



**HAL**  
open science

# Fonctions des protéines liant la matrice nucléaire dans l'organisation et l'intégrité du génome des cellules B

Morgane Thomas

► **To cite this version:**

Morgane Thomas. Fonctions des protéines liant la matrice nucléaire dans l'organisation et l'intégrité du génome des cellules B. Médecine humaine et pathologie. Université de Limoges, 2022. Français. NNT : 2022LIMO0071 . tel-04816690

**HAL Id: tel-04816690**

**<https://theses.hal.science/tel-04816690v1>**

Submitted on 3 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Université de Limoges**

**ED 652 – OMEGALIM Biologie, Chimie, Santé (BCS)**

**CNRS UMR 7276 INSERM U1262 – Laboratoire CRIBL : Contrôle de la Réponse Immune B et des Lymphoproliférations – B NATION : B cell Nuclear Architecture, Ig genes and Oncogenes**

Thèse pour obtenir le grade de

**Docteur de l'Université de Limoges**

Spécialité : Immunologie, oncologie, inflammation et infectiologie

Présentée et soutenue par

**Morgane Thomas**

Le 21 Octobre 2022

**Fonctions des protéines liant la matrice nucléaire dans  
l'organisation et l'intégrité du génome des cellules B.**

Thèse dirigée par le Dr. Eric PINAUD et le Dr. Sandrine LE NOIR

JURY :

Rapporteurs

Dr. Dominique PAYET-BORNET, Directeur de Recherche CNRS, CIML, Marseille

Dr. Sébastien STORCK, Maître de Conférence, INSERM U1151, Paris

Examineurs

Dr. Jean-Christophe ANDRAU, Directeur de Recherche CNRS, IGMM, Montpellier

Dr. Stéphane MANCINI, Chargé de Recherche, INSERM U1236, Rennes

Invités

Dr. Éric PINAUD, Directeur de Recherche CNRS, CRIBL, Limoges

Dr. Sandrine LE NOIR, Chargée de Recherche, CRIBL, Limoges









*A mes nièces, Naelle & Léa,  
Parce que l'innocence d'un enfant vous ramène à l'essentiel...*



*Face à la roche, le ruisseau l'emporte toujours,  
non pas par la force mais par la persévérance.  
H. Jackson Brown*

*Even when I lose I'm winning,  
Because I give all of me.  
J.L*

*Siempre que llueve escampa.*





## Remerciements

---

*Les remerciements ... Ces quelques pages qui font partie des petits plus de la thèse. Ces quelques paragraphes écrits en fin de période intensive de rédaction, ceux dont on dit « t'as plus que ça à écrire, c'est bon alors t'as fini » et pourtant... personne ne le dit mais c'est bien la partie qui sera la plus souvent lue et puis finalement, encore une de ses choses qu'on ne sait que lorsqu'on y est : pas sûr que ce soit la simple à rédiger...*

*Pour mes débuts au labo, Merci Christelle, la plus cool des maîtres de stage. Merci de m'avoir fait découvrir ce métier et de m'avoir si bien accueillie, merci pour ta disponibilité, ton écoute, ta simplicité, ta confiance et merci de ne pas avoir glissé à mes chefs actuels que je pouvais m'avérer être un boulet ☺. Et enfin merci de m'avoir donné envie de continuer. J'ai bien essayé de reproduire la même chose avec mes stagiaires mais je suis loin d'être sûre du résultat ☺.*

*La suite se poursuit avec vous... Éric, Sandrine, je commencerais, déjà, par vous dire merci de m'avoir donné cette opportunité. Merci Éric pour ta patience et ta confiance, pour avoir toléré mon impatience et mes manips clandestines, pour m'avoir toujours laissée l'occasion de dire ce que je pensais et de donner mon avis dans les bonnes périodes comme dans les mauvaises même si « Il est bon de parler et meilleur de se taire, mais tous deux sont mauvais alors qu'ils sont outrés » (JDLF). Merci d'avoir accepté mes blagues un peu passées même dans les moments sérieux, de m'avoir forcée à réfléchir autrement, de m'avoir laissé autant d'autonomie et d'avoir ponctué ton enseignement de toutes ces phrases qui restent en mémoire, je continuerai à me mettre dans mes tubes et surtout (pour le mot de la fin) à en\*\*\*\*\* les mouches. Sandrine, merci de m'avoir poussée à m'investir dans la vie du labo et tous les aspects au-delà de la science et merci d'avoir rendu la bio-info un peu moins obscure pour moi, même s'il reste du travail -'.*

*Bien sûr la vie au labo n'aurait pas été la même sans vous, Merci à toutes ces tatas de CRJBL dont la team mousquetaire (ou le bureau des plaintes / on pose nos vacances FRS ☺) :*

*Tata Carrion, merci de m'avoir éclairée sur la cyto, d'avoir parlé de la bouchure de temps en temps et d'avoir enrichi mon vocabulaire berrichon ! J'ai encore un panel de gag/blague/coup tordu à tester et comme tu as su réceptionner tout ça magnifiquement bien pendant mes années de thèse, n'oublie pas que post-soutenance, il me reste quelques mois de contrat et que le plus gros reste à venir et je sais que la relève est là, elle va assurer. Je ne voudrais pas que tu t'ennuies une fois que je serai partie ☺. Cendrine, D'Artagnan ☺ je te remercie pour ta formation LSI Sesque, nos bières apéritives et nos debriefs en tout genre. Merci d'avoir été une oreille attentive. Je suis sûre que les souris ne te manquent pas mais j'espère que tes anciens collègues te manquent un peu quand même.*

*Christelle D, le western trop rigolo (Dui on ne te l'a jamais faite celle-là, @copyright à Charlye pour Ficello, moi j'ai ajouté la touche western, et pour la réf c'est le fromage Ficello si jamais cette blague nécessitait des explications, bien que je trouve qu'elle se suffit à elle-même ☺). Merci pour ta patience et ta pédagogie incroyable lors de ces mises au point westernblotesque, je pense que la prot' ne sera jamais mon amie et j'avoue que je me suis découragée à rivaliser avec la qualité de tes blots mais j'ai survécu ☺. J'espère que tu conserveras longtemps notre petit totem du western et que tu auras de nombreuses occasions d'expliquer aux futurs doctorants CRJBLein « pourquoi ce totem » ☺. Merci à ma partenaire du Green, l'ultime*







*permise d'être pleinement moi dans cet endroit où l'on a passé beaucoup trop de temps pour laisser toute notre folie à la maison. On évalue les liens solides à leur résistance à la distance... on est toujours là presque deux ans après ton départ et qu'est-ce qu'il a été important pour moi que tu redeviennes limougeaude en cette fin de thèse. Merci de ton soutien inébranlable et de ta gaieté au quotidien. Je lève mon verre à cette équipe qu'on était, Tic & Tac, à nos événements apéro CRJBL (j'espère que la relève sera assurée), à ces discussions réconfortantes, à nos conseils en manip, à ta facilité de rigoler et de me suivre dans tous mes délires, à ta bonne humeur résistante à toutes épreuves (parce que je t'ai vu rire en pleurant ou pleurer en rigolant 😊), à nos sushis partagés mais surtout à tout ce qui nous attend maintenant et pour la suite.*

*Enfin, Ma petite Charlye, ce pilier gauche (dans tous les sens du terme 😊). Ça été une force de t'avoir à mes côtés ! Une force technique (on ne va pas se mentir) et même si on s'est toujours entraïdées je pense que je te dois encore quelques heures de manips. A l'heure où j'écris ce petit paragraphe, je ne regrette aucune de nos manips border Line : nos stims lancés à minuit, les extinctions du trieur à deux heures du matin et même nos grosses sessions dissections à 7h tapantes parce que quand t'en fais 30 t'es plus à 2 ou 3 près !! C'est dans toutes nos galères que j'ai construit mes meilleurs souvenirs de thèse, et même si on l'a déjà dit : t'as quand même eu de la chance de tomber sur aussi barrée que toi 😊. Je ne compte plus toutes les fois où tu m'as dit "attend je vais t'aider !", ni nos conversations qui dépassent l'entendement, parce qu'on sait des choses l'une sur l'autre qui n'ont vraiment aucun intérêt 😊. Merci d'avoir été mon premier rempart pour parer mes réactions excessives ! Je sais que tu n'as jamais écouté aucun de mes conseils 😊 mais si jamais tu t'ennuies et que t'as le temps de suivre mon dernier conseil de doctorante : cours 😊 ! Rédige comme un running, c'est juste "un pied après l'autre" et "si j'ai fait 7 km je peux pousser à 8" même sans les Hoka (j'ai dû modifier cette fin de phrase 😊). Et enfin n'oublies surtout pas que ceux qui font 800 mètres font toujours plus que ceux qui reculent devant la ligne de départ et que quoi qu'il arrive, on a deux bras, deux jambes et on respire. Ah et surtout Salsate, Bachatate et Kizoumbate 😊.*

*Je tiens aussi à remercier toutes ces personnes extérieures qui sont d'un soutien infini et qui croient parfois plus en nous qu'on ne saura jamais le faire nous-même. Ces personnes à qui on se lie sans trop savoir pourquoi au début puis que l'on choisit pour partager les bons côtés de la vie et se soutenir dans les mauvais. Un premier grand merci à tous mes amis...*

*Merci Mathilde, mon amie d'enfance, je n'ai pas besoin d'en dire long, une case de plus cochée ensemble, vivement la prochaine 😊.*

*Ma Nono ! 11 ans d'amitié sans faille, seulement et déjà à la fois ! Merci pour le plus fidèle des soutiens que tu m'as apportés, bien avant et pendant ma thèse. T'as été un indispensable à ma réussite, parce qu'on a commencé ensemble, à bachoter intensément en licence ! Merci pour nos soirées révisions, nos soirées tout court, nos discussions reboostantes, ton incroyable solidarité, ta maturité sur la vie et cette énorme solidité exemplaire. Sache que je suis fière de toi et heureuse de te voir épanouie. Et à bientôt petite Crevette 😊*

*Merci également à mes rencontres Australiennes : dans mes débuts compliqués à Perth, Clément, Baptiste et Ma team BAFMDCO, Evan, Coco et mon Adèle, Mes partenaires de Coinda, Adrien, Thomas et Lucas ! Vous êtes un morceau de la plus*



*belle aventure de ma vie, merci pour ce road trip hors norme ! Mon Australie n'aurait jamais pu être aussi belle sans vous. 9 mois intenses d'un panel d'émotions infinies, de rencontres, de nuit à la belle étoile (et qu'elles étaient belles ces étoiles !), de cette traversée de feu, de magnum almond, de ces kilomètres avalés, des pink lakes, cette panne de voiture, des déserts à perte de vue, des PJNACLES, de Perth à Melbourne en passant par Darwin et Uluru. Merci de m'avoir fait grandir, de m'avoir ouvert les yeux sur la beauté de ce monde, de m'avoir donné ces clés pour tout entreprendre, et merci pour cette folie raisonnée. Et un gros merci spécial à toi mon Adé qui m'a appris à tout relativiser, parce que rien n'est fini tant qu'on l'a pas choisi ☺*

*Un immense merci à une personne inattendue, à ajouter parmi les gens qui ont compté, Jay. Un soutien de dernière ligne droite mais pas des moindres. Je suis chanceuse qu'on ait transformé tes « 1.2.3- 5.6.7 chic chic chac » en cette amitié aussi enrichissante et bénéfique. Merci pour ton écoute, ta sagesse, tes encouragements et pour avoir partagé ta vision si serene sur les événements de la vie. Je te serais goldenement ☺ reconnaissante pour tout le soutien que tu m'as apporté et je suis toujours aussi impressionnée que tu connaisses le lymphocyte B (j'espère qu'après lecture de ma thèse tu auras appris deux ou trois trucs quand même ☺). Merci pour tes blagues pourries, tes GJFs pandas et nos running remotivant. Mais plus sérieusement, Merci pour ces mercredi-oxygène, ces bouffées d'air frais, ton besoin de partager ta passion et de donner envie, et par-dessus tout ton immense bienveillance. Enfin merci de m'avoir porté dans les coups durs. A notre amitié, devenue si précieuse à mes yeux.*

*Petit clin d'œil aussi à mes belles rencontres de l'année, mes partenaires de danse : Sharon, Brigitte, Jérôme, Chris et Patrick, cette fin de thèse était plus simple aussi grâce à vous, merci d'avoir contribué à me vider la tête en me bougeant les fesses ☺.*

*Deboho, merci pour ton soutien même au moment où c'est toi qui en avait le plus besoin, j'admire ta force et ton courage. Cynthia, merci de comprendre toutes mes tares et de les accepter ☺. Merci à vous deux d'être une écoute attentive et de faire des blagues cailloux aux bons moments ☺ Vous savez déjà pas mal de choses alors, à nos Mojitos et aux milliards qui nous attendent avec une bonne occasion ou non ☺. Petit clin d'œil et merci également à PA, j'espère que tes questions sont prêtes, mes réponses sont déjà toutes faites ☺.*

