Richards, H.W., Kondo, M., Taniuchi, I., Kohwi, Y., Kohwi-Shigematsu, T., et al. (2015). An anti-silencer- and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J. Exp. Med.* 212, 809–824.



Hardy, R.R., Kincade, P.W., and Dorshkind, K. (2007). The Protean Nature of Cells in the B Lymphocyte Lineage. *Immunity* 26, 703–714.

Hayakawa, K., Hardy, R.R., Herzenberg, L.A., and Herzenberg, L.A. (1985). Progenitors for Ly-1 B cells are distinct from progenitors for other B cells. *J. Exp. Med.* 161, 1554–1568.

Heng, H.H., Krawetz, S.A., Lu, W., Bremer, S., Liu, G., and Ye, C.J. (2001). Re-defining the chromatin loop domain. *Cytogenet. Cell Genet.* 93, 155–161.

Heng, H.H.Q., Goetze, S., Ye, C.J., Liu, G., Stevens, J.B., Bremer, S.W., Wykes, S.M., Bode, J., and Krawetz, S.A. (2004). Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J. Cell Sci.* 117, 999–1008.

Herrscher, R.F., Kaplan, M.H., Lelsz, D.L., Das, C., Scheuermann, R., and Tucker, P.W. (1995). The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev.* 9, 3067–3082.

Hewitt, S.L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., Salvaggio, G., Steinel, N., Ramsey, L.B., Ghysdael, J., et al. (2009). RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nat. Immunol.* 10, 655–664.

Hewitt, S.L., Chaumeil, J., and Skok, J.A. (2010). Chromosome dynamics and the regulation of V(D)J recombination. *Immunol. Rev.* 237, 43–54.

Hirano, M., Das, S., Guo, P., and Cooper, M.D. (2011). Chapter 4 - The Evolution of Adaptive Immunity in Vertebrates. In *Advances in Immunology*, F.W. Alt, ed. (Academic Press), pp. 125–157.

Hobeika, E., Thiemann, S., Storch, B., Jumaa, H., Nielsen, P.J., Pelanda, R., and Reth, M. (2006). Testing gene function early in the B cell lineage in mb1-cre mice. *Proc. Natl. Acad. Sci. U. S. A.* 103, 13789–13794.

Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. *Mol. Cell* 48, 365–374.

Holwerda, S.J.B., van de Werken, H.J.G., Ribeiro de Almeida, C., Bergen, I.M., de Bruijn, M.J.W., Versteegen, M.J.A.M., Simonis, M., Splinter, E., Wijchers, P.J., Hendriks, R.W., et al. (2013). Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. *Nucleic Acids Res.* 41, 6905–6916.

Huang, C., Hatzi, K., and Melnick, A. (2013). Lineage-specific functions of Bcl-6 in immunity and inflammation are mediated by distinct biochemical mechanisms. *Nat. Immunol.* 14, 380–388.

HWANG, J.K., ALT, F.W., and YEAP, L.-S. (2015). Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination. *Microbiol. Spectr.* 3.

## I

Igarashi, H., Gregory, S.C., Yokota, T., Sakaguchi, N., and Kincade, P.W. (2002). Transcription from the RAG1 locus marks the earliest lymphocyte progenitors in bone marrow. *Immunity* 17, 117–130.

Imler, J.L., Lemaire, C., Wasyluk, C., and Wasyluk, B. (1987). Negative regulation contributes to tissue specificity of the immunoglobulin heavy-chain enhancer. *Mol. Cell. Biol.* 7, 2558–2567.





Inlay, M.A., H.H. Gao, V.H. Odegard, T. Lin, D.G. Schatz, and Y. Xu. 2006. Roles of the Ig kappa *IgHt* chain intronic and 3' enhancers in Igk somatic hypermutation. *J. Immunol.* 177:1146–1151.

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1975–1980.

## J

Jackson, D.A., Dickinson, P., and Cook, P.R. (1990). The size of chromatin loops in HeLa cells. *EMBO J.* 9, 567–571.

Jacob, J., and Kelsoe, G. (1992). In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. II. A common clonal origin for periarteriolar lymphoid sheath-associated foci and germinal centers. *J. Exp. Med.* 176, 679–687.

Jacob, J., Kelsoe, G., Rajewsky, K., and Weiss, U. (1991). Intraclonal generation of antibody mutants in germinal centres. *Nature* 354, 389–392.

Jankovic, M., Robbiani, D.F., Dorsett, Y., Eisenreich, T., Xu, Y., Tarakhovsky, A., Nussenzweig, A., and Nussenzweig, M.C. (2010). Role of the translocation partner in protection against AID-dependent chromosomal translocations. *Proc. Natl. Acad. Sci. U. S. A.* 107, 187–192.

Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H., and de Wind, N. (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. *J. Exp. Med.* 203, 319–323.

Jardin, F., and Sahota, S.S. (2005). Targeted somatic mutation of the BCL6 proto-oncogene and its impact on lymphomagenesis. *Hematol. Amst. Neth.* 10, 115–129.

Jardin, F., P. Ruminy, C. Bastard, and H. Tilly. 2007. The BCL6 proto-oncogene: a leading role during germinal center development and lymphomagenesis. *Pathol. Biol.* 55:73–83. doi:10.1016/j.patbio.2006.04.001.

Jenuwein, T. (2001). Translating the Histone Code. *Science* 293, 1074–1080.

Jenuwein, T., Forrester, W.C., Fernández-Herrero, L.A., Laible, G., Dull, M., and Grosschedl, R. (1997a). Extension of chromatin accessibility by nuclear matrix attachment regions. *Nature* 385, 269–272.

Jenuwein, T., Forrester, W.C., Fernández-Herrero, L.A., Laible, G., Dull, M., and Grosschedl, R. (1997b). Extension of chromatin accessibility by nuclear matrix attachment regions. *Nature* 385, 269–272.

Jhunjunwala, S., van Zelm, M.C., Peak, M.M., Cutchin, S., Riblet, R., van Dongen, J.J.M., Grosveld, F.G., Knoch, T.A., and Murre, C. (2008). The 3D Structure of the Immunoglobulin Heavy-Chain Locus: Implications for Long-Range Genomic Interactions. *Cell* 133, 265–279.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., and Schatz, D.G. (2010). The In Vivo Pattern of Binding of RAG1 and RAG2 to Antigen Receptor Loci. *Cell* 141, 419–431.

Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* 7, 335–346.

Johnston, C.M., Wood, A.L., Bolland, D.J., and Corcoran, A.E. (2006). Complete Sequence Assembly and Characterization of the C57BL/6 Mouse Ig Heavy Chain V Region. *J. Immunol.* 176, 4221–4234.



Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F.W. (2006). Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* *24*, 541–570.

Jung, S., Rajewsky, K., and Radbruch, A. (1993). Shutdown of class switch recombination by deletion of a switch region control element. *Science* *259*, 984–987.

## K

Kadesch, T., Zervos, P., and Ruezinsky, D. (1986). Functional analysis of the murine IgH enhancer: evidence for negative control of cell-type specificity. *Nucleic Acids Res.* *14*, 8209–8221.

Kakugawa, K., Kojo, S., Tanaka, H., Seo, W., Endo, T.A., Kitagawa, Y., Muroi, S., Tenno, M., Yasmin, N., Kohwi, Y., et al. (2017). Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. *Cell Rep.* *19*, 1176–1188.

Kalouisi, A., and Soutoglou, E. (2016). Nuclear compartmentalization of DNA repair. *Curr. Opin. Genet. Dev.* *37*, 148–157.

Kaplan, M.H., Zong, R.-T., Herrscher, R.F., Scheuermann, R.H., and Tucker, P.W. (2001a). Transcriptional Activation by a Matrix Associating Region-binding Protein CONTEXTUAL REQUIREMENTS FOR THE FUNCTION OF BRIGHT. *J. Biol. Chem.* *276*, 21325–21330.

Kaplan, M.H., Zong, R.T., Herrscher, R.F., Scheuermann, R.H., and Tucker, P.W. (2001b). Transcriptional activation by a matrix associating region-binding protein. contextual requirements for the function of bright. *J. Biol. Chem.* *276*, 21325–21330.

Kato, L., Begum, N.A., Burroughs, A.M., Doi, T., Kawai, J., Daub, C.O., Kawaguchi, T., Matsuda, F., Hayashizaki, Y., and Honjo, T. (2012). Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 2479–2484.

Keim, C., Kazadi, D., Rothschild, G., and Basu, U. (2013). Regulation of AID, the B-cell genome mutator. *Genes Dev.* *27*, 1–17.

Kenter, A.L., Feldman, S., Wuerffel, R., Achour, I., Wang, L., and Kumar, S. (2012). Three-dimensional architecture of the IgH locus facilitates class switch recombination. *Ann. N. Y. Acad. Sci.* *1267*, 86–94.

Khair, L., Baker, R.E., Linehan, E.K., Schrader, C.E., and Stavnezer, J. (2015). Nbs1 ChIP-Seq Identifies Off-Target DNA Double-Strand Breaks Induced by AID in Activated Splenic B Cells. *PLoS Genet.* *11*, e1005438.

Khamlichi, A.A., Glaudet, F., Oruc, Z., Denis, V., Le Bert, M., and Cogné, M. (2004). Immunoglobulin class-switch recombination in mice devoid of any S mu tandem repeat. *Blood* *103*, 3828–3836.

Khan, W.N. (2009). B cell receptor and BAFF receptor signaling regulation of B cell homeostasis. *J. Immunol. Baltim. Md 1950* *183*, 3561–3567.

Kim, Y.-J., and Wilson, D.M. (2012). Overview of base excision repair biochemistry. *Curr. Mol. Pharmacol.* *5*, 3–13.

Kim, D., Probst, L., Das, C., and Tucker, P.W. (2007). REKLES is an ARID3-restricted multifunctional domain. *J. Biol. Chem.* *282*, 15768–15777.

Kim, D., Schmidt, C., Brown, M.A., and Tucker, H. (2017). Competitive Promoter-Associated Matrix Attachment Region Binding of the Arid3a and Cux1 Transcription Factors. *Dis. Basel Switz.* *5*.



- King, J.J., Manuel, C.A., Barrett, C.V., Raber, S., Lucas, H., Sutter, P., and Larijani, M. (2015). Catalytic pocket inaccessibility of activation-induced cytidine deaminase is a safeguard against excessive mutagenic activity. *Struct. Lond. Engl.* 1993 23, 615–627.
- Kisseljova, N.P., Dmitriev, P., Katargin, A., Kim, E., Ezerina, D., Markozashvili, D., Malysheva, D., Planche, E., Lemmers, R.J.L.F., van der Maarel, S.M., et al. (2014). DNA polymorphism and epigenetic marks modulate the affinity of a scaffold/matrix attachment region to the nuclear matrix. *Eur. J. Hum. Genet. EJHG* 22, 1117–1123.
- Kitano, M., Moriyama, S., Ando, Y., Hikida, M., Mori, Y., Kurosaki, T., and Okada, T. (2011). Bcl6 protein expression shapes pre-germinal center B cell dynamics and follicular helper T cell heterogeneity. *Immunity* 34, 961–972.
- Klein, U., and Dalla-Favera, R. (2008). Germinal centres: role in B-cell physiology and malignancy. *Nat. Rev. Immunol.* 8, 22–33.
- Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147, 95–106.
- Klymenko, T., Bloehdorn, J., Bahlo, J., Robrecht, S., Akylzhanova, G., Cox, K., Estenfelder, S., Wang, J., Edelmann, J., Strefford, J.C., et al. (2018). Lamin B1 regulates somatic mutations and progression of B-cell malignancies. *Leukemia* 32, 364–375.
- Kobayashi, M., Aida, M., Nagaoka, H., Begum, N.A., Kitawaki, Y., Nakata, M., Stanlie, A., Doi, T., Kato, L., Okazaki, I., et al. (2009). AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22375–22380.
- Kobayashi, M., Sabouri, Z., Sabouri, S., Kitawaki, Y., Pommier, Y., Abe, T., Kiyonari, H., and Honjo, T. (2011). Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 19305–19310.
- Koboldt, D. C., Zhang, D. E., Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22: 568–576.
- Kodgire, P., Mukkavar, P., Ratnam, S., Martin, T.E., and Storb, U. (2013). Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID. *J. Exp. Med.* 210, 1481–1492.
- Kohler, K.M., McDonald, J.J., Duke, J.L., Arakawa, H., Tan, S., Kleinstein, S.H., Buerstedde, J.-M., and Schatz, D.G. (2012). Identification of core DNA elements that target somatic hypermutation. *J. Immunol. Baltim. Md* 1950 189, 5314–5326.
- Kohli, R.M., Maul, R.W., Guminski, A.F., McClure, R.L., Gajula, K.S., Saribasak, H., McMahan, M.A., Siliciano, R.F., Gearhart, P.J., and Stivers, J.T. (2010). Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. *J. Biol. Chem.* 285, 40956–40964.
- Kohwi-Shigematsu, T., Maass, K., and Bode, J. (1997). A thymocyte factor SATB1 suppresses transcription of stably integrated matrix-attachment region-linked reporter genes. *Biochemistry* 36, 12005–12010.
- Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., and Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* 5, 1.



- Kosak, S.T. (2002). Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development. *Science* 296, 158–162.
- Kothapalli, N.R., Collura, K.M., Norton, D.D., and Fugmann, S.D. (2011). Separation of mutational and transcriptional enhancers in immunoglobulin genes. *J. Immunol. Baltim. Md 1950* 187, 3247–3255.
- Kottmann, A.H., Zevnik, B., Welte, M., Nielsen, P.J., and Köhler, G. (1994). A second promoter and enhancer element within the immunoglobulin heavy chain locus. *Eur. J. Immunol.* 24, 817–821.
- Kou, T., Marusawa, H., Kinoshita, K., Endo, Y., Okazaki, I.-M., Ueda, Y., Kodama, Y., Haga, H., Ikai, I., and Chiba, T. (2007). Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis. *Int. J. Cancer* 120, 469–476.
- Kozmik, Z., Wang, S., Dörfler, P., Adams, B., and Busslinger, M. (1992). The promoter of the CD19 gene is a target for the B-cell-specific transcription factor BSAP. *Mol. Cell. Biol.* 12, 2662–2672.
- Krijger, P.H.L., Langerak, P., van den Berk, P.C.M., and Jacobs, H. (2009). Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. *J. Exp. Med.* 206, 2603–2611.
- Krokan, H.E., and Bjørås, M. (2013). Base excision repair. *Cold Spring Harb. Perspect. Biol.* 5, a012583.
- Kunimoto, H., McKenney, A.S., Meydan, C., Shank, K., Nazir, A., Rapaport, F., Durham, B., Garrett-Bakelman, F.E., Pronier, E., Shih, A.H., et al. (2017). Aid is a key regulator of myeloid/erythroid differentiation and DNA methylation in hematopoietic stem/progenitor cells. *Blood* 129, 1779–1790.
- Kurkewich, J.L., Klopfenstein, N., Hallas, W.M., Wood, C., Sattler, R.A., Das, C., Tucker, H., Dahl, R., and Cowden Dahl, K.D. (2016). Arid3b Is Critical for B Lymphocyte Development. *PLoS ONE* 11.
- Kurosaki, T., Kometani, K., and Ise, W. (2015). Memory B cells. *Nat. Rev. Immunol.* 15, 149–159.

## L

- Laffleur, B., Basu, U., and Lim, J. (2017). RNA Exosome and Non-coding RNA-Coupled Mechanisms in AID-Mediated Genomic Alterations. *J. Mol. Biol.* 429, 3230–3241.
- Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* 39, 170–182.
- Lam, T., Thomas, L.M., White, C.A., Li, G., Pone, E.J., Xu, Z., and Casali, P. (2013). Scaffold functions of 14-3-3 adaptors in B cell immunoglobulin class switch DNA recombination. *PLoS One* 8, e80414.
- Langerak, P., Nygren, A.O.H., Krijger, P.H.L., van den Berk, P.C.M., and Jacobs, H. (2007). A/T mutagenesis in hypermutated immunoglobulin genes strongly depends on PCNAK164 modification. *J. Exp. Med.* 204, 1989–1998.
- Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.
- Larijani, M., and Martin, A. (2007). Single-stranded DNA structure and positional context of the target cytidine determine the enzymatic efficiency of AID. *Mol. Cell. Biol.* 27, 8038–8048.
- Lazorchak, A.S., Wojciechowski, J., Dai, M., and Zhuang, Y. (2006). E2A promotes the survival of precursor and mature B lymphocytes. *J. Immunol. Baltim. Md 1950* 177, 2495–2504.





- Le Noir, S., Boyer, F., Lecardeur, S., Brousse, M., Oruc, Z., Cook-Moreau, J., Denizot, Y., and Cogné, M. (2017). Functional anatomy of the immunoglobulin heavy chain 3' super-enhancer needs not only core enhancer elements but also their unique DNA context. *Nucleic Acids Res.* *45*, 5829–5837.
- Lebecque, S.G., and Gearhart, P.J. (1990). Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *J. Exp. Med.* *172*, 1717–1727.
- Ledford, A.W., Brantley, J.G., Kemeny, G., Foreman, T.L., Quaggin, S.E., Igarashi, P., Oberhaus, S.M., Rodova, M., Calvet, J.P., and Vanden Heuvel, G.B. (2002). Deregulated expression of the homeobox gene *Cux-1* in transgenic mice results in downregulation of p27(kip1) expression during nephrogenesis, glomerular abnormalities, and multiorgan hyperplasia. *Dev. Biol.* *245*, 157–171.
- Lee, J.-H., and Paull, T.T. (2007). Activation and regulation of ATM kinase activity in response to DNA double-strand breaks. *Oncogene* *26*, 7741–7748.
- Lemaître, C., Grabarz, A., Tsouroula, K., Andronov, L., Furst, A., Pankotai, T., Heyer, V., Rogier, M., Attwood, K.M., Kessler, P., et al. (2014). Nuclear position dictates DNA repair pathway choice. *Genes Dev.* *28*, 2450–2463.
- Lennon, G.G., and Perry, R.P. (1985). C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. *Nature* *318*, 475–478.
- Li, F., Yan, Y., Pieretti, J., Feldman, D.A., and Eckhardt, L.A. (2010a). Comparison of identical and functional Igh alleles reveals a nonessential role for  $\mu$  in somatic hypermutation and class-switch recombination. *J. Immunol. Baltim. Md 1950* *185*, 6049–6057.
- Li, G., White, C.A., Lam, T., Pone, E.J., Tran, D.C., Hayama, K.L., Zan, H., Xu, Z., and Casali, P. (2013). Combinatorial H3K9acS10ph Histone Modifications in IgH Locus S Regions Target 14-3-3 Adaptors and AID to Specify Antibody Class Switch DNA Recombination. *Cell Rep.* *5*, 702–714.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*: 2078–2079.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* *25*: 1754–1760.
- Li, W., Notani, D., and Rosenfeld, M.G. (2016a). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat. Rev. Genet.* *17*, 207–223.
- Li, Y., Gupta, G., Molofsky, A., Xie, Y., Shihabi, N., McCormick, J., and Jaffe, E.S. (2018). B Lymphoblastic Leukemia/Lymphoma With Burkitt-like Morphology and IGH/MYC Rearrangement: Report of 3 Cases in Adult Patients. *Am. J. Surg. Pathol.* *42*, 269–276.
- Li, Z., B. J. Abraham, A. Berezovskaya, N. Farah, Y. Liu, T. Leon, A. Fielding, S. H. Tan, T. Sanda, A. S. Weintraub, B. Li, S. Shen, J. Zhang, M. R. Mansour, R. A. Young, and A. T. Look. 2017. APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL. *Leukemia* *31*: 2057–2064.
- Liang, G., Kitamura, K., Wang, Z., Liu, G., Chowdhury, S., Fu, W., Koura, M., Wakae, K., Honjo, T., and Muramatsu, M. (2013). RNA editing of hepatitis B virus transcripts by activation-induced cytidine deaminase. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 2246–2251.



- Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211.
- Lim, J., Giri, P.K., Kazadi, D., Laffleur, B., Zhang, W., Grinstein, V., Pefanis, E., Brown, L.M., Ladewig, E., Martin, O., et al. (2017). Nuclear Proximity of Mtr4 to RNA Exosome Restricts DNA Mutational Asymmetry. *Cell* **169**, 523–537.e15.
- Lin, Y.C., and Murre, C. (2013). Nuclear Location and the Control of Developmental Progression. *Curr. Opin. Genet. Dev.* **23**, 104–108.
- Lin, D., Ippolito, G.C., Zong, R.-T., Bryant, J., Koslovsky, J., and Tucker, P. (2007). Bright/ARID3A contributes to chromatin accessibility of the immunoglobulin heavy chain enhancer. *Mol. Cancer* **6**, 23.
- Lin, M.M., Green, N.S., Zhang, W., and Scharff, M.D. (1998). The effects of E mu, 3'alpha (hs 1,2) and 3'kappa enhancers on mutation of an Ig-VDJ-Cgamma2a Ig heavy gene in cultured B cells. *Int. Immunol.* **10**, 1121–1129.
- Lin, S.G., Guo, C., Su, A., Zhang, Y., and Alt, F.W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1815–1820.
- Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in nuclear positioning of genes and intra- and inter-domain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* **13**, 1196–1204.
- Liu, M., and Schatz, D.G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* **30**, 173–181.
- Liu, F., Wu, D., and Wang, X. (2018). Roles of CTCF in conformation and functions of chromosome. *Semin. Cell Dev. Biol.*
- Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**, 841–845.
- Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., and Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* **27**, 561–571.
- Liubchenko, G.A., Appleberry, H.C., Holers, V.M., Banda, N.K., Willis, V.C., and Lyubchenko, T. (2012). Potentially autoreactive naturally occurring transitional T3 B lymphocytes exhibit a unique signaling profile. *J. Autoimmun.* **38**, 293–303.
- Lovely, G.A., Brewster, R.C., Schatz, D.G., Baltimore, D., and Phillips, R. (2015). Single-molecule analysis of RAG-mediated V(D)J DNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1715–E1723.
- Luby, T.M., Schrader, C.E., Stavnezer, J., and Selsing, E. (2001a). The mu switch region tandem repeats are important, but not required, for antibody class switch recombination. *J. Exp. Med.* **193**, 159–168.
- Luby, T.M., Sigurdardottir, D., Berger, E.D., and Selsing, E. (2001b). Sequences associated with the mouse Smu switch region are important for immunoglobulin heavy chain transgene expression in B cell development. *Eur. J. Immunol.* **31**, 2866–2875.
- Ludérus, M.E., de Graaf, A., Mattia, E., den Blaauwen, J.L., Grande, M.A., de Jong, L., and van Driel, R. (1992). Binding of matrix attachment regions to lamin B1. *Cell* **70**, 949–959.



Luo, H., and Tian, M. (2010). Transcription factors PU.1 and IRF4 regulate activation induced cytidine deaminase in chicken B cells. *Mol. Immunol.* *47*, 1383–1395.

## M

Maccarthy, T., S. Roa, M. D. Scharff, and A. Bergman. 2009. SHMTool: a webserver for comparative analysis of somatic hypermutation datasets. *DNA Repair (Amst.)* *8*: 137–141.

MacLennan, I.C.M., Toellner, K.-M., Cunningham, A.F., Serre, K., Sze, D.M.-Y., Zúñiga, E., Cook, M.C., and Vinuesa, C.G. (2003). Extrafollicular antibody responses. *Immunol. Rev.* *194*, 8–18.

Madisen, L., and Groudine, M. (1994). Identification of a locus control region in the immunoglobulin heavy-chain locus that deregulates c-myc expression in plasmacytoma and Burkitt's lymphoma cells. *Genes Dev.* *8*, 2212–2226.

Maeda, K., Singh, S.K., Eda, K., Kitabatake, M., Pham, P., Goodman, M.F., and Sakaguchi, N. (2010). GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *J. Biol. Chem.* *285*, 23945–23953.

Maier, H., and Hagman, J. (2002). Roles of EBF and Pax-5 in B lineage commitment and development. *Semin. Immunol.* *14*, 415–422.

Manis, J.P., van der Stoep, N., Tian, M., Ferrini, R., Davidson, L., Bottaro, A., and Alt, F.W. (1998). Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers. *J. Exp. Med.* *188*, 1421–1431.

Manis, J.P., Morales, J.C., Xia, Z., Kutok, J.L., Alt, F.W., and Carpenter, P.B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. *Nat. Immunol.* *5*, 481–487.

Marquet, M., Garot, A., Bender, S., Carrion, C., Rouaud, P., Lecardeur, S., Denizot, Y., Cogné, M., and Pinaud, E. (2014). The E $\mu$  Enhancer Region Influences H Chain Expression and B Cell Fate without Impacting IgVH Repertoire and Immune Response In Vivo. *J. Immunol.* *193*, 1171–1183.

Mårtensson, I.-L., and Ceredig, R. (2000). Role of the surrogate light chain and the pre-B-cell receptor in mouse B-cell development. *Immunology* *101*, 435–441.

Mårtensson, I.-L., Rolink, A., Melchers, F., Mundt, C., Licence, S., and Shimizu, T. (2002). The pre-B cell receptor and its role in proliferation and Ig heavy chain allelic exclusion. *Semin. Immunol.* *14*, 335–342.

Martin, F., Oliver, A.M., and Kearney, J.F. (2001). Marginal zone and B1 B cells unite in the early response against T-independent blood-borne particulate antigens. *Immunity* *14*, 617–629.

Martin, O.A., A. Garot, S. Le Noir, J.-C. Aldigier, M. Cogné, E. Pinaud, and F. Boyer. 2018. Detecting Rare AID-Induced Mutations in B-Lineage Oncogenes from *HlgH*-Throughput Sequencing Data Using the Detection of Minor Variants by Error Correction Method. *J. Immunol.* doi:10.4049/jimmunol.1800203.

Martins, R.P., Ostermeier, G.C., and Krawetz, S.A. (2004). Nuclear Matrix Interactions at the Human Protamine Domain A WORKING MODEL OF POTENTIATION. *J. Biol. Chem.* *279*, 51862–51868.

Masani, S., Han, L., and Yu, K. (2013). Apurinic/apyrimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination. *Mol. Cell. Biol.* *33*, 1468–1473.



- Matsumoto, Y., Marusawa, H., Kinoshita, K., Niwa, Y., Sakai, Y., and Chiba, T. (2010). Up-regulation of activation-induced cytidine deaminase causes genetic aberrations at the CDKN2b-CDKN2a in gastric cancer. *Gastroenterology* *139*, 1984–1994.
- Matthews, A.G.W., Kuo, A.J., Ramón-Maiques, S., Han, S., Champagne, K.S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D.N., et al. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* *450*, 1106–1110.
- Matthews, A.J., Husain, S., and Chaudhuri, J. (2014). Binding of AID to DNA does not correlate with mutator activity. *J. Immunol. Baltim. Md 1950* *193*, 252–257.
- Matthias, P., and Baltimore, D. (1993). The immunoglobulin heavy chain locus contains another B-cell-specific 3'enhancer close to the alpha constant region. *Mol. Cell. Biol.* *13*, 1547–1553.
- Maul, R.W., Cao, Z., Venkataraman, L., Giorgetti, C.A., Press, J.L., Denizot, Y., Du, H., Sen, R., and Gearhart, P.J. (2014). Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo-activated cells. *J. Exp. Med.* *211*, 2297–2306.
- Maul, R.W., Saribasak, H., Cao, Z., and Gearhart, P.J. (2015). Topoisomerase I deficiency causes RNA polymerase II accumulation and increases AID abundance in immunoglobulin variable genes. *DNA Repair* *30*, 46–52.
- Maul, R.W., MacCarthy, T., Frank, E.G., Donigan, K.A., McLenigan, M.P., Yang, W., Saribasak, H., Huston, D.E., Lange, S.S., Woodgate, R., et al. (2016). DNA polymerase  $\iota$  functions in the generation of tandem mutations during somatic hypermutation of antibody genes. *J. Exp. Med.* *213*, 1675–1683.
- McBride, K.M., Gazumyan, A., Woo, E.M., Barreto, V.M., Robbiani, D.F., Chait, B.T., and Nussenzweig, M.C. (2006). Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 8798–8803.
- McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J. Exp. Med.* *205*, 2585–2594.
- McDonald, J.P., Frank, E.G., Plosky, B.S., Rogozin, I.B., Masutani, C., Hanaoka, F., Woodgate, R., and Gearhart, P.J. (2003). 129-derived strains of mice are deficient in DNA polymerase  $\iota$  and have normal immunoglobulin hypermutation. *J. Exp. Med.* *198*, 635–643.
- McIntyre, J., Vidal, A.E., McLenigan, M.P., Bomar, M.G., Curti, E., McDonald, J.P., Plosky, B.S., Ohashi, E., and Woodgate, R. (2013). Ubiquitin mediates the physical and functional interaction between human DNA polymerases  $\eta$  and  $\iota$ . *Nucleic Acids Res.* *41*, 1649–1660.
- Medina, K.L., Pongubala, J.M.R., Reddy, K.L., Lancki, D.W., Dekoter, R., Kieslinger, M., Grosschedl, R., and Singh, H. (2004). Assembling a gene regulatory network for specification of the B cell fate. *Dev. Cell* *7*, 607–617.
- Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I.M., Schwickert, T.A., Novatchkova, M., Sun, Q., Huis In 't Veld, P.J., Guo, C., Yoon, H.S., et al. (2013). Flexible long-range loops in the VH gene region of the Igh locus facilitate the generation of a diverse antibody repertoire. *Immunity* *39*, 229–244.
- Melchers, F., Karasuyama, H., Haasner, D., Bauer, S., Kudo, A., Sakaguchi, N., Jameson, B., and Rolink, A. (1993). The surrogate light chain in B-cell development. *Immunol. Today* *14*, 60–68.