*Et pour finir merci à la meilleure des familles, parce que tout seul on ne va jamais très loin. Merci à mes oncles, mes tantes et à mes cousins, les amis d'une vie malgré la distance (Steph, Alex, Marion et Maé, Louis, Anaïs et Pauline).*

*Je veux également remercier ma super Mamie pour ton immense bienveillance et de t'inquiéter pour nous tous parfois même beaucoup plus que nous même. Ça révèle tout l'amour que tu nous portes et je te le rendrai au centuple.*

*Merci à mon demi-frère, Jordan et à la plus déjantée des demi-sœur, Manon, parce qu'on se dit rarement tout l'amour qu'on se porte, mais les liens du cœur peuvent être aussi fort que les liens du sang et crois-moi tu remplis à merveille ton rôle de petite sœur chiante ☺. J'ai aussi une immense reconnaissance pour mon Beau père Didou/Redbull ! Parce que t'es un exemple de force mentale inébranlable et de réussite. Merci de faire autant pour nous, de nous supporter et de nous accepter. Enfin merci à la meilleure maman du monde d'avoir su me donner deux fois une famille merveilleuse, de m'avoir appris que le bonheur se construit et qu'il faut*





*parfois tout lâcher pour mieux recommencer, de croire en moi plus que personne et de donner autant pour les siens quitte à s'oublier parfois.*

*Merci aussi à mon merveilleux Papa qui ne doute jamais de moi. On est loin d'être toujours d'accord mais je sais que tu donnerais corps et âme pour ta famille et que tu me fais confiance bien plus que quiconque. Merci d'être ce pilier dans ma vie.*

*Merci à mes parents d'accepter et de me suivre dans toutes mes aventures 😊.*

*Merci à mon frère Teuteu préféré, le premier ami de ma vie, on est tellement et complètement différents, mais malgré tout ce qui nous oppose je suis fière de notre relation, merci d'être cette épaule dans ma vie, une épaule muette 😊 mais des fois la présence suffit. Merci aussi d'avoir été aussi sélectif lorsque tu as choisi la personne qui partage ta vie, plus qu'une belle sœur tu m'as fait gagner une véritable amie. Alors merci à toi Ma'hie, ma belle-sœur préférée, et pas que parce que tu compenses les silences de mon frère 😊, tu es un véritable soutien à part entière et une oreille attentive et pleine de compassion. T'es une des rares personnes à savoir réellement ce qui se passe dans mes journées de travail, et ça reflète toute ta générosité. Et bien sûr, infiniment merci pour ton talent orthographique et le temps que tu m'as consacré 😊.*

*Et il n'y a pas de plus belle conclusion que de finir par Mes nièces, Cléa & Naëlle.*

*Merci mes princesses de mettre autant de soleil dans ma vie.*

*Peu importe l'issue, il n'y a aucun regret à avoir lorsqu'on vit pleinement chaque instant. Je suis parfois dans l'excès, mais ça me permet de ne rien regretter, j'ai aimé tous ces moments, même ceux sous la pluie qui rappellent à quel point les petits rayons de soleil sont importants et je repars plus riche. Riche de connaissances, riche de ses rencontres mais par-dessus tout riche de toutes ces émotions qui font sentir vivant.*

*Une pensée aussi pour mon Indhi...*



## 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

---

Préambule .....	2
Partie I. Il était une fois l'ontogenèse B .....	5
I.1. Développement précoce : acquisition d'un récepteur B spécifique .....	5
I.1.1. De la cellule souche à l'engagement dans la lignée lymphocytaire B.....	5
I.1.2. La cellule pro-B : premiers remaniements géniques.....	6
A. La recombinaison VDJ.....	7
B. Contrôle du mécanisme de recombinaison VDJ.....	8
I.1.3. La cellule pré-B : naissance du BCR .....	10
I.1.4. La cellule B immature : acquisition de la tolérance du soi.....	12
I.2. Développement tardif : maturation du BCR, acquisition de fonction effectrice .....	12
I.2.1. Les cellules transitionnelles.....	13
I.2.2. Les cellules de la zone marginale .....	14
I.2.3. Cellules folliculaires : Coopération lymphocytaire.....	15
I.2.4. Centre germinatif : maturation finale des gènes d'Immunoglobulines .....	15
A. L'hypermutation somatique.....	17
B. La commutation de classe.....	21
I.2.5. Différenciation plasmocytaire : vers une usine à immunoglobulines.....	23
I.2.6. Les cellules B mémoires .....	25
I.2.7. Les cellules B1.....	26
Partie II. De l'Immunoglobuline aux gènes d'Ig.....	28
II.1. Les Immunoglobulines.....	28
II.1.1. Structure des immunoglobulines .....	28
A. Structure globale .....	28
B. Région Variable.....	29
C. Régions Constantes .....	29
II.1.2. Fonctions des Immunoglobulines.....	29
II.2. Origine des Immunoglobulines .....	30
II.2.1. Le locus $Ig\lambda$ .....	30
II.2.2. Le locus $Ig\kappa$ .....	31
II.2.3. Le locus des chaînes lourdes $IgH$ .....	33
Partie III. Les protéines de liaison aux régions MAR .....	37
III.1. Famille de protéine SATB (Special A-T rich Binding) .....	37
III.1.1. SATB1.....	37
A. Structure génomique et protéique de SATB1 .....	37
B. Régulation de l'expression et de l'activité de SATB1 .....	39
C. Implication de SATB1 dans le système immunitaire .....	42
D. Lien SATB1-pathologie / développement cancéreux .....	46
III.1.2. SATB2.....	48
A. Structure génomique et protéique de SATB2 .....	48
B. Régulation de l'expression et de l'activité de SATB2 .....	49
C. Implication de SATB2 dans l'organisme .....	50
D. Lien SATB2-pathologie / développement cancéreux .....	50



III.2. Famille de protéine A-T Rich Interaction Domain (ARID).....	51
III.2.1. ARID3A.....	52
A. Structure génomique et protéique de ARID3A .....	52
B. Mode d'action et régulation de ARID3A .....	52
C. Implication de ARID3A dans le système immunitaire .....	54
D. Lien ARID3A-pathologie / développement cancéreux .....	57
III.2.2. ARID3B.....	59
A. Structure génomique et protéique de ARID3B .....	59
B. Mode d'action et régulation de ARID3B.....	59
C. Implication de ARID3B dans l'organisme .....	60
D. Lien ARID3B-pathologie / développement cancéreux .....	61
III.3. CUX1 (Cut Like Homeobox 1) .....	61
Objectifs .....	66
Résultats .....	70
Discussion .....	82
Références bibliographiques .....	93
Annexes .....	126





## *Abbréviations*

---

<b>A</b>	<b>C</b>	<b>D</b>
A : adenine	C : cytosine	D: diversité
AA : acide aminé	C : constant	DICE : downstream immunoglobulin control element
ACF : ATP dependant chromatin assembly factor	CBP : CCAAT binding protein	DL : delta like
ADN : acide désoxyribonucléique	CBP : CREB binding protein	DN : double négatif
Ag : antigène	CC : coiled coil	DP : double positif
AID : activation induced cytidine deaminase	CD : cluster de différenciation	DSB : double strand break
APE : apurinique/apyrimidique endonucléase	CDP : CAAT displacement protein	
APOBEC : apolipoprotein B	CDR : complementary determining region	<b>E</b>
ARID3 : A-T rich interaction domain 3	CER : contracting element for recombination	eARID : extended ARID
ARN : acide ribonucléique	CG : centre germinatif	EBF : early B cell factor
ASE : anti-silencer element	ChIP : chromatin immunoprecipitation	EGR : early growth response
ASO : antisens oligonucleotide	CMH : complexe d'histocompatibilité	EHMT : euchromatic histone lysine methyltransferase
ATAC : assay for transposase-accessible chromatin	CPA : cellule présentatrice d'antigène	ELP : early lymphoid progenitor
	CSH : cellule souche hématopoïétique	EdU : ethynyl deoxyuridine
<b>B</b>	CSR : classe switch recombinaison	<b>F</b>
BCL : B cell lymphoma	CtBP : C –termina binding protein	Fab : fragment antigen binding
BCR : B cell receptor	CTCF : CTC- binding factor	FADD : Fas associates protein with death domain
BER : base excision repair	CTCL : cutaneous T cell lymphoma	Fas-L : Fas-ligand
BRG1 : brama related gene 1	CTGF : connective tissue growth factor	Fc : fragment cristallisable
BRIGHT : B cell regulator of IgH transcription	CtIP : Ct-interaction protein	FEN : flap endonuclease
BTK : Bruton's tyrosine kinase	CUX : cut-ike homeobox	FIGO : federation international de gynécologie et d'obstétrique
BUR : base unpairing region		FISH : Fluorescence <i>in situ</i> hybridization
		FO : folliculaire



FR : framework

## G

G : guanine

## H

H : Heavy

H2AX : histone 2A member X

HDAC : histone déacétylase

HES : hairy and enhancer of split

HMGB : high mobility group box

HS : hypersensibility

HV : hypervariable

## I

I : intervening

IEGs : immediate early gene

IFN : interferon

Ig : Immunoglobuline

IgH : locus des chaines Lourdes d'immunoglobuline

IgL : locus des chaines légères d'immunoglobuline

Ig $\kappa$  : locus de la chaine légère kappa

Ig $\lambda$  : locus de la chaine légère lambda

IL : interleukine

IRF : interferon regulatory factor

IRIS : inverted repeated intervening sequence

## J

J : jonction

JAK : janus kinase

JARID : jumonji-ARID

## K

Kb : kilobase

KD : knock down

kDa : kilo Dalton

KO : knock out

## L

L : leader

L : Light

LAD : Lamin Associated Domain

LAGC : lymphome anaplasique à grande cellule T

LB : Lymphocyte B

LED : lupus érythémateux disséminé

LLC : leucémie chronique lymphoïde

LPD : lymphoprolifération primitive cutané

LPS : lipopolysaccharide

LS : like switch

LSR : locus suicid recombination

LT : Lymphocyte T

LyP : papulose lymphomatoïde

## M

MAR : Matrix attachment region (région d'attachement à la matrice)

MAR BD : MAR binding domain

MCL : myeloid cell leukemia

MEF : myocyte enhancer factor

MLH : MutL homologie

MMR : mismatch repair

MSH : mutS homolog

MTA : metastasis associated protein

MZ : B de la zone marginale

## N

NES : nuclear export signal

NHEJ : non homologous end joining

NK : natural killer

NLS : nuclear localization sequence

NMD : nonsense mediated decay

NMTS : nuclear matrix targeting sequence

NP-CGG : nitrophenylacetyl-chicken gammaglobuline

NURD : nucleosomal remodeling deacetylase

## O

OGG : 8 oxo-guanine glycosylase



**P**

PAIR : pax5 activated intergenic repeat

PARP : poly- ADP ribose polymerase

PAX : paired box protein

PCAF : P300/CBP associated factor

PC-KLH : phosphatidylcholine-keyhole limpet hemocyanine

PCNA : proliferating cell nuclear antigen

PD1 : programmed death protein 1

PDZ-L : PDZ like

PH : pleckstrin homologie domain

PHD : plant homeodomain

PK : protéinase K

PKC : protéine kinase C

PLC : progéniteur lymphoïde commun

PMA : phorbol, myristate, acetate

PML : promyelocytic leukemia

PMP : progéniteur multipotent

PSM : postmeiotic segregation

**R**

RAG : recombination activating gene

**S**

S : switch

SAS : syndrome associé à SATB2

SATB : Special A-T rich binding

SCF : stem cell factor

SHM : somatic hypermutation

SIS : silencer in the intervening sequence

SPT5 : suppressor of ty5 homolog

SRSF : serine/arginine splicing factor

SSR : séquence signal de recombinaison

STAT : signal transducer and activator of transcription

SUMO : small ubiquitin related modifier

**T**

T : thymine

TADs : topologically associated domains

T1/2/3 : B transitionnel

TCL : T cell leukemia

TCR : T cell receptor

TdT : terminal desoxynucleotidyl transferase

TFH : T folliculaire helper

TGF : transforming growth factor

TI : thymo independent

TLR : toll like receptor

Treg : T régulateur

**U**

U : uracil

UNG : uracyl- DNA glycosylase

**V**

V : variable

VEGF : vascular endothelial growth factor

VIH : virus immunodéficience humaine

**W**

WT : wild type

**X**

XRCC : X-ray repair cross complementing

**Z**

ZC : zone claire

ZS : zone sombre



## Table des illustrations

---

Figure 1 : Structure d'une immunoglobuline

Figure 2 : Etapes du développement précoce du lymphocyte B dans la moelle osseuse.