- Meng, F.-L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.-R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberger, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* *159*, 1538–1548.
- Meng, X., Qi, X., Guo, H., Cai, M., Li, C., Zhu, J., Chen, F., Guo, H., Li, J., Zhao, Y., et al. (2015). Novel role for non-homologous end joining in the formation of double minutes in methotrexate-resistant colon cancer cells. *J. Med. Genet.* *52*, 135–144.
- Methot, S. P., and J. M. Di Noia. 2017. Molecular Mechanisms of Somatic Hypermutation and Class Switch Recombination. *Adv. Immunol.* *133*: 37–87.
- Michael, N., Shen, H.M., Longerich, S., Kim, N., Longacre, A., and Storb, U. (2003). The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity* *19*, 235–242.
- Michaelson, J.S., Giannini, S.L., and Birshtein, B.K. (1995). Identification of 3' alpha-hs4, a novel Ig heavy chain enhancer element regulated at multiple stages of B cell differentiation. *Nucleic Acids Res.* *23*, 975–981.
- Migliazza, A., Martinotti, S., Chen, W., Fusco, C., Ye, B.H., Knowles, D.M., Offit, K., Chaganti, R.S., and Dalla-Favera, R. (1995). Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* *92*, 12520–12524.
- Mikkola, I., Heavey, B., Horcher, M., and Busslinger, M. (2002). Reversion of B cell commitment upon loss of Pax5 expression. *Science* *297*, 110–113.
- Milstein, C., Neuberger, M.S., and Staden, R. (1998). Both DNA strands of antibody genes are hypermutation targets. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 8791–8794.
- Mombaerts, P., Iacomini, J., Johnson, R.S., Herrup, K., Tonegawa, S., and Papaioannou, V.E. (1992). RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* *68*, 869–877.
- Monroe, J.G., and Dorshkind, K. (2007). Fate Decisions Regulating Bone Marrow and Peripheral B Lymphocyte Development. In *Advances in Immunology*, (Academic Press), pp. 1–50.
- Morgado, M.G., Cam, P., Gris-Liebe, C., Cazenave, P.A., and Jouvin-Marche, E. (1989). Further evidence that BALB/c and C57BL/6 gamma 2a genes originate from two distinct isotypes. *EMBO J.* *8*, 3245–3251.
- Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* *279*, 52353–52360.
- Morshead, K.B., Ciccone, D.N., Taverna, S.D., Allis, C.D., and Oettinger, M.A. (2003). Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 11577–11582.
- Morvan, C.L., Pinaud, E., Decourt, C., Cuvillier, A., and Cogné, M. (2003). The immunoglobulin heavy-chain locus hs3b and hs4 3' enhancers are dispensable for VDJ assembly and somatic hypermutation. *Blood* *102*, 1421–1427.
- Motoyama, N., Miwa, T., Suzuki, Y., Okada, H., and Azuma, T. (1994). Comparison of somatic mutation frequency among immunoglobulin genes. *J. Exp. Med.* *179*, 395–403.
- Mouly, E., Ghamlouch, H., Della-Valle, V., Scourzic, L., Quivoron, C., Roos-Weil, D., Pawlikowska, P., Saada, V., Diop, M.K., Lopez, C.K., et al. (2018). B-cell tumor development in Tet2-deficient mice. *Blood Adv.* *2*, 703–714.



Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol. Cell* *51*, 606–617.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* *274*, 18470–18476.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell* *102*, 553–563.

Müschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., and Küppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* *192*, 1833–1840.

## N

Nabel, C.S., Lee, J.W., Wang, L.C., and Kohli, R.M. (2013). Nucleic acid determinants for selective deamination of DNA over RNA by activation-induced deaminase. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 14225–14230.

Nagaoka, H., Muramatsu, M., Yamamura, N., Kinoshita, K., and Honjo, T. (2002). Activation-induced deaminase (AID)-directed hypermutation in the immunoglobulin Smu region: implication of AID involvement in a common step of class switch recombination and somatic hypermutation. *J. Exp. Med.* *195*, 529–534.

Nakayasu, H., and Berezney, R. (1991). Nuclear matrins: identification of the major nuclear matrix proteins. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 10312–10316.

Nambu, Y., Sugai, M., Gonda, H., Lee, C.-G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* *302*, 2137–2140.

Neuberger, M.S., Di Noia, J.M., Beale, R.C.L., Williams, G.T., Yang, Z., and Rada, C. (2005). Somatic hypermutation at A.T pairs: polymerase error versus dUTP incorporation. *Nat. Rev. Immunol.* *5*, 171–178.

Niuro, H., and Clark, E.A. (2002). Regulation of B-cell fate by antigen-receptor signals. *Nat. Rev. Immunol.* *2*, 945–956.

Nitschke, L., Kestler, J., Tallone, T., Pelkonen, S., and Pelkonen, J. (2001). Deletion of the DQ52 element within the Ig heavy chain locus leads to a selective reduction in VDJ recombination and altered D gene usage. *J. Immunol. Baltim. Md 1950* *166*, 2540–2552.

Niu, H., Cattoretti, G., and Dalla-Favera, R. (2003). BCL6 controls the expression of the B7-1/CD80 costimulatory receptor in germinal center B cells. *J. Exp. Med.* *198*, 211–221.

Nixon, J.C., Rajaiya, J., and Webb, C.F. (2004). Mutations in the DNA-binding Domain of the Transcription Factor Bright Act as Dominant Negative Proteins and Interfere with Immunoglobulin Transactivation. *J. Biol. Chem.* *279*, 52465–52472.

Nixon, J.C., Ferrell, S., Miner, C., Oldham, A.L., Hochgeschwender, U., and Webb, C.F. (2008). Transgenic mice expressing dominant-negative bright exhibit defects in B1 B cells. *J. Immunol. Baltim. Md 1950* *181*, 6913–6922.



Nojima, T., Haniuda, K., Moutai, T., Matsudaira, M., Mizokawa, S., Shiratori, I., Azuma, T., and Kitamura, D. (2011). In-vitro derived germinal centre B cells differentially generate memory B or plasma cells in vivo. *Nat. Commun.* *2*, 465.

Norvell, A., Mandik, L., and Monroe, J.G. (1995). Engagement of the antigen-receptor on immature murine B lymphocytes results in death by apoptosis. *J. Immunol. Baltim. Md 1950* *154*, 4404–4413.

Numata, S., Claudio, P.P., Dean, C., Giordano, A., and Croce, C.M. (1999). Bdp, a New Member of a Family of DNA-binding Proteins, Associates with the Retinoblastoma Gene Product. *Cancer Res.* *59*, 3741–3747.

Nutt, S.L., and Kee, B.L. (2007). The Transcriptional Regulation of B Cell Lineage Commitment. *Immunity* *26*, 715–725.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* *401*, 556–562.

Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M., and Corcoran, L.M. (2015). The generation of antibody-secreting plasma cells. *Nat. Rev. Immunol.* *15*, 160–171.

## O

Ochiai, K., Maienschein-Cline, M., Simonetti, G., Chen, J., Rosenthal, R., Brink, R., Chong, A.S., Klein, U., Dinner, A.R., Singh, H., et al. (2013). Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity* *38*, 918–929.

Oduor, C.I., Kaymaz, Y., Chelimo, K., Otieno, J.A., Ong'echa, J.M., Moormann, A.M., and Bailey, J.A. (2017). Integrative microRNA and mRNA deep-sequencing expression profiling in endemic Burkitt lymphoma. *BMC Cancer* *17*, 761.

Okazaki, I., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.* *197*, 1173–1181.

Okazaki, I., Kotani, A., and Honjo, T. (2007). Role of AID in Tumorigenesis. In *Advances in Immunology*, (Academic Press), pp. 245–273.

Oldham, A.L., Miner, C.A., Wang, H.-C., and Webb, C.F. (2011). The Transcription Factor Bright Plays a Role in Marginal Zone B Lymphocyte Development and Autoantibody Production. *Mol. Immunol.* *49*, 367–379.

O'Riordan, M., and Grosschedl, R. (1999). Coordinate Regulation of B Cell Differentiation by the Transcription Factors EBF and E2A. *Immunity* *11*, 21–31.

Orthwein, A., and Di Noia, J.M. (2012). Activation induced deaminase: how much and where? *Semin. Immunol.* *24*, 246–254.

Osborne, C.S., Chakalova, L., Mitchell, J.A., Horton, A., Wood, A.L., Bolland, D.J., Corcoran, A.E., and Fraser, P. (2007). Myc Dynamically and Preferentially Relocates to a Transcription Factory Occupied by Igh. *PLoS Biol.* *5*.

## P

Packard, T.A., and Cambier, J.C. (2013). B lymphocyte antigen receptor signaling: initiation, amplification, and regulation. *F1000Prime Rep.* *5*.



- Panchakshari, R.A., Zhang, X., Kumar, V., Du, Z., Wei, P.-C., Kao, J., Dong, J., and Alt, F.W. (2018). DNA double-strand break response factors influence end-joining features of IgH class switch and general translocation junctions. *Proc. Natl. Acad. Sci. U. S. A.* *115*, 762–767.
- Pannunzio, N.R., and Lieber, M.R. (2018). Concept of DNA Lesion Longevity and Chromosomal Translocations. *Trends Biochem. Sci.*
- Park, S.-K., Xiang, Y., Feng, X., and Garrard, W.T. (2014). Pronounced cohabitation of active immunoglobulin genes from three different chromosomes in transcription factories during maximal antibody synthesis. *Genes Dev.* *28*, 1159–1164.
- Parsa, J.-Y., Ramachandran, S., Zaheen, A., Nepal, R.M., Kapelnikov, A., Belcheva, A., Berru, M., Ronai, D., and Martin, A. (2012). Negative supercoiling creates single-stranded patches of DNA that are substrates for AID-mediated mutagenesis. *PLoS Genet.* *8*, e1002518.
- Parslow, T.G., Blair, D.L., Murphy, W.J., and Granner, D.K. (1984). Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc. Natl. Acad. Sci. U. S. A.* *81*, 2650–2654.
- Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R.S., Klein, U., Küppers, R., Rajewsky, K., et al. (1998). BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 11816–11821.
- Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Küppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* *412*, 341–346.
- Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006). PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 395–400.
- Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Morse, H.C., Nussenzweig, M.C., and Dalla-Favera, R. (2008). AID is required for germinal center-derived lymphomagenesis. *Nat. Genet.* *40*, 108–112.
- Patenaude, A.-M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzi, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat. Struct. Mol. Biol.* *16*, 517–527.
- Pavan Kumar, P., Purbey, P.K., Sinha, C.K., Notani, D., Limaye, A., Jayani, R.S., and Galande, S. (2006). Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. *Mol. Cell* *22*, 231–243.
- Pavlicek, J.W., Lyubchenko, Y.L., and Chang, Y. (2008). Quantitative Analyses of RAG-RSS Interactions and Conformations Revealed by Atomic Force Microscopy. *Biochemistry (Mosc.)* *47*, 11204–11211.
- Pavri, R., and Nussenzweig, M.C. (2011). Chapter 1 - AID Targeting in Antibody Diversity. In *Advances in Immunology*, F.W. Alt, K.F. Austen, T. Honj, F. Melchers, J.W. Uhr, and E.R. Unanue, eds. (Academic Press), pp. 1–26.
- Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* *143*, 122–133.