Figure 3 : Représentation schématique de la recombinaison VDJ.

Figure 4 : Structure globale de la rate.

Figure 5 : Etapes du développement B tardif dans les organes lymphoïdes secondaires.

Figure 6 : Composition générale d'un centre germinatif.

Figure 7 : Hypermutation somatique (SHM).

Figure 8 : Prise en charge de l'uracile généré par AID lors du processus de SHM.

Figure 9 : Exemple du processus de la commutation de classe entre C $\mu$  et C $\epsilon$ (CSR).

Figure 10 : Voies de réparations initiées pendant le processus de CSR.

Figure 12 : Structure détaillée d'une immunoglobuline.

Figure 13 : Position des enhancers sur les loci des gènes des chaînes légères.

Figure 14 : Positions des enhancers sur IgH.

Figure 15 : Tableau récapitulatif des modèles murins de délétion de la région E $\mu$ MAR et les phénotypes associés ( $\nabla$  normal  $\searrow$  diminution).

Figure 16 : Représentation schématique des structures protéiques de SATB1 et SATB2.

Figure 17 : Représentation schématique de la structure protéique de ARID3A.

Figure 18 : Tableau récapitulatif des modèles murins de délétion de la protéine ARID3A et les conséquences sur le développement B.

Figure 19 : Représentation schématique de la structure protéique de CUX1 et de ses différentes isoformes

Figure 20 : Impact de la délétion de SATB1 sur les gènes AID OFF TARGET.

Figure 21 : Illustration de l'organisation nucléaire sous forme de TAD/LAD.

Figure 22 : Représentation schématique de la formation de boucle de chromatine par l'association des MAR BP avec les régions MAR et la matrice nucléaire.

Figure 23 : Analyse de pathway Gene ontology – cellular component à partir des données de RNA séquençage de cellules B spléniques mature et activées in vitro deux jours au LPS des souris SATB1 déficientes comparés à des souris WT.

Figure 24 : Test de l'efficacité de l'oligonucléotide antisens ciblant ARID3A.









# ***Préambule***



## Préambule

---

La capacité du corps humain à lutter contre l'agression d'un corps étranger repose sur le système immunitaire qui regroupe une diversité de cellules importante se répartissant dans l'ensemble de l'organisme et exerçant chacune des fonctions précises permettant la dégradation de l'agent pathogène. L'organisme bénéficie de barrières anatomiques et chimiques telles que la peau et les muqueuses, faisant office de premières lignes de défenses. Lorsque ces barrières sont franchies par un agent pathogène, les acteurs du système immunitaire prennent le relais pour limiter l'invasion et dégrader le pathogène.

Lors de l'invasion d'un corps étranger, le système immunitaire a recours à une réponse qui se divise en deux phases. La première phase fait appel aux acteurs du système immunitaire dit inné. Les mécanismes employés par le système inné, tels que l'induction de la lyse membranaire de bactérie ou la phagocytose d'agents infectieux, reposent sur des cellules exprimant des récepteurs et des molécules dont l'expression est héritée. Cette phase utilise des mécanismes généraux non spécifiques du pathogène qui peuvent s'avérer insuffisants. Lorsque c'est le cas, certains acteurs innés contribuent à l'activation de la seconde phase de réponse du système : la réponse adaptative.

Celle-ci apporte une action plus précise vis à vis de l'infection car elle est spécifique ou adaptée au pathogène rencontré. Elle se base sur deux types de cellules : les lymphocytes T, fournissant une réponse adaptative cellulaire et les lymphocytes B, producteurs d'anticorps (ou immunoglobulines), acteurs d'une réponse adaptative humorale.

Cette phase adaptative contribue également à établir une mémoire immunologique permettant une rapidité de réponse accrue lors d'une seconde infection par un même pathogène. C'est sur ces capacités de mémoire immunologique et de spécificité du pathogène que repose le principe de la vaccination. Toutefois, l'enjeu de la compréhension des mécanismes de ce système va bien au-delà du domaine de la vaccination. Compte-tenu du champs d'action étendu du système immunitaire, permettant à la fois une surveillance totale de l'organisme, et la distinction entre les molécules du soi et les molécules étrangères, l'appréhension de ce système est un indispensable à la compréhension des mécanismes d'échappement à sa réponse mis en place par certains pathogènes ou observés lors de développement cancéreux ou de maladie auto-immune.

Pour générer une réponse spécifique les lymphocytes ont recours à des événements de remaniement génique qui consistent à la modification des gènes (recombinaisons et mutations), permettant à la cellule d'exprimer un récepteur spécifique du pathogène. Dans le cas du lymphocyte B, ce récepteur est une immunoglobuline (Ig) membranaire qui portera la même spécificité que les Ig sécrétées. La cellule B est capable de produire une large diversité d'immunoglobulines, qui varient de par leur reconnaissance des pathogènes mais aussi de par leur fonction, à partir de seulement 3 gènes distincts.



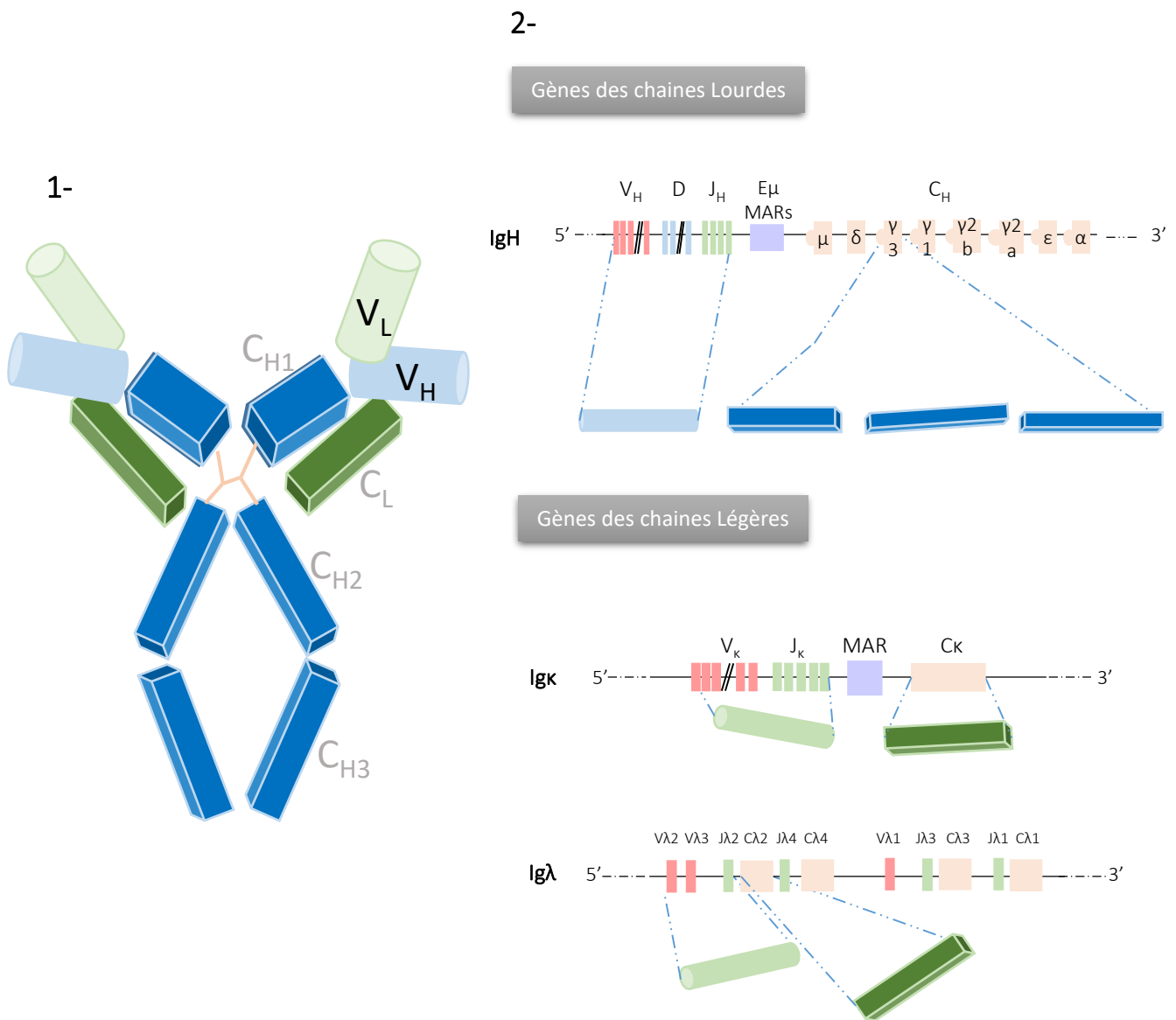
Cette caractéristique unique et propre à la cellule B la rend aussi sujette à certains évènements de remaniement indésirables pouvant conduire à l'émergence de pathologie. A l'état physiologique la cellule orchestre donc ces remaniements de façon très précise. Les gènes d'Ig, cibles de ces évènements, sont constitués de régions régulatrices permettant d'ordonner, limiter et contrôler ces modifications géniques. Un domaine génique composé d'un core enhancer, appelé E $\mu$ , encadré par deux régions d'attachement à la matrice nucléaire (MAR) contribue aux contrôles de ces évènements. Mes travaux de thèse consistent à étudier le rôle de facteurs de transcription possédant la capacité de se lier aux régions MAR. Diverses molécules, regroupées sous le nom de MAR binding protein (ou protéine de liaison aux régions MAR) ont la capacité de se lier sur ces régions et pourraient contribuer à différents processus spécifiques de la cellule B. A partir de deux modèles murins dépourvus de l'expression de SATB1 ou ARID3A, nous avons étudié l'implication de ces deux facteurs dans le développement des cellules B, la différenciation des plasmocytes et les remaniements géniques associés à la production d'immunoglobulines spécifiques.





# ***Synthèse Bibliographique***





**Figure 1 : Structure d'une immunoglobuline**

1- Structure d'une IgG composée de deux chaînes lourdes identiques (bleu) constituées d'un domaine variable  $V_H$  (clair) et de trois domaines constants  $C_H$  (foncé) ainsi que de deux chaînes légères identiques (vert) composées également d'un domaine variable  $V_L$  (clair) et d'un domaine constant  $C_L$  (foncé).

2- Représentation des loci des gènes des immunoglobulines (IgH, Igk et Igλ) repositionnant schématiquement les gènes V(D)J codant les régions variables et les gènes C codant les régions constantes ainsi que les régions MAR.

## **PARTIE I. IL ETAIT UNE FOIS L'ONTOGENESE B**

---

L'ontogénèse du lymphocyte B est basée sur une succession d'étapes toutes mises en place pour que la cellule mature ait acquis la capacité de produire des immunoglobulines matures, affines et fonctionnelles. Les Ig sont composées de quatre chaînes peptidiques (2 chaînes lourdes identiques et deux chaînes légères également identiques) toutes organisées en deux parties : une région variable et une région constante (**Figure 1**). Ces chaînes peptidiques sont issues de l'expression de deux gènes groupés en loci : IgH (Heavy chain immunoglobulin) pour la chaîne lourde et Igk ou Igl pour la chaîne légère. Ces gènes sont composés de nombreux segments géniques remaniés au cours du développement de la cellule afin de permettre la production d'une diversité d'immunoglobulines très conséquente. La structure globale des gènes d'Ig se compose d'une portion permettant l'expression de la région variable : il s'agit d'une multitude de segments codants nommés V (variable), D (diversité) et J (jonction) (**Figure 1**). C'est l'association entre un segment de chaque catégorie qui permet de coder la région variable. Cette association est la première modification génique subie par les gènes d'Ig, on parle alors de la recombinaison VDJ. La région variable sera ensuite maturée au cours du développement afin de devenir spécifique d'un antigène donné, il s'agit là du mécanisme d'hypermutation somatique (SHM).

La portion constante, quant à elle, confère la fonction effectrice des immunoglobulines et détermine l'isotype. L'isotype IgM est le premier synthétisé, ce dernier sera modifié au cours de l'activation des cellules B suite à une stimulation antigénique par un mécanisme appelé commutation de classe (CSR) sur le locus de la chaîne lourde. Ces trois mécanismes seront détaillés dans ce chapitre en retraçant le développement des cellules B.