- Pawlitzky, I., Angeles, C.V., Siegel, A.M., Stanton, M.L., Riblet, R., and Brodeur, P.H. (2006). Identification of a Candidate Regulatory Element within the 5' Flanking Region of the Mouse Igh Locus Defined by Pro-B Cell-Specific Hypersensitivity Associated with Binding of PU.1, Pax5, and E2A. *J. Immunol.* *176*, 6839–6851.
- Pefanis, E., and Basu, U. (2015). RNA exosome regulates AID DNA mutator activity in the B cell genome. *Adv. Immunol.* *127*, 257–308.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* *514*, 389–393.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.-P., et al. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* *161*, 774–789.
- Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D. (2008). The biochemistry of somatic hypermutation. *Annu. Rev. Immunol.* *26*, 481–511.
- Peng, C., and Eckhardt, L.A. (2013). Role of the IgH Intronic Enhancer E $\mu$  in Clonal Selection at the Pre-B to Immature B Cell Transition. *J. Immunol. Baltim. Md* *1950* *191*.
- Perlot, T., and Alt, F.W. (2008). Cis-regulatory elements and epigenetic changes control genomic rearrangements of the IgH locus. *Adv. Immunol.* *99*, 1–32.
- Perlot, T., Alt, F.W., Bassing, C.H., Suh, H., and Pinaud, E. (2005). Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 14362–14367.
- Perlot, T., Li, G., and Alt, F.W. (2008). Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 3843–3848.
- Perlot, T., Pawlitzky, I., Manis, J.P., Zarrin, A.A., Brodeur, P.H., and Alt, F.W. (2010). Analysis of mice lacking DNaseI hypersensitive sites at the 5' end of the IgH locus. *PloS One* *5*, e13992.
- Péron, S., Laffleur, B., Denis-Lagache, N., Cook-Moreau, J., Tinguely, A., Delpy, L., Denizot, Y., Pinaud, E., and Cogné, M. (2012). AID-driven deletion causes immunoglobulin heavy chain locus suicide recombination in B cells. *Science* *336*, 931–934.
- Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R., et al. (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* *414*, 660–665.
- Petro, J.B., Gerstein, R.M., Lowe, J., Carter, R.S., Shinnars, N., and Khan, W.N. (2002). Transitional type 1 and 2 B lymphocyte subsets are differentially responsive to antigen receptor signaling. *J. Biol. Chem.* *277*, 48009–48019.
- Pettersson, S., Cook, G.P., Brüggemann, M., Williams, G.T., and Neuberger, M.S. (1990). A second B cell-specific enhancer 3' of the immunoglobulin heavy-chain locus. *Nature* *344*, 165–168.
- Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* *424*, 103–107.
- Phan, R.T., and Dalla-Favera, R. (2004). The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* *432*, 635–639.



Pichugin, A., Iarovaia, O.V., Gavrillov, A., Sklyar, I., Barinova, N., Barinov, A., Ivashkin, E., Caron, G., Aoufouchi, S., Razin, S.V., et al. (2017). The IGH locus relocates to a “recombination compartment” in the perinucleolar region of differentiating B-lymphocytes. *Oncotarget* 8, 40079–40089.

Pike, B.L., Boyd, A.W., and Nossal, G.J. (1982). Clonal anergy: the universally anergic B lymphocyte. *Proc. Natl. Acad. Sci. U. S. A.* 79, 2013–2017.

Pillai, S., and Cariappa, A. (2009). The follicular versus marginal zone B lymphocyte cell fate decision. *Nat. Rev. Immunol.* 9, 767–777.

Pillai, S., Cariappa, A., and Moran, S.T. (2005). Marginal zone B cells. *Annu. Rev. Immunol.* 23, 161–196.

Pinaud, E., Aupetit, C., Chauveau, C., and Cogné, M. (1997). Identification of a homolog of the C alpha 3'/hs3 enhancer and of an allelic variant of the 3'IgH/hs1,2 enhancer downstream of the human immunoglobulin alpha 1 gene. *Eur. J. Immunol.* 27, 2981–2985.

Pinaud, E., Khamlichi, A.A., Morvan, C.L., Drouet, M., Nalesso, V., Bert, M.L., and Cogné, M. (2001). Localization of the 3' IgH Locus Elements that Effect Long-Distance Regulation of Class Switch Recombination. *Immunity* 15, 187–199.

Pinaud, E., Marquet, M., Fiancette, R., Péron, S., Vincent-Fabert, C., Denizot, Y., and Cogné, M. (2011). The IgH locus 3' regulatory region: pulling the strings from behind. *Adv. Immunol.* 110, 27–70.

Pipek, O., D. Ribli, J. Molnár, Á. Póti, M. Krzystanek, A. Bodor, G. E. Tusnády, Z. Szallasi, I. Csabai, and D. Szüts. 2017. Fast and accurate mutation detection in whole genome sequences of multiple isogenic samples with IsoMut. *BMC Bioinformatics* 18: 73.

Poltoratsky, V., Prasad, R., Horton, J.K., and Wilson, S.H. (2007). Down-regulation of DNA polymerase beta accompanies somatic hypermutation in human BL2 cell lines. *DNA Repair* 6, 244–253.

Pommier, Y., Sun, Y., Huang, S.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* 17, 703–721.

Porter, R.R. (1973). Structural studies of immunoglobulins. *Science* 180, 713–716.

Proudhon, C., B. Hao, R. Raviram, J. Chaumeil, and J.A. Skok. 2015. Long-Range Regulation of V(D)J Recombination. *Adv. Immunol.* 128:123–182. doi:10.1016/bs.ai.2015.07.003.

Pucella, J.N., and Chaudhuri, J. (2017). AID Invited to the G4 Summit. *Mol. Cell* 67, 355–357.

## Q

Qiao, Q., Wang, L., Meng, F.-L., Hwang, J.K., Alt, F.W., and Wu, H. (2017). AID Recognizes Structured DNA for Class Switch Recombination. *Mol. Cell* 67, 361–373.e4.

Quinodoz, S., and Guttman, M. (2014). Long non-coding RNAs: An emerging link between gene regulation and nuclear organization. *Trends Cell Biol.* 24, 651–663.

## R

Rada, C., Ehrenstein, M.R., Neuberger, M.S., and Milstein, C. (1998). Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* 9, 135–141.



- Rada, C., Jarvis, J.M., and Milstein, C. (2002a). AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 7003–7008.
- Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002b). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol. CB* *12*, 1748–1755.
- Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell* *16*, 163–171.
- Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.-T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., et al. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature* *440*, 105–109.
- Ranuncolo, S.M., Polo, J.M., Dierov, J., Singer, M., Kuo, T., Grealley, J., Green, R., Carroll, M., and Melnick, A. (2007). Bcl-6 mediates the germinal center B cell phenotype and lymphomagenesis through transcriptional repression of the DNA-damage sensor ATR. *Nat. Immunol.* *8*, 705–714.
- Ratliff, M.L., Templeton, T.D., Ward, J.M., and Webb, C.F. (2014a). The Bright Side of Hematopoiesis: Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. *Front. Immunol.* *5*.
- Raybuck, A.L., Cho, S.H., Li, J., Rogers, M.C., Lee, K., Williams, C.L., Shlomchik, M., Thomas, J.W., Chen, J., Williams, J.V., et al. (2018). B Cell-Intrinsic mTORC1 Promotes Germinal Center-Defining Transcription Factor Gene Expression, Somatic Hypermutation, and Memory B Cell Generation in Humoral Immunity. *J. Immunol. Baltim. Md 1950* *200*, 2627–2639.
- Razin, S.V., Iarovaia, O.V., and Vassetzky, Y.S. (2014). A requiem to the nuclear matrix: from a controversial concept to 3D organization of the nucleus. *Chromosoma* *123*, 217–224.
- Rebhandl, S., Huemer, M., Greil, R., and Geisberger, R. (2015). AID/APOBEC deaminases and cancer. *Oncoscience* *2*, 320–333.
- Reth, M.G., and Alt, F.W. (1984). Novel immunoglobulin heavy chains are produced from DJH gene segment rearrangements in lymphoid cells. *Nature* *312*, 418–423.
- Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A., et al. (2000). Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2). *Cell* *102*, 565–575.
- Reynaud, D., Demarco, I.A., Reddy, K.L., Schjerven, H., Bertolino, E., Chen, Z., Smale, S.T., Winandy, S., and Singh, H. (2008). Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nat. Immunol.* *9*, 927–936.
- Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A.E., Sattentau, Q., Meinhart, A., and Proudfoot, N.J. (2018). RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol. Cell* *70*, 650–662.e8.
- Ritland Politz, J.C., Scalzo, D., and Groudine, M. (2016). The redundancy of the mammalian heterochromatic compartment. *Curr. Opin. Genet. Dev.* *37*, 1–8.
- Roa, S., Kuang, F.L., and Scharff, M.D. (2008a). Does antisense make sense of AID targeting? *Proc. Natl. Acad. Sci. U. S. A.* *105*, 3661–3662.



- Roa, S., Avdievich, E., Peled, J.U., Maccarthy, T., Werling, U., Kuang, F.L., Kan, R., Zhao, C., Bergman, A., Cohen, P.E., et al. (2008b). Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 16248–16253.
- Robbiani, D.F., and Nussenzweig, M.C. (2013). Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. *Annu. Rev. Pathol.* *8*, 79–103.
- Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cell. Mol. Life Sci. CMLS* *66*, 981–993.
- Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. Integrative genomics viewer. *Nat. Biotechnol.* *29*: 24–26.
- Rocha, P.P., Micsinai, M., Kim, J.R., Hewitt, S.L., Souza, P.P., Trimarchi, T., Strino, F., Parisi, F., Kluger, Y., and Skok, J.A. (2012). Close proximity to Igh is a contributing factor to AID-mediated translocations. *Mol. Cell* *47*, 873–885.
- Rogakou, E.P., Boon, C., Redon, C., and Bonner, W.M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J. Cell Biol.* *146*, 905–916.
- Roldán, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., and Skok, J.A. (2005). Locus “decontraction” and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat. Immunol.* *6*, 31–41.
- Romanello, M., Schiavone, D., Frey, A., and Sale, J.E. (2016). Histone H3.3 promotes IgV gene diversification by enhancing formation of AID-accessible single-stranded DNA. *EMBO J.* *35*, 1452–1464.
- Romanow, W.J., Langerak, A.W., Goebel, P., Wolvers-Tettero, I.L.M., Dongen, J.J.M. van, Feeney, A.J., and Murre, C. (2000). E2A and EBF Act in Synergy with the V(D)J Recombinase to Generate a Diverse Immunoglobulin Repertoire in Nonlymphoid Cells. *Mol. Cell* *5*, 343–353.
- Ronai, D., Berru, M., and Shulman, M.J. (1999). Variegated expression of the endogenous immunoglobulin heavy-chain gene in the absence of the intronic locus control region. *Mol. Cell. Biol.* *19*, 7031–7040.
- Ronai, D., Iglesias-Ussel, M.D., Fan, M., Shulman, M.J., and Scharff, M.D. (2005a). Complex regulation of somatic hypermutation by cis-acting sequences in the endogenous IgH gene in hybridoma cells. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 11829–11834.
- Ronai, D., Iglesias-Ussel, M.D., Fan, M., Li, Z., Martin, A., and Scharff, M.D. (2007). Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *J. Exp. Med.* *204*, 181–190.
- Rother, M.B., Palstra, R.-J., Jhunjunwala, S., van Kester, K.A.M., van Ijcken, W.F.J., Hendriks, R.W., van Dongen, J.J.M., Murre, C., and van Zelm, M.C. (2016). Nuclear positioning rather than contraction controls ordered rearrangements of immunoglobulin loci. *Nucleic Acids Res.* *44*, 175–186.
- Rothschild, G., and Basu, U. (2017). Lingering Questions about Enhancer RNA and Enhancer Transcription-Coupled Genomic Instability. *Trends Genet.* *33*, 143–154.
- Rouaud, P., Vincent-Fabert, C., Fiancette, R., Cogne, M., Pinaud, E., and Denizot, Y. (2012). Enhancers Located in Heavy Chain Regulatory Region (hs3a, hs1,2, hs3b, and hs4) Are Dispensable for Diversity of VDJ Recombination. *J. Biol. Chem.* *287*, 8356–8360.