### **I.1. Développement précoce : acquisition d'un récepteur B spécifique**

#### **I.1.1. De la cellule souche à l'engagement dans la lignée lymphocytaire B**

Les remaniements géniques nécessaires à la génération des immunoglobulines par le lymphocyte B imposent un développement et une maturation contrôlée et en diverses étapes. A l'origine de ces cellules, on trouve les cellules souches hématopoïétiques (CSH), cellules qui naissent au stade embryonnaire et qui résident tout au long de la vie de l'individu au sein de la moelle osseuse. Elles donnent naissance aux cellules sanguines et ont, comme toutes cellules souches, une capacité d'auto-renouvellement. Elles perdent cette capacité lorsqu'elles se différencient en cellules progénitrices multipotentes (PMP) qui seront à leurs tours capables de donner naissance aux progéniteurs



lymphoïdes communs (PLC) ou aux progéniteurs communs des lignées myéloïde et érythroïdique (Chao et al., 2008). Ce sont les signaux générés par le microenvironnement médullaire qui permettent aux cellules souches d'entrer dans une phase de différenciation. Pour exemple, l'engagement des progéniteurs multipotents dans la différenciation en cellules progénitrices lymphoïdes passe entre autre par la signalisation induite par le récepteur tyrosine-kinase FLT3 (Sitnicka et al., 2002). Les PMP expriment FLT3 qui sera activé par la fixation de son ligand porté par les cellules stromales de la moelle. Suite à l'activation de la signalisation de FLT3, certains récepteurs cytokiniques seront produits à leur tour, permettant la réception d'autres signaux et la poursuite de la cellule dans les différents stades de la différenciation lymphoïde. Les PLC sont capables de se différencier en trois types cellulaires différents : les lymphocytes B, les lymphocytes T et les cellules NK (Natural Killer).

La lymphopoïèse B se déroule en deux phases : la première phase constitue l'engagement des progéniteurs dans la lignée jusqu'à la formation d'une cellule B immature portant un récepteur fonctionnel et spécifique : le BCR (B cell receptor). Lors de cette phase, indépendante de l'antigène, la cellule restera au sein de la moelle osseuse en effectuant un déplacement progressif au fur et à mesure de sa différenciation afin de favoriser sa sortie de la moelle. La seconde étape, dépendante de l'antigène, s'effectuera après migration de la cellule immature au sein des organes lymphoïdes secondaires.

### **I.1.2. La cellule pro-B : premiers remaniements géniques**

La cellule pro-B est le premier stade cellulaire engagé dans la différenciation B, elle perd alors toute capacité de multipotence et son engagement dans la lignée B est définitif. Ce stade est permis par l'expression de certains facteurs dont l'absence engendre l'inhibition du développement normal de la cellule et un arrêt précoce dans la différenciation. Les premiers facteurs importants du développement B sont les facteurs Ikaros et PU.1, exprimés du stade de progéniteur jusqu'au stade de la cellule B immature et permettant la transition entre chaque stade de la différenciation médullaire (**Figure 2**) (Hardy et al., 2007; Monroe and Dorshkind, 2007; Yoshida et al., 2006). Le récepteur à tyrosine kinase, c-Kit, est également essentiel pour la progression de la cellule PLC au stade pro-B, en interagissant avec le facteur SCF (Stem Cell Factor) présent à la surface des cellules stromales (Hardy et al., 2007). Un autre facteur, indispensable à la différenciation des cellules progénitrices en cellules pro-B et dont l'expression est en partie médiée par Ikaros et PU.1 est le facteur de transcription E2A.

E2A induit la production de EBF (Early B-cell Factor) qui induit à son tour l'expression de différents acteurs, facteurs de transcription, marqueur de surface et certaines enzymes catalysant les premiers remaniements géniques de la cellule B (O'Riordan and Grosschedl, 1999; Ye and Graf, 2007). Parmi ces différents acteurs, EBF permet la production de PAX5 (Paired box protein 5) (Mikkola et al., 2002), un



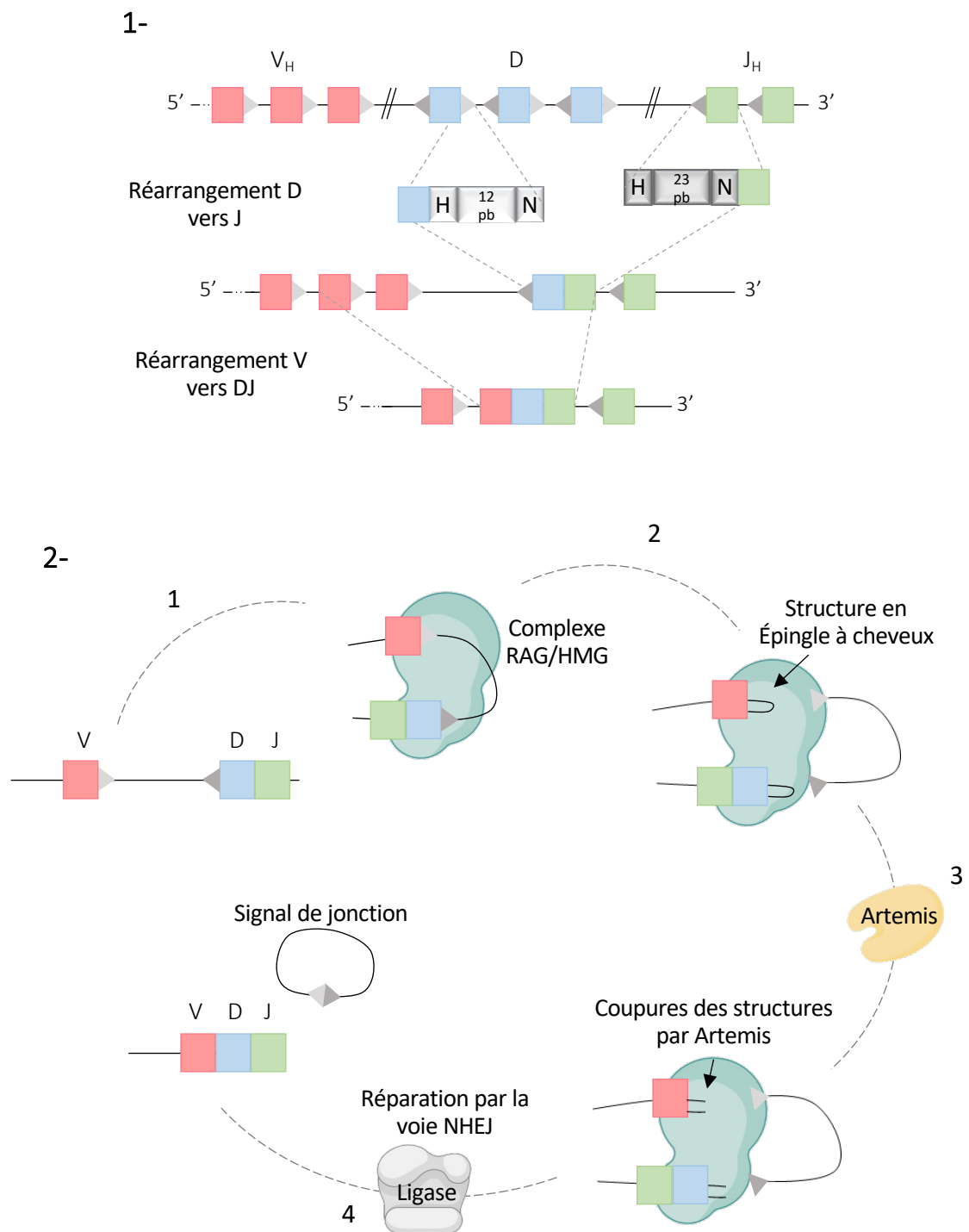
facteur de transcription indispensable à la progression de la cellule dans son développement médullaire (Maier and Hagman, 2002; Nutt et al., 1997). Il sera, par exemple, à l'origine de l'expression des protéines transmembranaires de CD19 et CD79a (aussi connu sous le nom de Ig $\alpha$  ou MB1), associés au BCR et assurant des fonctions respectives de co-récepteur et de signalisation (Cobaleda et al., 2007).

Au stade pro-B, la cellule initie les premiers évènements de recombinaison des gènes formant la région variable des Ig afin d'aboutir à la production d'un BCR spécifique et fonctionnel. On distingue alors un premier stade « pro-B précoce » au cours duquel s'effectue l'association d'un segment D à un segment J puis un stade « pro-B tardif » au cours duquel s'effectue un second réarrangement d'un segment V vers un segment D-J préalablement formé.

### **A. La recombinaison VDJ.**

Afin de guider ces réarrangements, les séquences d'ADN juxtaposées juste en aval des segments V, de part et d'autre des segments D et en amont des segments J, sont des séquences d'ADN à composition particulière, hautement conservées et non codantes, appelées séquence signal de recombinaison (SSR) (Early et al., 1980). Ces séquences se constituent d'un heptamère de nucléotide conservé 5' CACAGCT 3', séparé d'un nonamère 5' ACAAAAACC 3' par d'autres séquences appelées espaceurs. Les espaceurs sont des séquences non conservées de 12 ou 23 paires de bases. Ces derniers jouent un rôle important dans le mécanisme du réarrangement. Un segment V, D ou J associé à un espaceur de 12 nucléotides ne pourra être associé qu'avec une région ayant un espaceur de 23 nucléotides (Schatz and Ji, 2011). On parle alors de la règle 12-23 (**Figure 3**).

Le facteur EBF initie la production des protéines RAG1 et RAG2 (Recombination Activating Gene) qui s'associent sous forme d'hétérotétramère afin de reconnaître et rapprocher deux séquences SSR. Les deux protéines RAG1 reconnaissent chacune l'heptamère et le nonamère de la séquence SSR cible (Oettinger et al., 1990). Leurs dimérisations et leurs coopérations avec RAG2, HMGB1 et HMGB2 (High Mobility Group Box) permettent le rapprochement physique des deux SSR et la formation d'une synapse (Fugmann et al., 2000; van Gent, 1997; Shinkai, 1992). Les protéines RAG créent une cassure simple brin sur chacune des régions à associer, libérant ainsi une extrémité nucléophile qui génère à son tour une cassure sur le brin opposé et entraînant la formation d'une structure en épingle à cheveux (Schatz and Ji, 2011). L'ADN situé entre les deux séquences SSR et formant l'extrémité de la synapse est pris en charge par un complexe qui ligature les extrémités afin de former ce que l'on appelle la jonction des signaux. En ce qui concerne les extrémités des jonctions codantes (V, D ou J) à associer, elles sont prises en charge par les acteurs de la voie de réparation NHEJ (Non Homologous End Joining) spécialisés dans la réparation de cassure double brin. Les extrémités, toujours maintenues par la



**Figure 3 : Représentation schématique de la recombinaison VDJ.**

1- La Règle d'association 12/23 permet d'ordonner les réarrangements de D vers J puis de V vers le DJ préalablement réarrangé. Elle consiste en l'association d'une séquence SSR (triangle gris clair) composée d'un heptamère (H), d'un espaceur de 12 nucléotides et d'un nonamère (N) avec une séquence SSR dont l'heptamère et le nonamère sont séparés par un espaceur de 23 nucléotides (triangle gris foncé).

2- Représentation des étapes de la recombinaison d'un V vers le segment DJ préalablement réarrangé au locus IgH. Le mécanisme est initié par le complexe RAG/HMG qui crée des cassures dans l'ADN initiant l'intervention de la protéine ARTEMIS et de la voie de réparation NHEJ.

présence du complexe recombinaise, vont être liées par la protéine Ku (formé par l'association de Ku70 et Ku80) qui a pour rôle de recruter l'ADN-PK (protéine kinase dépendante de l'ADN), la protéine de réparation ATM ainsi que l'endonucléase ARTEMIS (Bredemeyer et al., 2006; Malu et al., 2012; Nussenzweig et al., 1996). L'association ADN-PK et ARTEMIS permet d'ouvrir la structure en épingle à cheveux (**Figure 3**) (Ma et al., 2002). Cette coupure, exercée de façon aléatoire, permet de libérer une extrémité simple brin, juxtaposant ainsi des nucléotides qui étaient préalablement complémentaires dans la séquence germinale. Ces derniers, annotés nucléotide P, forment de courtes séquences palindromiques et contribuent, par le caractère aléatoire du site de coupure, à la diversité de la séquence. Cette diversité est également augmentée par l'intervention d'une autre enzyme : la désoxynucléotidyl transférase terminale (TdT) qui ajoute des nucléotides aléatoires supplémentaires appelés N (Nontemplated) à chaque extrémité des brins à ligaturer (Desiderio et al., 1984). Le rôle de la TdT dans cette diversité jonctionnelle est donc important, cependant sa délétion n'éteint pas entièrement cette diversité. Un modèle murin, dépourvu de l'expression de la protéine, présente tout de même une diversité jonctionnelle résiduelle (Gilfillan et al., 1993).

Enfin, par l'intermédiaire de la protéine XRCC4, l'ADN-ligase IV est recrutée afin de lier les extrémités de chaque région codante (Li et al., 1995).

Le réarrangement de D vers J au locus IgH s'effectue sur les deux allèles puis la cellule orchestre par la suite, et de la même façon, un réarrangement d'un segment V vers DJ. Dans deux cas sur trois, les premiers réarrangements effectués sont hors du cadre de lecture et ne permettent pas la production d'une chaîne lourde. De nouveaux réarrangements sont alors initiés sur le second allèle.

Le mécanisme de recombinaison, nécessitant des étapes de cassures et ligatures de portions de séquence du génome, doit être impérativement contrôlé de façon stricte pour éviter tout évènements de recombinaison non désirable.

### ***B. Contrôle du mécanisme de recombinaison VDJ.***

La recombinaison VDJ est finement régulée pour assurer un contrôle temporel et spatial du mécanisme afin de limiter ces remaniements à un stade cellulaire (pro-B) et à l'échelle d'une portion de chromosome (locus IgH). Cette régulation s'effectue dans un premier temps par le contrôle de l'expression des protéines effectrices du mécanisme. Si les acteurs de la voie de réparation de l'ADN NEHJ sont des protéines ubiquitaires, l'expression des protéines RAG ainsi que de la TdT sont, quant à elles, restreintes aux stades précoces des cellules T (Bassing et al., 2000; Grawunder et al., 1995) (qui procèdent également aux réarrangements géniques de son récepteur, T cell receptor) et des cellules B, limitant ainsi l'apparition de cassures de brin d'ADN au sein d'autres types cellulaires.





Un autre contrôle est préétabli par la composition des séquences SSR et la règle 12/23. Ces séquences participent aux guidages des enzymes RAG sur les segments V, D et J, s'assurant ainsi qu'un segment V du locus IgH ne sera pas directement recombinaison avec un segment J (supprimant D) dont la SSR sera composé d'un espaceur de taille égale à celui porté par le segment V (Alt et al., 1984).

Des modèles de transfection de fibroblaste ont montré que l'expression seule des protéines RAG1 et RAG2 ne suffit pas à induire la recombinaison des gènes d'Ig endogènes. En revanche, ces modèles ont la capacité de réarranger des segments de gènes portés par un plasmide (Oettinger et al., 1990). Ces expériences mettent en avant la nécessité de rendre les segments à recombinaison et leurs SSR accessibles aux différents acteurs du mécanisme et le contrôle de cette accessibilité est donc un point essentiel à la régulation des recombinaisons VDJ. Dans un premier temps, il s'établit des modifications de conformation de la chromatine par modulation des marques épigénétiques des histones (protéines de compaction de l'ADN) (Chakraborty et al., 2007; Sen and Oltz, 2006). La méthylation semble jouer un rôle important notamment pour le recrutement de RAG2. Le domaine PHD (Plant HomeoDomain) de cette protéine reconnaît entre autre les marques de méthylation des histones H3, notamment lorsque celle-ci est tri-méthylée sur la lysine 4 (Liu et al., 2007; Matthews et al., 2007). Des mutations au sein de ce domaine entraînent une altération du mécanisme de recombinaison. C'est notamment le cas chez les patients atteints du syndrome d'Omenn, causé en partie par l'absence de reconnaissance de cette triméthylation. Ceci entraîne une perte de la capacité de recombinaison des cellules B et un blocage de leur maturation illustré par l'absence de cellules B circulantes chez les patients (Villa et al., 1998).

L'hyperacétylation des histones contribue aussi à l'accessibilité des gènes en favorisant la transcription. Aux loci des gènes d'Ig, les marques d'acétylation de l'histone H3 lysine 9 et H3 lysine 14 sont fortement enrichies sur les régions transcriptionnellement actives et favorisent l'accessibilité des séquences SSR aux protéines RAG (Johnson et al., 2003; McMurry and Krangel, 2000). Cette distribution de l'acétylation est hautement corrélée à la présence de la protéine BRG1 (Brama related gene 1) (Morshead et al., 2003). Celle-ci participe à la formation du complexe de remodelage de la chromatine SWI/SNF qui permet d'initier la transcription germinale des segments à recombinaison. Cette transcription, elle aussi, optimisera l'accessibilité des gènes. Les premiers événements de transcription au locus IgH s'établissent au promoteur pdQ52, situé en amont du dernier segment D, et entraînent la production d'un transcrit sens non codant nommé  $\mu 0$  (Alessandrini and Desiderio, 1991). La production de ce transcrit est accompagnée par celle de deux autres transcrits :  $\mu$  initiée en aval de l'enhancer  $E_{\mu}$  d'IgH (situé entre les segments J et les gènes constants) ainsi qu'un transcrit anti-sens initié en amont de cet activateur et favorisant la recombinaison de D vers J (Perlot and Alt, 2008). Lorsque la recombinaison D-J est établie, la transcription se met en place au promoteur du V sélectionné pour être réarrangé, menant à la production de transcrits non codants sens et anti-sens (Bolland et al.,



2004). Les modifications des marques épigénétiques contribuant à l'activation de la transcription s'établissent de façon ordonnée et séquentielle (sur DJ puis sur V) et participent au contrôle de l'accessibilité des segments.

Il est indispensable pour la cellule que les réarrangements VDJ soient fonctionnels et permettent la production d'une chaîne lourde d'immunoglobuline. Celle-ci est essentielle à la réception d'un signal de survie qui permettra à la cellule de poursuivre sa maturation (détaillé au paragraphe suivant). Si l'ensemble des essais de réarrangement reste non productif, la cellule sera éliminée par apoptose.

### **I.1.3. La cellule pré-B : naissance du BCR**

Le stade pré-B consiste pour la cellule à acquérir un signal de survie via l'expression à sa surface d'un récepteur qui inclut sa chaîne lourde nouvellement réarrangée. Les facteurs E2A et EBF, cités précédemment, permettent à ce stade du développement l'expression de protéines substitutives à la chaîne légère. Cette dernière n'étant pas encore réarrangée, l'expression d'un BCR complet n'est pas réalisable. La cellule exprime alors une pseudo chaîne légère constituée de protéines appelées  $\lambda 5$  et VpréB qui présentent respectivement des caractéristiques communes avec la portion constante de la chaîne  $\lambda$  et du domaine variable des chaînes légères (Mårtensson and Ceredig, 2000; Melchers et al., 1993). Bien que la pseudo chaîne légère n'inclut pas de vraie région variable, son association à la chaîne lourde permet à la cellule d'exprimer un pré-BCR qui présente tout de même certaines spécificités de reconnaissances portées par la chaîne lourde seule. Il est également probable que l'assemblage de cette pseudo chaîne légère à la chaîne lourde donne une conformité au récepteur qui favorise une interaction entre plusieurs pré-BCR (Bankovich et al., 2007). Il est aussi établi que le pré-BCR est capable d'interagir avec un ligand des cellules stromales, la Galectin-1 qui contribue également à la signalisation (Gauthier et al., 2002). Ces interactions (inter pré-BCR et Galectin-1) participent à la transmission du signal de survie qui est relégué en intracellulaire par l'intermédiaire de  $Ig\alpha$  (CD79a) et  $Ig\beta$  (CD79b), molécules de co-stimulation. Cette signalisation fait aussi appel à plusieurs facteurs incluant la Btk (Bruton tyrosine kinase), dont l'expression est indispensable à la poursuite du développement B ; dans un premier temps via son rôle dans cette signalisation puis nous verrons par la suite qu'elle participe également à la transcription de la chaîne lourde au sein des cellules plus matures (de Weers et al., 1993). Sa déficience induit un blocage total du développement B dès le stade pré-B chez les patients atteints de l'agammaglobulinémie de Bruton (Lederman and Winkelstein, 1985). Le signal perçu suite à l'engagement du pré-BCR entraîne une inhibition de l'expression de la pseudo chaîne légère ainsi qu'une internalisation du pré-BCR expliquant ainsi son expression transitoire (limitée au stade pré-B), le temps d'acquérir le signal de survie qui induit la prolifération des cellules. On parle alors de grande cellule pré-B (Burrows et al., 2002). Après cette phase de



prolifération, qui aboutit à une modeste expansion clonale de cellules exprimant une chaîne lourde identique, la cellule pré-B devient une petite cellule pré-B dont le rôle est d'orchestrer le réarrangement d'une chaîne légère fonctionnelle afin d'exprimer un BCR complet (**Figure 2**) (Mårtensson et al., 2002). La recombinaison des chaînes légères utilise le même mécanisme que la chaîne lourde, néanmoins, la chaîne légère ne possédant pas de segment D, le V sélectionné est directement recombinaison avec un segment J (Weigert et al., 1980). Dans le cas de la chaîne  $\kappa$ , une portion non négligeable des segments V est orientée en sens inverse. Cette configuration induit des recombinaisons conduisant à l'inversion de la portion d'ADN située entre les deux segments et à l'inclusion du signal de jonction (normalement éliminée dans une recombinaison classique) en amont du VJ réarrangé (Aoki-Ota et al., 2012).

On peut aussi noter que la diversité générée par la TdT est moindre sur les chaînes légères que sur la chaîne lourde, l'enzyme étant moins exprimée à ce stade du développement cellulaire; c'est particulièrement vrai chez la souris où la présence de nucléotides additionnels est très rare sur une chaîne légère. La diversité du BCR nouvellement acquis est aussi apportée par la possibilité de recombinaison l'une ou l'autre des deux chaînes légères :  $\kappa$  et  $\lambda$ . Leur recombinaison est également orchestrée de façon ordonnée : dans un premier temps le réarrangement aura lieu sur un allèle  $Ig\kappa$  puis sur le second si le premier n'est pas fonctionnel, et enfin sur  $Ig\lambda$  (Arakawa et al., 1996). Ainsi, si les réarrangements d' $IgH$  conduisent à une perte d'un bon nombre de cellules pro-B incapables de produire une chaîne lourde, l'habileté de produire deux chaînes légères à partir de deux gènes distincts permet à la majorité des cellules pré-B de poursuivre leur développement.

La séquence ordonnée des réarrangements sur les deux loci  $\kappa$  et  $\lambda$  permet un phénomène d'exclusion isotypique qui consiste, dans le cas de la production d'une chaîne  $\kappa$  fonctionnelle, à l'arrêt du mécanisme de recombinaison pour bloquer un éventuel réarrangement au locus  $Ig\lambda$  (Arakawa et al., 1996).

Si l'on revient à la cellule pro-B et son réarrangement VDJ, le même phénomène est mis en place ; on parle cependant ici d'exclusion allélique. Le mécanisme de recombinaison est inhibé pour bloquer de potentiels réarrangements sur le second allèle, lorsque le premier est fonctionnel, afin de permettre à la cellule de ne présenter qu'un seul type de chaîne lourde. Un modèle murin dépourvu de l'expression de la chaîne  $\lambda 5$  montre que celle-ci contribuerait de manière importante au phénomène d'exclusion allélique d' $IgH$  (Löffert et al., 1996). Cette exclusion est notamment exercée par un arrêt transitoire de l'expression des protéines RAG ainsi que par la dégradation de RAG2, initiée par sa phosphorylation lors de la phase de prolifération, au stade de grande cellule pré-B (Geier and Schlessel, 2006). L'expression des protéines sera ensuite réactivée pour permettre le réarrangement de la chaîne légère. Chez l'Homme et la souris, les réarrangements sur  $Ig\kappa$  sont majoritairement fonctionnels, les deux



espèces présentent alors respectivement 65 et 95 % de cellules B matures exprimant un BCR à chaîne K.

L'association de la chaîne légère réarrangée à la chaîne lourde permet de produire une immunoglobuline de type M (IgM) exprimée à la surface de la cellule qui, associée aux molécules de co-signalisation (Ig $\alpha$  et Ig $\beta$ ), formera le récepteur spécifique du lymphocyte B qu'est le BCR. La cellule atteint alors le statut de cellule B immature.