Rouaud, P., Vincent-Fabert, C., Saintamand, A., Fiancette, R., Marquet, M., Robert, I., Reina-San-Martin, B., Pinaud, E., Cogné, M., and Denizot, Y. (2013). The IgH 3' regulatory region controls somatic hypermutation in germinal center B cells. *J. Exp. Med.* *210*, 1501–1507.

Roy, A.L., Sen, R., and Roeder, R.G. (2011). Enhancer-promoter communication and transcriptional regulation of Igh. *Trends Immunol.* *32*, 532–539.

## S

Saintamand, A., Rouaud, P., Saad, F., Rios, G., Cogné, M., and Denizot, Y. (2015). Elucidation of IgH 3' region regulatory role during class switch recombination via germline deletion. *Nat. Commun.* *6*, 7084.

Saintamand, A., Vincent-Fabert, C., Garot, A., Rouaud, P., Oruc, Z., Magnone, V., Cogné, M., and Denizot, Y. (2016). Deciphering the importance of the palindromic architecture of the immunoglobulin heavy-chain 3' regulatory region. *Nat. Commun.* *7*, 10730.

Saito, M., Gao, J., Basso, K., Kitagawa, Y., Smith, P.M., Bhagat, G., Pernis, A., Pasqualucci, L., and Dalla-Favera, R. (2007). A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma. *Cancer Cell* *12*, 280–292.

Saito, M., Novak, U., Piovan, E., Basso, K., Sumazin, P., Schneider, C., Crespo, M., Shen, Q., Bhagat, G., Califano, A., et al. (2009). BCL6 suppression of BCL2 via Miz1 and its disruption in diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 11294–11299.

Sakai, E., Bottaro, A., Davidson, L., Sleckman, B.P., and Alt, F.W. (1999a). Recombination and transcription of the endogenous Ig heavy chain locus is effected by the Ig heavy chain intronic enhancer core region in the absence of the matrix attachment regions. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 1526–1531.

Sakai, E., Bottaro, A., and Alt, F.W. (1999b). The Ig heavy chain intronic enhancer core region is necessary and sufficient to promote efficient class switch recombination. *Int. Immunol.* *11*, 1709–1713.

Saribasak, H., and Gearhart, P.J. (2012). Does DNA repair occur during somatic hypermutation? *Semin. Immunol.* *24*, 287–292.

Saribasak, H., Maul, R.W., Cao, Z., Yang, W.W., Schenten, D., Kracker, S., and Gearhart, P.J. (2012). DNA polymerase  $\zeta$  generates tandem mutations in immunoglobulin variable regions. *J. Exp. Med.* *209*, 1075–1081.

Sasaki, Y., Casola, S., Kutok, J.L., Rajewsky, K., and Schmidt-Supprian, M. (2004). TNF family member B cell-activating factor (BAFF) receptor-dependent and -independent roles for BAFF in B cell physiology. *J. Immunol. Baltim. Md 1950* *173*, 2245–2252.

Satoh, Y., Yokota, T., Sudo, T., Kondo, M., Lai, A., Kincade, P.W., Kouro, T., Iida, R., Kokame, K., Miyata, T., et al. (2013). The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. *Immunity* *38*, 1105–1115.

Savarese, F., Dávila, A., Nechanitzky, R., De La Rosa-Velazquez, I., Pereira, C.F., Engelke, R., Takahashi, K., Jenuwein, T., Kohwi-Shigematsu, T., Fisher, A.G., et al. (2009). Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. *Genes Dev.* *23*, 2625–2638.

Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* *11*, 251–263.



- Scheuermann, R.H., and Chen, U. (1989). A developmental-specific factor binds to suppressor sites flanking the immunoglobulin heavy-chain enhancer. *Genes Dev.* *3*, 1255–1266.
- Scheuermann, R.H., and Garrard, W.T. (1999). MARs of antigen receptor and co-receptor genes. *Crit. Rev. Eukaryot. Gene Expr.* *9*, 295–310.
- Schlissel, M.S. (2004). Regulation of activation and recombination of the murine Igkappa locus. *Immunol. Rev.* *200*, 215–223.
- Schlissel, M.S., Corcoran, L.M., and Baltimore, D. (1991). Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. *J. Exp. Med.* *173*, 711–720.
- Schroeder, H.W., and Cavacini, L. (2010). Structure and Function of Immunoglobulins. *J. Allergy Clin. Immunol.* *125*, S41–S52.
- Schwickert, T.A., Tagoh, H., Gültekin, S., Dakic, A., Axelsson, E., Minnich, M., Ebert, A., Werner, B., Roth, M., Cimmino, L., et al. (2014). Stage-specific control of early B cell development by the transcription factor Ikaros. *Nat. Immunol.* *15*, 283–293.
- Seifert, M., Scholtysik, R., and Küppers, R. (2013). Origin and pathogenesis of B cell lymphomas. *Methods Mol. Biol. Clifton NJ* *971*, 1–25.
- Sen, R., and Oltz, E. (2006). Genetic and epigenetic regulation of IgH gene assembly. *Curr. Opin. Immunol.* *18*, 237–242.
- Serwe, M., and Sablitzky, F. (1993). V (D) J recombination in B cells is impaired but not blocked by targeted deletion of the immunoglobulin heavy chain intron enhancer. *EMBO J.* *12*, 2321.
- Shaffer, A.L., Yu, X., He, Y., Boldrick, J., Chan, E.P., and Staudt, L.M. (2000). BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity* *13*, 199–212.
- Shaknovich, R., Cerchietti, L., Tsikitas, L., Kormaksson, M., De, S., Figueroa, M.E., Ballon, G., Yang, S.N., Weinhold, N., Reimers, M., et al. (2011). DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* *118*, 3559–3569.
- Shankar, M., Nixon, J.C., Maier, S., Workman, J., Farris, A.D., and Webb, C.F. (2007). Anti-Nuclear Antibody Production and Autoimmunity in Transgenic Mice that Over-Express the Transcription Factor Bright. *J. Immunol. Baltim. Md 1950* *178*, 2996–3006.
- Shen, H. M., A. Peters, B. Baron, X. Zhu, and U. Storb. 1998. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* *280*: 1750–1752.
- Shen, H.M., and Storb, U. (2004). Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 12997–13002.
- Shen, H.M., Tanaka, A., Bozek, G., Nicolae, D., and Storb, U. (2006). Somatic Hypermutation and Class Switch Recombination in Msh6<sup>-/-</sup>Ung<sup>-/-</sup> Double-Knockout Mice. *J. Immunol.* *177*, 5386–5392.
- Sheppard, E.C., Morrish, R.B., Dillon, M.J., Leyland, R., and Chahwan, R. (2018). Epigenomic Modifications Mediating Antibody Maturation. *Front. Immunol.* *9*.
- Shimamoto, R., Amano, N., Ichisaka, T., Watanabe, A., Yamanaka, S., and Okita, K. (2014). Generation and Characterization of Induced Pluripotent Stem Cells from Aid-Deficient Mice. *PLoS ONE* *9*.



- Shinkai, Y., Rathbun, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., and Stall, A.M. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855–867.
- Shinkura, R., Tian, M., Smith, M., Chua, K., Fujiwara, Y., and Alt, F.W. (2003). The influence of transcriptional orientation on endogenous switch region function. *Nat. Immunol.* **4**, 435–441.
- Sinclair, A.M., Lee, J.A., Goldstein, A., Xing, D., Liu, S., Ju, R., Tucker, P.W., Neufeld, E.J., and Scheuermann, R.H. (2001). Lymphoid apoptosis and myeloid hyperplasia in CCAAT displacement protein mutant mice. *Blood* **98**, 3658–3667.
- Skok, J.A., Brown, K.E., Azuara, V., Caparros, M.-L., Baxter, J., Takacs, K., Dillon, N., Gray, D., Perry, R.P., Merckenschlager, M., et al. (2001). Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. *Nat. Immunol.* **2**, 848–854.
- Skowronska-Krawczyk, D., and Rosenfeld, M.G. (2015). Nuclear matrix revisited? *Cell Cycle Georget. Tex* **14**, 1487–1488.
- Smith, K.G., Hewitson, T.D., Nossal, G.J., and Tarlinton, D.M. (1996). The phenotype and fate of the antibody-forming cells of the splenic foci. *Eur. J. Immunol.* **26**, 444–448.
- Song, H., and Cerny, J. (2003). Functional Heterogeneity of Marginal Zone B Cells Revealed by Their Ability to Generate Both Early Antibody-forming Cells and Germinal Centers with Hypermutation and Memory in Response to a T-dependent Antigen. *J. Exp. Med.* **198**, 1923–1935.
- Souabni, A., Cobaleda, C., Schebesta, M., and Busslinger, M. (2002). Pax5 promotes B lymphopoiesis and blocks T cell development by repressing Notch1. *Immunity* **17**, 781–793.
- Spilianakis, C.G., M.D. Lalioti, T. Town, G.R. Lee, and R.A. Flavell. 2005. Interchromosomal associations between alternatively expressed loci. *Nature*. **435**:637–645. doi:10.1038/nature03574.
- Staszewski, O., Baker, R.E., Ucher, A.J., Martier, R., Stavnezer, J., and Guikema, J.E.J. (2011). Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig loci in activated B cells. *Mol. Cell* **41**, 232–242.
- Stavnezer, J., J. E. J. Guikema, and C. E. Schrader. 2008. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* **26**: 261–292.
- Stavnezer, J., and Schrader, C.E. (2014). IgH chain class switch recombination: mechanism and regulation. *J. Immunol. Baltim. Md 1950* **193**, 5370–5378.
- Stavnezer, J., Linehan, E.K., Thompson, M.R., Habboub, G., Ucher, A.J., Kadungure, T., Tsuchimoto, D., Nakabeppu, Y., and Schrader, C.E. (2014). Differential expression of APE1 and APE2 in germinal centers promotes error-prone repair and A:T mutations during somatic hypermutation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9217–9222.
- van Steensel, B., and Belmont, A.S. (2017). Lamina-associated domains: links with chromosome architecture, heterochromatin and gene repression. *Cell* **169**, 780–791.
- Storb, U. (2014). Chapter Seven - Why Does Somatic Hypermutation by AID Require Transcription of Its Target Genes? In *Advances in Immunology*, F.W. Alt, ed. (Academic Press), pp. 253–277.