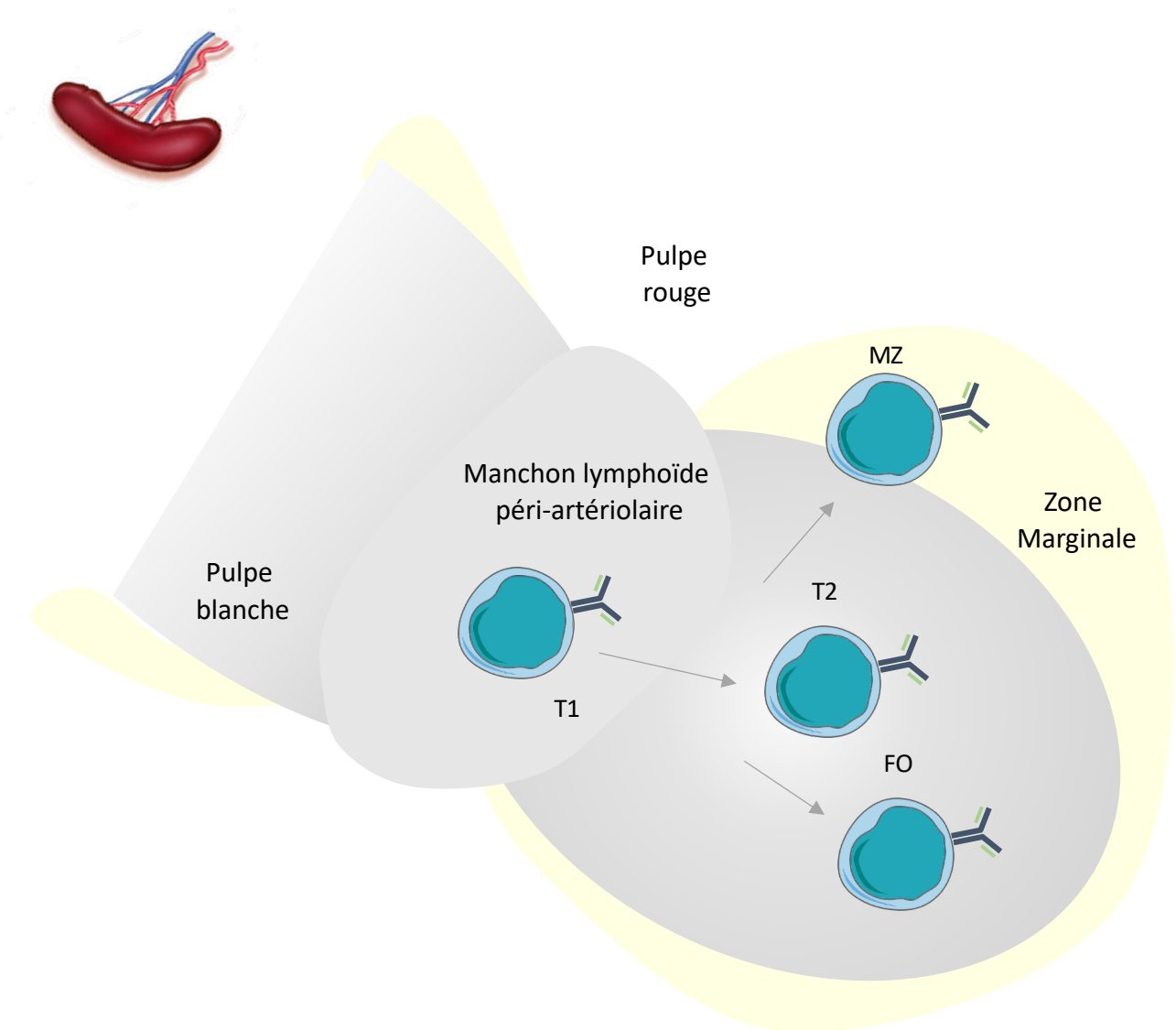
#### **I.1.4. La cellule B immature : acquisition de la tolérance du soi**

Suite à l'expression de son BCR chaque cellule B, dite immature, doit soumettre son récepteur à un contrôle de tolérance du soi avant d'effectuer sa sortie de la moelle osseuse. Cette étape permet de s'assurer que les IgM produites par chaque lymphocyte n'auront pas la capacité de se lier aux antigènes de l'organisme et d'initier une réponse auto-immune. Si les réarrangements conduisent à l'obtention d'une Ig auto-réactive, le lymphocyte peut alors effectuer une révision de son BCR (Melamed et al., 1998; Radic and Zouali, 1996). Ce mécanisme fait intervenir de nouveaux réarrangements sur l'allèle fonctionnel (recombinaisons entre segments encadrant le VJ préalablement réarrangé) et offre la possibilité à la cellule d'exprimer une nouvelle chaîne légère qui sera associée avec la chaîne lourde initiale (Chen et al., 1997). Le nouveau BCR sera de nouveau testé, les récepteurs autoréactifs conduiront la cellule à une entrée en phase apoptotique, alors que les BCR non réactifs fourniront aux cellules un signal permettant la poursuite de leur développement. Ce signal favorise alors la sortie des lymphocytes B en dehors de la moelle osseuse et ils migreront, via le réseau sanguin, jusqu'aux organes lymphoïdes secondaires. Parmi ces derniers, on retrouve principalement la rate, les ganglions lymphatiques et les tissus lymphoïdes associés aux muqueuses. Au sein de la rate, les lymphocytes vont subir différentes phases de maturation qui rendront leur BCR plus spécifique et leur permettront d'atteindre le stade final de leur développement.

#### **I.2. Développement tardif : maturation du BCR, acquisition de fonction effectrice**

Les organes lymphoïdes secondaires font office de point de rencontre entre l'antigène et les cellules lymphoïdes. C'est dans ces organes que les cellules B finissent leur développement, notamment via la maturation finale de leurs gènes d'Ig médiée par la coopération avec d'autres types cellulaires, afin de devenir un plasmocyte sécréteur. Les organes secondaires sont plus ou moins organisés de façon similaire avec des structures composées de différentes cellules.





**Figure 4 : Structure globale de la rate.**

Représentation schématique des différentes zones de localisation des cellules B au sein de la rate. T : transitionnel, MZ : B de la zone marginale, FO : cellule B folliculaire.

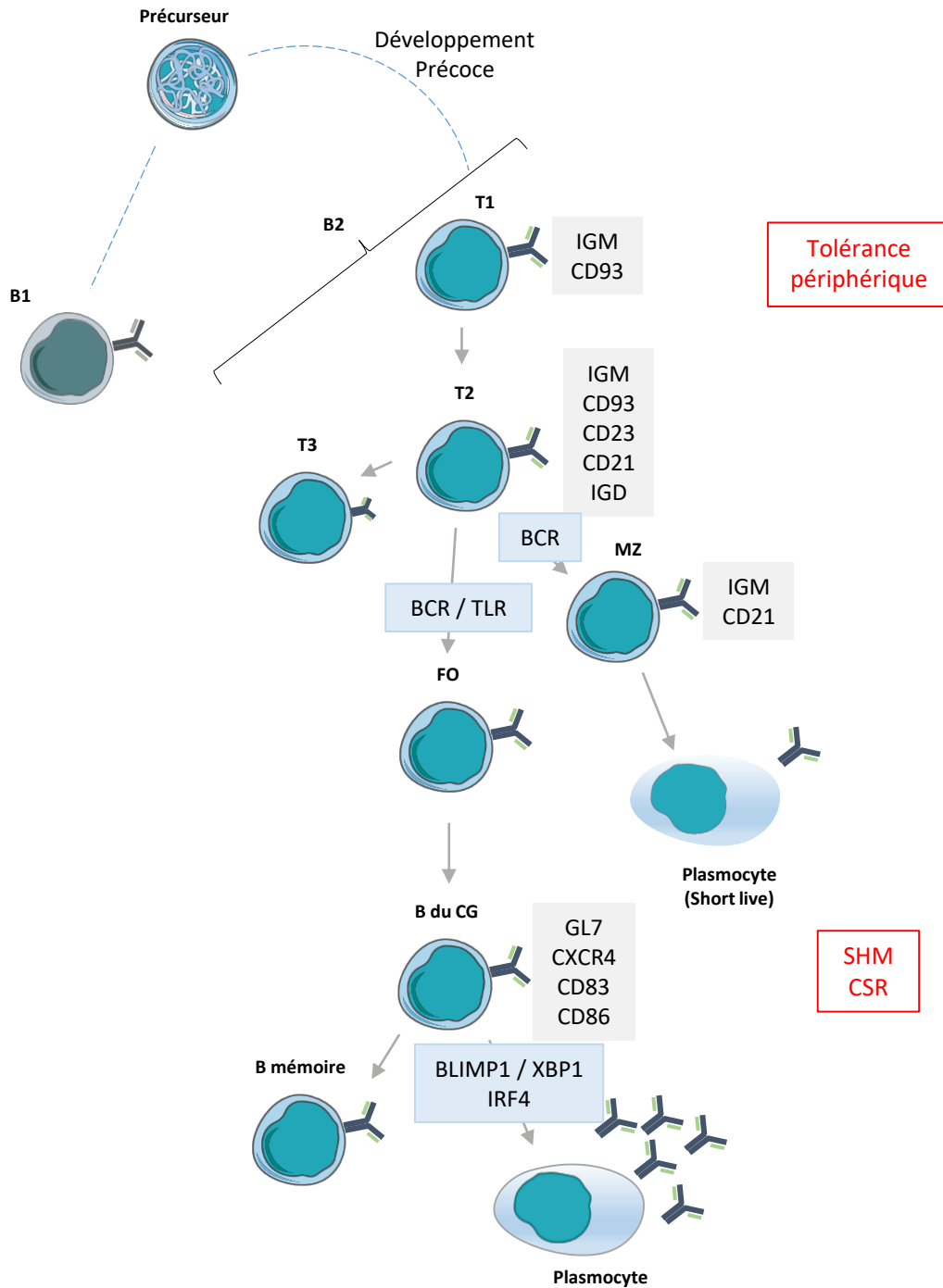
La rate est l'organe lymphoïde secondaire principal qui permet la capture d'antigènes issus de la circulation sanguine. Elle se compose de deux tissus : la pulpe rouge qui est constituée de sinus veineux et de capillaires et la pulpe blanche qui contient les tissus lymphoïdes. Ces derniers sont finement organisés en plusieurs zones. Une première zone, occupant le centre de la pulpe blanche, est définie comme une zone à cellules T. Aux extrémités de la pulpe, se forment des zones de cellules B, appelées follicules lymphoïdes primaires, entre lesquels s'établissent les manchons lymphoïdes péri-artériolaires. Enfin, l'ensemble de la pulpe blanche est délimitée par la zone marginale (**Figure 4**). Chacune de ces zones regroupent des populations de cellules lymphoïdes, dendritiques et stromales qui participent à la maturation et à l'activation des lymphocytes B par échanges de signaux chimiokines et d'interactions cellule-cellule (Mebius and Kraal, 2005).

### **1.2.1. Les cellules transitionnelles**

Les cellules immatures colonisent la rate dans le manchon lymphoïde péri-artériolaire sous forme de lymphocyte B transitionnel de niveau 1 (T1) et sont de nouveau soumises à une phase de sélection. Celle-ci, appelée tolérance périphérique, permet d'éliminer les cellules portant un BCR auto-réactif à un antigène du soi qui ne serait pas présenté au sein de la moelle. Les cellules dont l'activation du BCR est importante durant cette phase de sélection seront éliminées (Goodnow et al., 1989; Petro et al., 2002). En revanche les cellules dont l'activation du récepteur est faible, recevront un signal de survie qui sera aussi, en partie, fourni par des récepteurs membranaires de la cellule B liant la protéine BAFF (B Cell Activating Factor) exprimée par les cellules dendritiques folliculaires. A ce stade, la réception des signaux émis par les récepteurs de BAFF (BAFF-R, BCMA et TACI) et la faible activation du BCR sont indispensables à la poursuite de la maturation du lymphocyte (Rolink et al., 2002; Stadanlick and Cancro, 2008). Une modification de l'expression de ces récepteurs et plus particulièrement du BAFF-R conduit à une perturbation de la sélection cellulaire avec un phénotype de maladie auto-immune illustrée notamment dans le cas du syndrome de Sjögren (Thompson et al., 2016). Ce syndrome est caractérisé par la production d'autoanticorps dirigés contre les glandes sécrétrices de l'organisme. Les patients présentent une infiltration de lymphocyte au sein des glandes (lacrymales et salivaires) ainsi qu'une sécheresse buccale et oculaire importante.

Les cellules T1 caractérisées par une expression importante d'IgM et de CD93 (AA4.1), sélectionnées par la tolérance périphérique, migrent alors vers les follicules et deviennent des lymphocytes transitionnels T2 (**Figure 4 et 5**).

Les T2 acquièrent l'expression de nouveaux facteurs dont CD23 et CD21, indispensable à leur maturation. Ce type cellulaire est également caractérisé par une expression importante de l'isotype IgD à sa surface, résultant d'un épissage alternatif qui retire le gène codant la portion constante d'une



**Figure 5 : Etapes du développement B tardif dans les organes lymphoïdes secondaires.**

Cette représentation indique les différents stades cellulaires du développement tardif du lymphocyte B avec les marqueurs associés (noir), les facteurs indispensables à la progression de la cellule dans le développement (cadre bleuté) et les différents mécanismes de sélection et de maturation de la cellule (rouge). T : transitionnel, MZ : B de la zone marginale, FO : cellule B folliculaire, CG : centre germinatif, BCR : récepteur spécifique de la cellule B, TLR : Toll like receptor.

immunoglobuline de type M. Le rôle de l'IgD n'est pas clairement défini : dans certaines études, il est associé à la mise en place de signaux apoptotiques, dans d'autres, au contraire, il est décrit comme étant un amplificateur de la réponse au BCR, favorisant la survie et la prolifération cellulaire (Chen and Cerutti, 2011). En effet, les cellules T2 obtiennent une capacité de prolifération fournie par l'activation de leur BCR, qui, à ce stade, contribuera à la progression des cellules dans leur développement mais également à la détermination de leur devenir, entraînant une différenciation en cellule folliculaire B (FO) ou en cellule B de la zone marginale (MZ) (**Figure 5**) (Niir and Clark, 2002). D'autres récepteurs, les TLR (Toll-Like Receptor) participent également à cette prise de décision du devenir de la cellule. Ces derniers, présents sur divers types cellulaires, permettent la reconnaissance de molécules bactériennes, fongiques ou virales. Ils sont hautement conservés et contribuent aussi bien à l'immunité innée qu'à l'immunité adaptative. Les plus connus, exprimés par la cellule B, sont les TLR4 et 9 reconnaissant respectivement le lipopolysaccharide (composant de membrane bactérienne) et l'ADN portant des CpG non méthylés, retrouvé chez les bactéries et les virus (Lund et al., 2003; Park et al., 2009). Leur activation entraîne la mise en place d'une signalisation intracellulaire via le recrutement de molécules adaptatrices telles que MYD88, MAL, TRIF et TRAM (Kawai et al., 2004; Takeda and Akira, 2004). La signalisation d'un TLR associée à celle du BCR contribue à induire une différenciation de la cellule T2 en cellule Folliculaire. En revanche, une signalisation médiée seulement par le BCR favorisera une différenciation en cellule MZ (B de la zone marginale) (Pillai and Cariappa, 2009). Une troisième population de cellules transitionnelles, T3, est également répertoriée comme étant une population de cellules B légèrement auto-réactive, rendues anergiques et exprimant très faiblement une IgM de surface (Liubchenko et al., 2012; Merrell et al., 2006; Teague et al., 2007).

### 1.2.2. Les cellules de la zone marginale

La zone marginale, décrite précédemment, comme une zone de jonction entre la pulpe rouge et la pulpe blanche, fournit une réserve de cellule B ayant la capacité de répondre rapidement à la présence d'un pathogène mais d'une façon moins spécifique. Cette population exprime de façon importante une IgM de surface, au détriment de l'IgD, et est caractérisée par une forte expression de CD21 ainsi que par la perte de l'expression de CD23 (**Figure 5**) (Cerutti et al., 2013; Pillai et al., 2005). Ce type cellulaire se différencie des B folliculaires par sa capacité à initier une première réponse B de façon indépendante à une coopération avec d'autres types cellulaires et ce notamment par leur capacité de différenciation rapide en plasmocyte à courte durée de vie qui leur permet de sécréter rapidement des IgM (Balázs et al., 2002; Martin et al., 2001). Même si leur rôle est initialement décrit comme étant une première défense anti-pathogène il semblerait que ces cellules soient également aptes à répondre à une stimulation via un lymphocyte T (comme il sera décrit ultérieurement) et qu'elles participeraient



également au transport des antigènes au sein des follicules (Cinamon et al., 2008; Song and Cerny, 2003).

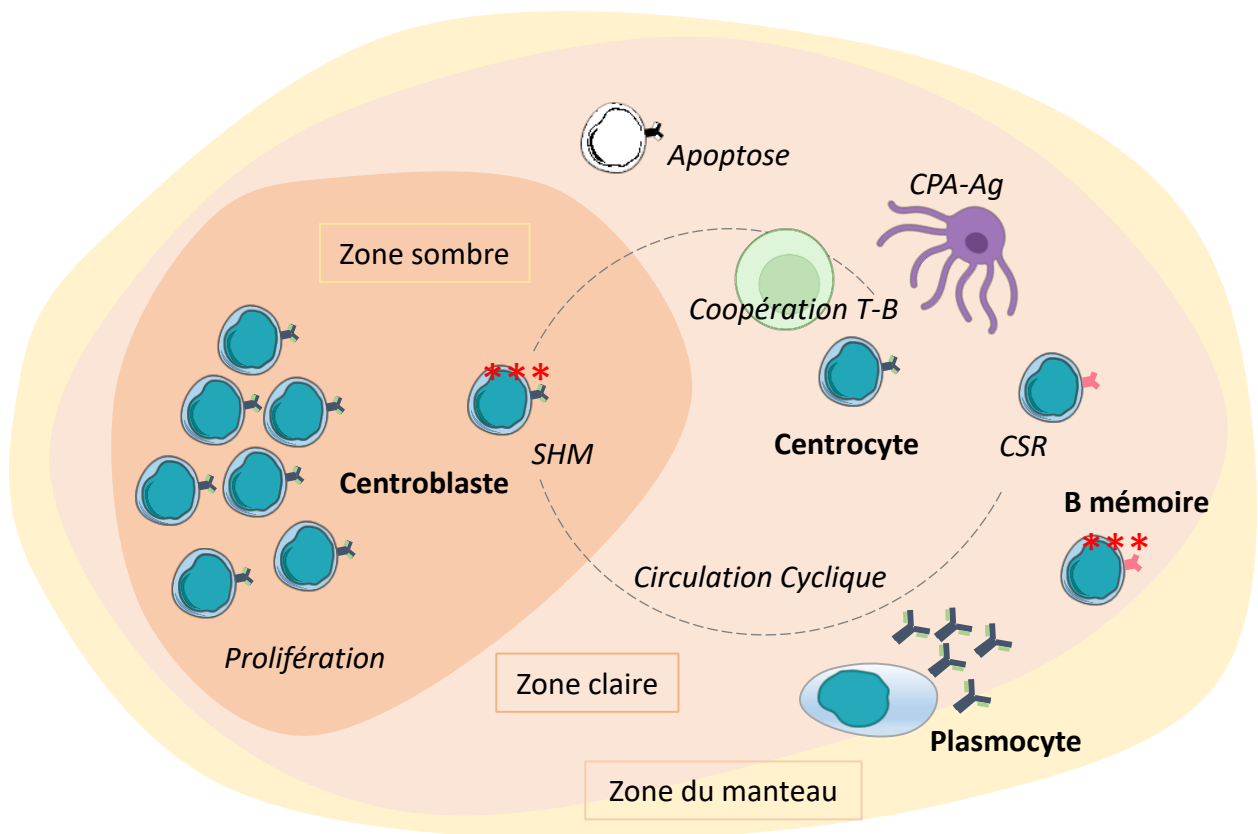
### **I.2.3. Cellules folliculaires : Coopération lymphocytaire**

La pulpe blanche est le lieu de formation des follicules qui abriteront des foyers de maturation des lymphocytes B appelés centres germinatifs (Mebius and Kraal, 2005). La répartition des cellules B et la formation des différentes zones au sein de la pulpe blanche est gérée par la rencontre avec un Ag mais aussi et toujours via des interactions chimiokines qui permettent la migration des cellules. Un lymphocyte B nouvellement entré au sein de l'organe, dit B naïf, exprime le facteur CXCR5 qui lui permet d'intégrer un follicule primaire guidé par les cellules dendritiques folliculaires et les cellules stromales qui sécrètent en quantité importante CXCL13, le ligand de CXCR5 (Okada and Cyster, 2006). Après sa rencontre avec un antigène, de façon directe ou médiée par une cellule présentatrice d'antigène (CPA), la cellule B initie son activation et exprime à sa surface différentes molécules qui lui permettront d'interagir avec un lymphocyte T situé dans la zone à cellule T. Il s'agit là d'un échange entre types lymphocytaires, le B s'engage à présenter l'Ag au T contribuant ainsi à son activation. Le T, en contrepartie, fournit à la cellule B différents signaux qui lui permettent de survivre et de finaliser, elle aussi, son activation. Cet échange a lieu par des signaux de co-stimulation qui s'effectue par l'interaction de récepteurs de surface ou de molécules sécrétées. Le signal majeur est médié par la présentation de l'Ag par la cellule B, via des molécules du complexe d'histocompatibilité (CMH), aux lymphocytes T qui réceptionnent le signal par leur TCR (T cell receptor). Les cellules B expriment aussi la protéine ICOS-L, le ligand du marqueur ICOS, présent à la surface des T. Cette interaction entraîne une signalisation au sein de la cellule T indispensable à sa différenciation en T Folliculaires Helper ( $T_{FH}$ ) (Yoshinaga et al., 1999). Ainsi les cellules T exposent le facteur CD40L à leur surface qui lie le récepteur CD40 présent sur la membrane des lymphocytes B. La réception de ce signal par la cellule B permet de générer d'autres signaux, notamment anti-apoptotiques et favorise sa survie (Mitchison, 2004). Le lymphocyte T sécrète également des cytokines telles que l'IL21, le TGF- $\beta$ , l'IFN- $\gamma$  et l'IL4 qui seront réceptionnées par le lymphocyte B afin de guider sa maturation.

La cellule B activée pourra alors entrer dans le centre germinatif afin de finaliser la maturation de ces gènes d'Ig.

### **I.2.4. Centre germinatif : maturation finale des gènes d'Immunoglobulines**

Les centres germinatifs (CG) sont formés au sein d'un follicule primaire qui deviendra alors un follicule secondaire. L'activation des cellules B par la rencontre avec l'antigène et la coopération avec les



**Figure 6 : Composition générale d'un centre germinatif.**

Cette figure positionne les différentes zones constituant un centre germinatif (zone sombre / zone claire) ; sont également répartis les divers types cellulaires ainsi que les mécanismes subis par chaque type (prolifération / CSR /SHM)

cellules T induisent leur prolifération et la formation de ces centres germinatifs (Allen et al., 2007). Les cellules B non activées vont être repoussées vers l'extérieur, formant ainsi la zone du manteau du CG. Ce centre est formé de deux autres zones, la zone sombre (ZS) et la zone claire (ZC), distinguables de par leur composition cellulaire. La ZS est une zone à forte densité cellulaire qui abrite des cellules B appelées centroblastes, proliférant de façon intensive et caractérisées par les marqueurs  $GL7^+ CXCR4^{hi} CD86^{low} CD83^{low}$  ainsi qu'une réduction importante de l'expression des IgM et IgD à leur surface. La ZS fait opposition à la ZC dans laquelle la densité cellulaire est moindre et la prolifération des B est diminuée. Les lymphocytes B de cette zone, appelés centrocytes, sont caractérisés par  $GL7^+ CXCR4^{low} CD86^{hi} CD83^{hi}$  et une forte expression des Ig de surface (Victoria et al., 2010). **(Figure 6)**

Bien qu'un centre germinatif soit divisé en deux, les deux zones ne sont pas exclusives et il s'y établit une dynamique de circulation cyclique des cellules B. Les centroblastes sont maintenus dans la zone sombre par les cellules stromales produisant la chimiokine CXCL12 qui se lie à la surface des B au récepteur CXCR4. Au cours du temps, le centroblaste réduit sa prolifération, éteint l'expression de CXCR4 et exprime uniquement CXCR5. Cette modification d'expression de récepteur chimiokinique favorise sa migration vers la zone claire, où CXCR5 liera de nouveau la chimiokine CXCL13 ; on parle alors de centrocyte (Allen et al., 2004).

Différents facteurs, initiés par les signaux extérieurs provenant des cellules environnantes, participent à la formation et le maintien du centre germinatif.

Le facteur de transcription MYC est un acteur indispensable à la formation du CG, sa délétion dans les cellules B activées induit une incapacité à former des centres germinatifs (Calado et al., 2012). Son rôle, souvent associé à un aspect prolifératif, n'est pas entièrement connu au sein des CG, d'autant plus que son profil d'expression dans les cellules B activées du centre pose question quant à son implication dans la fonction de prolifération de ces cellules. MYC est en effet exprimé de façon précoce lors de la formation d'un centre germinatif, il est éteint dans les centroblastes (très proliférant) puis de nouveau exprimé dans les centrocytes (moins proliférant) regagnant la ZS (Dominguez-Sola et al., 2012).

Un autre facteur de transcription soutenant l'aspect hautement proliférant des centroblastes est E2A : ce dernier favorise la progression des cellules dans le cycle cellulaire, notamment via l'expression de la cycline D3. E2A favorise également la production du BCR en participant à la transcription des gènes d'Ig (Kwon et al., 2008; Wöhner et al., 2016).

Le principal régulateur de la formation des centres germinatifs est le facteur de transcription BCL6 (B Cell Lymphoma 6), sa délétion entraînant une altération de la formation de ces CG (Cattoretti et al., 1995; Huang et al., 2014). BCL6 intervient à plusieurs niveaux. Il favorise, dans un premier temps, la migration des cellules afin d'initier la formation du CG (Huang et al., 2013; Kitano et al., 2011). Au niveau cellulaire, ce facteur permet la prolifération des cellules en inhibant l'expression de facteur





inhibiteur du cycle tel que CDKN1A ou celle de facteur anti-apoptotique comme BCL2 (Saito et al., 2009). Il cadre également la mise en place de la fonction effectrice des lymphocytes B en inhibant l'expression de certains TLR comme le TLR9 et il contribue aux voies de signalisation en aval du BCR et de son co-récepteur CD40 (Basso et al., 2010). BCL6 contribue aussi à favoriser les mécanismes de remaniements géniques établis dans le centre germinatif notamment via la régulation de certains acteurs de la réparation de l'ADN comme ATM, ATR et TP53 (Phan and Dalla-Favera, 2004; Ranuncolo et al., 2007). Enfin, BCL6 restreint également l'expression de gènes tels que BLIMP1 qui favorise la différenciation des cellules en plasmocytes et limite, de ce fait, une maturation trop précoce (Tunayaplin et al., 2004).