Storb, U., Haasch, D., Arp, B., Sanchez, P., Cazenave, P.A., and Miller, J. (1989). Physical linkage of mouse lambda genes by pulsed-field gel electrophoresis suggests that the rearrangement process favors proximate target sequences. *Mol. Cell. Biol.* *9*, 711–718.

Storb, U., Shen, H.M., Michael, N., and Kim, N. (2001). Somatic hypermutation of immunoglobulin and non-immunoglobulin genes. *Philos. Trans. R. Soc. Lond. Ser. B* *356*, 13–19.

Strongin, D.E., Groudine, M., and Politz, J.C.R. (2014). Nucleolar tethering mediates pairing between the IgH and Myc loci. *Nucl. Austin Tex* *5*, 474–481.

Stucki, M., Pascucci, B., Parlanti, E., Fortini, P., Wilson, S.H., Hübscher, U., and Dogliotti, E. (1998). Mammalian base excision repair by DNA polymerases delta and epsilon. *Oncogene* *17*, 835–843.

Su, T.T., Guo, B., Wei, B., Braun, J., and Rawlings, D.J. (2004). Signaling in transitional type 2 B cells is critical for peripheral B-cell development. *Immunol. Rev.* *197*, 161–178.

## T

Ta, V.-T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., et al. (2003). AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat. Immunol.* *4*, 843–848.

Tanaka, A., Shen, H.M., Ratnam, S., Kodgire, P., and Storb, U. (2010). Attracting AID to targets of somatic hypermutation. *J. Exp. Med.* *207*, 405–415.

Tangye, S.G. (2011). Staying alive: regulation of plasma cell survival. *Trends Immunol.* *32*, 595–602.

Tantin, D., Tussie-Luna, M.I., Roy, A.L., and Sharp, P.A. (2004). Regulation of Immunoglobulin Promoter Activity by TFII-I Class Transcription Factors. *J. Biol. Chem.* *279*, 5460–5469.

Taylor, J.J., Pape, K.A., and Jenkins, M.K. (2012). A germinal center-independent pathway generates unswitched memory B cells early in the primary response. *J. Exp. Med.* *209*, 597–606.

Teague, B.N., Pan, Y., Mudd, P.A., Nakken, B., Zhang, Q., Szodoray, P., Kim-Howard, X., Wilson, P.C., and Farris, A.D. (2007). Cutting edge: Transitional T3 B cells do not give rise to mature B cells, have undergone selection, and are reduced in murine lupus. *J. Immunol. Baltim. Md 1950* *178*, 7511–7515.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* *28*, 621–629.

Teves, S.S., and Henikoff, S. (2014). DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus* *5*, 211–218.

Tidwell, J.A., Schmidt, C., Heaton, P., Wilson, V., and Tucker, P.W. (2011). Characterization of a new ARID family transcription factor (Brightlike/ARID3C) that co-activates Bright/ARID3A-mediated immunoglobulin gene transcription. *Mol. Immunol.* *49*, 260–272.

Tinguely, A., Chemin, G., Péron, S., Sirac, C., Reynaud, S., Cogné, M., and Delpy, L. (2012). Cross talk between immunoglobulin heavy-chain transcription and RNA surveillance during B cell development. *Mol. Cell. Biol.* *32*, 107–117.





Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress *Aicda* by overcoming the effects of silencers. *Nat. Immunol.* *11*, 148–154.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* *329*, 689–693.

Tsukada, S., Sugiyama, H., Oka, Y., and Kishimoto, S. (1990). Estimation of D segment usage in initial D to JH joinings in a murine immature B cell line. Preferential usage of DFL16.1, the most 5' D segment and DQ52, the most JH-proximal D segment. *J. Immunol. Baltim. Md 1950* *144*, 4053–4059.

Tumas-Brundage, K., and Manser, T. (1997). The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. *J. Exp. Med.* *185*, 239–250.

## U

Uchimura, Y., Barton, L.F., Rada, C., and Neuberger, M.S. (2011). REG- $\gamma$  associates with and modulates the abundance of nuclear activation-induced deaminase. *J. Exp. Med.* *208*, 2385–2391.

## V

Vaidyanathan, B., Yen, W.-F., Pucella, J.N., and Chaudhuri, J. (2014). AIDing Chromatin and Transcription-Coupled Orchestration of Immunoglobulin Class-Switch Recombination. *Front. Immunol.* *5*.

Victoria, G.D., Schwickert, T.A., Fooksman, D.R., Kamphorst, A.O., Meyer-Hermann, M., Dustin, M.L., and Nussenzweig, M.C. (2010). Germinal Center Dynamics Revealed by Multiphoton Microscopy Using a Photoactivatable Fluorescent Reporter. *Cell* *143*, 592–605.

Vincent-Fabert, C., Truffinet, V., Fiancette, R., Cogné, N., Cogné, M., and Denizot, Y. (2009). Ig Synthesis and Class Switching Do Not Require the Presence of the hs4 Enhancer in the 3' IgH Regulatory Region. *J. Immunol.* *182*, 6926–6932.

Vincent-Fabert, C., Fiancette, R., Pinaud, E., Truffinet, V., Cogné, N., Cogné, M., and Denizot, Y. (2010). Genomic deletion of the whole IgH 3' regulatory region (hs3a, hs1,2, hs3b, and hs4) dramatically affects class switch recombination and Ig secretion to all isotypes. *Blood* *116*, 1895–1898.

Volpi, S.A., Verma-Gaur, J., Hassan, R., Ju, Z., Roa, S., Chatterjee, S., Werling, U., Hou, H., Will, B., Steidl, U., et al. (2012). Germline deletion of *Igh* 3' regulatory region elements hs5-7 affects B cell specific regulation, rearrangement and insulation of the *Igh* locus. *J. Immunol. Baltim. Md 1950* *188*, 2556–2566.

Vuong, B.Q., Lee, M., Kabir, S., Irimia, C., Macchiarulo, S., McKnight, G.S., and Chaudhuri, J. (2009). Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat. Immunol.* *10*, 420–426.

## W

Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.* *206*, 1817–1830.

Wang, T.-Y., Han, Z.-M., Chai, Y.-R., and Zhang, J.-H. (2010). A mini review of MAR-binding proteins. *Mol. Biol. Rep.* *37*, 3553–3560.



- Wang, X., Fan, M., Kalis, S., Wei, L., and Scharff, M.D. (2014). A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. *Nat. Commun.* *5*, 4137.
- Wang, Z., Goldstein, A., Zong, R.T., Lin, D., Neufeld, E.J., Scheuermann, R.H., and Tucker, P.W. (1999a). Cux/CDP homeoprotein is a component of NF- $\mu$ NR and represses the immunoglobulin heavy chain intronic enhancer by antagonizing the bright transcription activator. *Mol. Cell. Biol.* *19*, 284–295.
- Wang, Z., Goldstein, A., Zong, R.-T., Lin, D., Neufeld, E.J., Scheuermann, R.H., and Tucker, P.W. (1999b). Cux/CDP Homeoprotein Is a Component of NF- $\mu$ NR and Represses the Immunoglobulin Heavy Chain Intronic Enhancer by Antagonizing the Bright Transcription Activator. *Mol. Cell. Biol.* *19*, 284–295.
- Ward, I.M., Reina-San-Martin, B., Oлару, A., Minn, K., Tamada, K., Lau, J.S., Cascalho, M., Chen, L., Nussenzweig, A., Livak, F., et al. (2004). 53BP1 is required for class switch recombination. *J. Cell Biol.* *165*, 459–464.
- Webb, C.F., Das, C., Eneff, K.L., and Tucker, P.W. (1991). Identification of a matrix-associated region 5' of an immunoglobulin heavy chain variable region gene. *Mol. Cell. Biol.* *11*, 5206–5211.
- Webb, C.F., Smith, E.A., Medina, K.L., Buchanan, K.L., Smithson, G., and Dou, S. (1998). Expression of Bright at two distinct stages of B lymphocyte development. *J. Immunol.* *160*, 4747–4754.
- Webb, C.F., Bryant, J., Popowski, M., Allred, L., Kim, D., Harriss, J., Schmidt, C., Miner, C.A., Rose, K., Cheng, H.-L., et al. (2011a). The ARID Family Transcription Factor Bright Is Required for both Hematopoietic Stem Cell and B Lineage Development. *Mol. Cell. Biol.* *31*, 1041–1053.
- Webb, C.F., Bryant, J., Popowski, M., Allred, L., Kim, D., Harriss, J., Schmidt, C., Miner, C.A., Rose, K., Cheng, H.-L., et al. (2011b). The ARID Family Transcription Factor Bright Is Required for both Hematopoietic Stem Cell and B Lineage Development. *Mol. Cell. Biol.* *31*, 1041–1053.
- Weill, J.-C., and Reynaud, C.-A. (2008). DNA polymerases in adaptive immunity. *Nat. Rev. Immunol.* *8*, 302–312.
- Wiersma, E.J., Ronai, D., Berru, M., Tsui, F.W., and Shulman, M.J. (1999). Role of the intronic elements in the endogenous immunoglobulin heavy chain locus. Either the matrix attachment regions or the core enhancer is sufficient to maintain expression. *J. Biol. Chem.* *274*, 4858–4862.
- Will, B., Vogler, T.O., Bartholdy, B., Garrett-Bakelman, F., Mayer, J., Barreyro, L., Pandolfi, A., Todorova, T.I., Okoye-Okafor, U.C., Stanley, R.F., et al. (2013). *Satb1* regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. *Nat. Immunol.* *14*, 437–445.
- Williams, A., C.G. Spilianakis, and R.A. Flavell. 2010. Interchromosomal association and gene regulation in trans. *Trends Genet.* *26*:188–197. doi:10.1016/j.tig.2010.01.007.
- Woo, C.J., Martin, A., and Scharff, M.D. (2003). Induction of Somatic Hypermutation Is Associated with Modifications in Immunoglobulin Variable Region Chromatin. *Immunity* *19*, 479–489.
- Wright, B.E., Schmidt, K.H., Davis, N., Hunt, A.T., and Minnick, M.F. (2008). II. Correlations between secondary structure stability and mutation frequency during somatic hypermutation. *Mol. Immunol.* *45*, 3600–3608.
- Wright, B.E., Schmidt, K.H., Minnick, M.F., and Davis, N. (2008). VH gene transcription creates stabilized secondary structures for coordinated mutagenesis during somatic hypermutation. *Mol. Immunol.* *45*, 3589–3599.



Wu, H., Deng, Y., Feng, Y., Long, D., Ma, K., Wang, X., Zhao, M., Lu, L., and Lu, Q. (2018). Epigenetic regulation in B-cell maturation and its dysregulation in autoimmunity. *Cell. Mol. Immunol.*

Wuerffel, R., Wang, L., Grigera, F., Manis, J., Selsing, E., Perlot, T., Alt, F.W., Cogne, M., Pinaud, E., and Kenter, A.L. (2007). S-S Synapsis during Class Switch Recombination Is Promoted by Distantly Located Transcriptional Elements and Activation-Induced Deaminase. *Immunity* 27, 711–722.

## X

Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qahtani, A., White, C.A., Park, S.-R., et al. (2010). 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat. Struct. Mol. Biol.* 17, 1124–1135.

Xu, Z., Zan, H., Pone, E.J., Mai, T., and Casali, P. (2012). Immunoglobulin class switch DNA recombination: induction, targeting and beyond. *Nat. Rev. Immunol.* 12, 517–531.

Xue, K., Rada, C., and Neuberger, M.S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*<sup>-/-</sup> *ung*<sup>-/-</sup> mice. *J. Exp. Med.* 203, 2085–2094.

## Y

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol.* 12, 62–69.

Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P., et al. (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* 449, 478–482.

Yancopoulos, G.D., and Alt, F.W. (1985). Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* 40, 271–281.

Yang, Q., Riblet, R., and Schildkraut, C.L. (2005). Sites that direct nuclear compartmentalization are near the 5' end of the mouse immunoglobulin heavy-chain locus. *Mol. Cell. Biol.* 25, 6021–6030.

Yang, S.Y., Fugmann, S.D., and Schatz, D.G. (2006). Control of gene conversion and somatic hypermutation by immunoglobulin promoter and enhancer sequences. *J. Exp. Med.* 203, 2919–2928.

Yang, Y., Ghosn, E.E.B., Cole, L.E., Obukhanych, T.V., Sadate-Ngatchou, P., Vogel, S.N., Herzenberg, L.A., and Herzenberg, L.A. (2012). Antigen-specific memory in B-1a and its relationship to natural immunity. *Proc. Natl. Acad. Sci. U. S. A.* 109, 5388–5393.

Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002). SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419, 641–645.

Ye, M., and Graf, T. (2007). Early decisions in lymphoid development. *Curr. Opin. Immunol.* 19, 123–128.

de Yébenes, V.G., Belver, L., Pisano, D.G., González, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J. Exp. Med.* 205, 2199–2206.

Yi, M., Wu, P., Trevorrow, K.W., Claflin, L., and Garrard, W.T. (1999a). Evidence that the I $\kappa$  gene MAR regulates the probability of premature V-J joining and somatic hypermutation. *J. Immunol. Baltim. Md* 1950 162, 6029–6039.



Ying, C.Y., Dominguez-Sola, D., Fabi, M., Lorenz, I.C., Hussein, S., Bansal, M., Califano, A., Pasqualucci, L., Basso, K., and Dalla-Favera, R. (2013). MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nat. Immunol.* *14*, 1084–1092.

Yokota, T., and Kanakura, Y. (2014). Role of tissue-specific AT-rich DNA sequence-binding proteins in lymphocyte differentiation. *Int. J. Hematol.* *100*, 238–245.

Yoshida, T., Ng, S.Y.-M., Zuniga-Pflucker, J.C., and Georgopoulos, K. (2006). Early hemopoietic lineage restrictions directed by Ikaros. *Nat. Immunol.* *7*.

Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* *4*, 442–451.

Yusufzai, T.M., and Felsenfeld, G. (2004). The 5'-HS4 chicken  $\beta$ -globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc. Natl. Acad. Sci.* *101*, 8620–8624.

Zaller, D.M., and Eckhardt, L.A. (1985). Deletion of a B-cell-specific enhancer affects transfected, but not endogenous, immunoglobulin heavy-chain gene expression. *Proc. Natl. Acad. Sci. U. S. A.* *82*, 5088–5092.

## Z

Zan, H., and Casali, P. (2013). Regulation of Aicda expression and AID activity. *Autoimmunity* *46*, 83–101.

Zentner, G.E., Tesar, P.J., and Scacheri, P.C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* *21*, 1273–1283.

Zhang, J., Bottaro, A., Li, S., Stewart, V., and Alt, F.W. (1993). A selective defect in IgG2b switching as a result of targeted mutation of the I gamma 2b promoter and exon. *EMBO J.* *12*, 3529–3537.

Zhang, Z.Z., Pannunzio, N.R., Han, L., Hsieh, C.-L., Yu, K., and Lieber, M.R. (2014). The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites. *Cell Rep.* *8*, 557–569.

Zhao, K., Harel, A., Stuurman, N., Guedalia, D., and Gruenbaum, Y. (1996). Binding of matrix attachment regions to nuclear lamin is mediated by the rod domain and depends on the lamin polymerization state. *FEBS Lett.* *380*, 161–164.

Zhao, Y., Dunn-Walters, D.K., Barone, F., and Spencer, J. (2009). Antisense transcripts of V(D)J rearrangements; artifacts caused by false priming? *Mol. Immunol.* *46*, 2357–2362.

Zheng, S., Vuong, B.Q., Vaidyanathan, B., Lin, J.-Y., Huang, F.-T., and Chaudhuri, J. (2015). Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA. *Cell* *161*, 762–773.

Zhuang, Y., Soriano, P., and Weintraub, H. (1994). The helix-loop-helix gene E2A is required for B cell formation. *Cell* *79*, 875–884.

Zong, R.T., and Scheuermann, R.H. (1995). Mutually exclusive interaction of a novel matrix attachment region binding protein and the NF- $\mu$ NR enhancer repressor. Implications for regulation of immunoglobulin heavy chain expression. *J. Biol. Chem.* *270*, 24010–24018.





## Résumé

---

En introduisant fréquemment des mutations ponctuelles dans les régions variables des gènes d'immunoglobulines (Ig), le processus d'hypermutation somatique (SHM, initié par la déaminase AID) est essentiel pour augmenter l'affinité des anticorps. En marge de ses cibles physiologiques (les gènes d'Ig), AID peut induire des « dommages collatéraux » au niveau de cibles « illégitimes » qui sont appelées « off targets » (dont certains oncogènes, tel que *Bcl6* fréquemment muté dans les lymphomes B). Le risque élevé de dommages collatéraux dans le génome des cellules B implique que les remaniements géniques soient précisément surveillés. Parmi les éléments *cis*-régulateurs impliqués dans cette surveillance, on compte l'activateur cE $\mu$  au locus des chaînes lourdes des Ig (*IgH*) et ses régions flanquantes d'attachement à la matrice nucléaire MARs<sub>E $\mu$</sub>  (étudiés en détails dans nos modèles de souris KO). Nous montrons que la délétion des régions MARs<sub>E $\mu$</sub>  diminue non seulement les mutations au locus des chaînes lourdes des Ig (effet physiologique en *cis*) mais également au locus des chaînes légères *Ig $\kappa$*  situé sur un chromosome différent (effet en *trans*). A l'aide d'un outil bio-informatique (DeMinEr) que nous avons développé dans le but d'identifier des mutations rares, nous montrons également que les régions MARs<sub>E $\mu$</sub>  sont impliquées dans les dommages collatéraux infligés aux « off targets » des cellules B. Grâce à la technique de FISH 3D, nous proposons que les régions MARs<sub>E $\mu$</sub>  participent à la régulation de la SHM en influençant la position des cibles de AID dans le noyau des cellules B. Notre étude met en évidence un niveau de régulation spatiale du processus de SHM médié par les régions MARs<sub>E $\mu$</sub>  du locus *IgH*.

---

Mots-clés : B lymphocytes, *IgH* locus, MARs<sub>E $\mu$</sub>  regions, somatic hypermutation.

## Abstract

---

By introducing frequent point mutations into the variable regions of immunoglobulin (Ig) genes, somatic hypermutation (SHM, initiated by the AID deaminase) is a driving force for antibody affinity maturation. It is now admitted that AID-induced mutations in germinal centre B cells could affect, in parallel to their Ig genes physiological targets, illegitimate targets (including oncogenes) so called « off targets » (such as *Bcl6* with frequent point mutation in B lymphomas). The high risk of « collateral damage » in the B cell genome implies that remodeling events are precisely surveyed. Among *cis*-regulatory elements involved (transcriptional enhancers and chromatin isolators and anchors...), one best candidate is the intronic region including the cE $\mu$  enhancer and its flanking MARs<sub>E $\mu$</sub>  regions, that we have been studying extensively using mouse KO model. We recently showed that MARs<sub>E $\mu$</sub>  deletion decreases SHM not only at Ig Heavy chain locus *IgH* (physiological *cis* effect) but surprisingly also at the Ig Light chain Kappa locus *Ig $\kappa$* , located on a different chromosome (*trans* effect). To extend the study of this intriguing *trans* effect, we developed a bioinformatic tool called DeMinEr that unveiled that MARs<sub>E $\mu$</sub>  regions were also involved in AID-induced collateral damages to « off-targets ». Using FISH 3D, we show that MARs<sub>E $\mu$</sub>  regions harboured the potential not only to locally recruit SHM but also to cause dynamic changes of nuclear structures. The surprising *cis* and *trans* effect of MARs<sub>E $\mu$</sub>  deletion, impacting simultaneously nuclear positioning and SHM, revealed an additional level of regulation for targeting mutations to Ig and « off-targets » genes.

---

Keywords : B lymphocytes, *IgH* locus, MARs<sub>E $\mu$</sub>  regions, somatic hypermutation.