D'autres acteurs coopèrent afin d'établir la formation et le maintien des centres germinatifs tels que MCL1 (myeloid cell leukaemia 1), IRF8, IRF4, MEF2B et MEF2C et plusieurs d'entre eux (IRF4, IRF8 et MEF2B) participent, entre autre, à la régulation de l'expression de BCL6 (De Silva et al., 2012; Vikstrom et al., 2010; Ying et al., 2013).

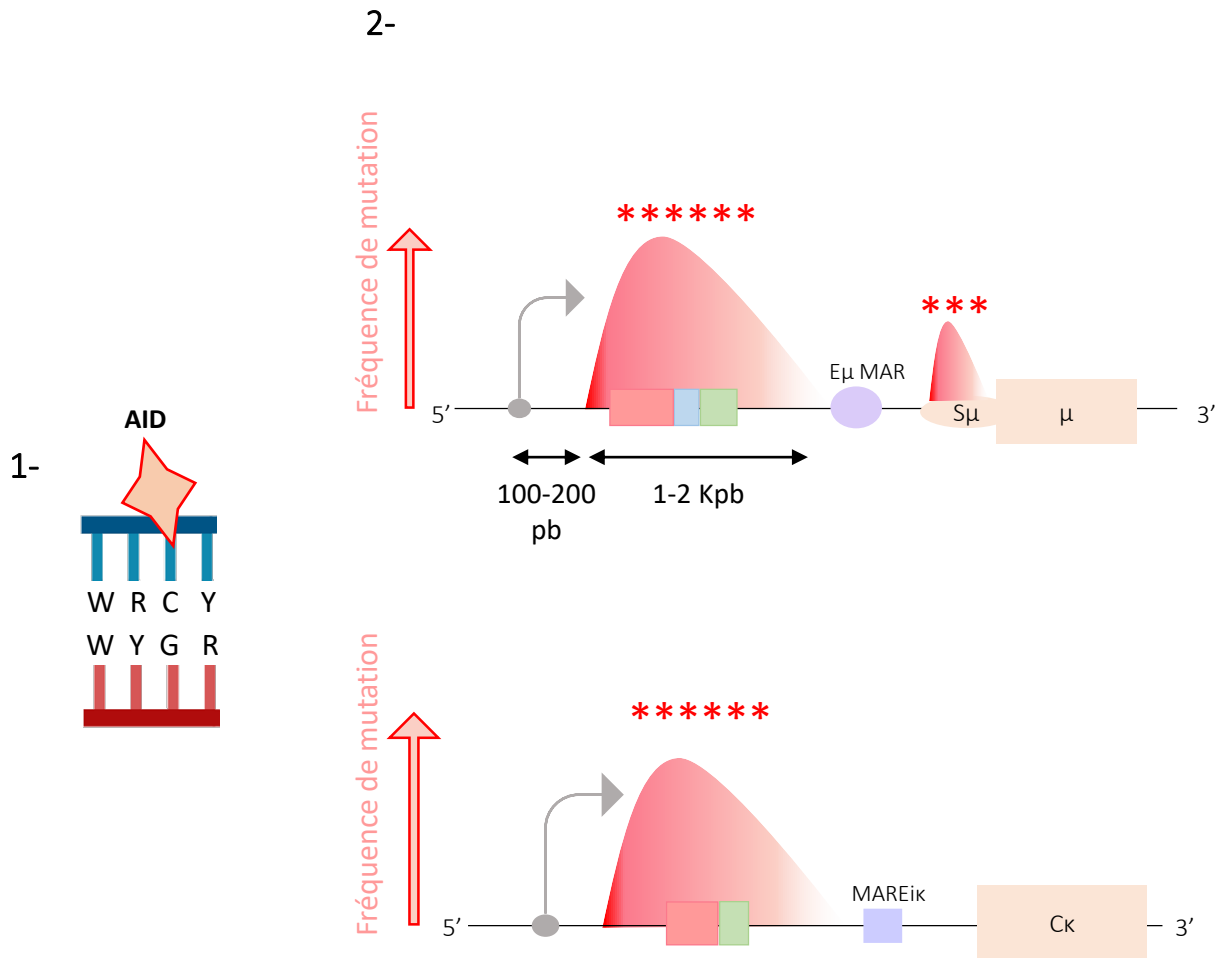
Dans le centre germinatif, les centroblastes subiront le mécanisme d'hypermutation somatique (SHM : somatic hypermutation) qui consiste à optimiser l'affinité des immunoglobulines pour un antigène.

#### **A. L'hypermutation somatique.**

La SHM est un mécanisme spécifique de la cellule B, consistant en l'introduction de nombreuses mutations au sein des régions variables des chaînes d'immunoglobulines (**Figure 7**). Ce processus a pour but de modifier la structure des domaines hypervariables appelés CDR (complementary determining region) afin d'accroître la liaison de l'Ig avec son antigène cible.

La SHM est habituellement décrite en deux phases : une phase d'initiation du mécanisme, qui nécessite un contexte transcriptionnel particulier pour favoriser l'intervention de l'enzyme initiateur qu'est AID (Activation Induced cytidine Deaminase) et une phase d'introduction des mutations qui fait appel à l'intervention des voies de réparation de l'ADN (Rada et al., 1998).

AID est une enzyme de 24 kDa, produite à partir du gène *Aicda*, principalement exprimée dans les cellules B activées du centre germinatif au sein desquelles elle est indispensable à la mise en place de la SHM et du mécanisme de commutation de classe (CSR : class switch recombination) décrite ultérieurement. Cette enzyme est un membre de la famille de protéine APOBEC (Apolipoprotein B mRNA editing catalytic component) qui regroupe des enzymes catalysant la déamination des cytosines (Rebhandl et al., 2015). AID se compose de plusieurs régions commençant par une séquence NLS (Nuclear Localisation Sequence) en position N-terminale suivie par un domaine catalytique de déamination, d'une région linker, d'un domaine APOBEC-like puis d'une séquence d'exportation NES (nuclear export signal). Des mutations induites sur diverses régions indiquent que les activités d'AID,



**Figure 7 : Hypermutation somatique (SHM).**

1- Séquence consensus ciblée par AID lors du processus de SHM : WRCY (W : A/T, R : A/G, Y : T/C). L'enzyme procède à la déamination de la cytosine du motif WRCY en uracile.

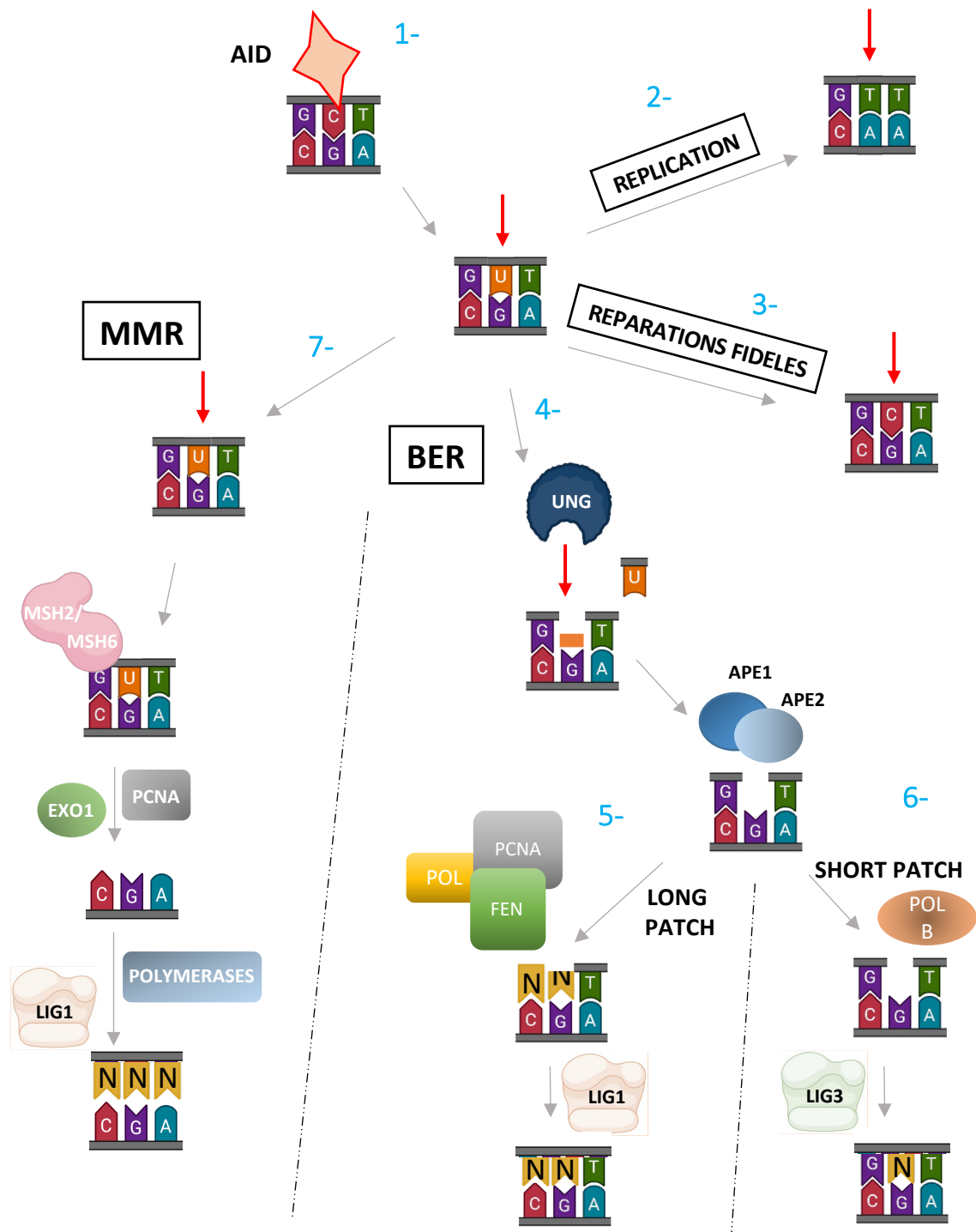
2- Profil de la répartition des mutations somatiques sur les loci IgH et Igk (les sites d'initiation de la transcription sont représentés par des flèches grises et les mutations par des étoiles rouges).

l'initiation de la SHM ainsi que de la CSR, sont portées par différentes régions de sa structure protéique : les mutations inhibant un mécanisme ne limitent pas la réalisation du second.

Les protéines de la famille APOBEC sont, pour la plupart, des déaminases de l'ARN. L'activité catalytique d'AID, quant à elle, s'exerce sur une région d'ADN simple brin dans un motif WRCY (W : A/T, R : A/G, Y : C/T) (**Figure 7**) (Dickerson et al., 2003; Hwang\* et al., 2015; Larijani and Martin, 2007). Ce motif est libéré des liaisons hydrogènes qui l'associe au brin opposé lors de la transcription. L'ARN polymérase II qui initie la transcription au promoteur du segment V réarrangé permet d'exposer le brin d'ADN non matrice pour faciliter l'accès à AID (Bransteitter et al., 2003). Le complexe formé par l'ARN pol II et le facteur d'élongation de la transcription SPT5 (suppressor of ty5 homolog) favorise le recrutement d'AID (Maul et al., 2014; Pavri and Nussenzweig, 2011). Cette nécessité d'avoir un substrat simple brin pour AID contribue à contrôler son activité qui s'exerce sur une région commençant environ 100 pb en aval du site de transcription et qui s'étend sur 1Kb.

Différents paramètres permettent le contrôle et aident au bon déroulement de l'étape d'initiation de la SHM. Le mécanisme est optimisé par la dégradation de l'ARN nouvellement synthétisé, médiée par le complexe RNA exosome et par l'initiation d'une transcription anti-sens qui permettent l'accessibilité du brin matrice à AID (Pefanis and Basu, 2015; Pefanis et al., 2014; Perlot et al., 2008). Cette accessibilité serait également favorisée par la présence de l'Histone 3.3 (variant de l'histone 3) (Romanello et al., 2016). Il semblerait aussi que les modifications épigénétiques des histones, telles que l'acétylation, favorisent le mécanisme d'hypermutation (Sheppard et al., 2018; Woo et al., 2003). Enfin, la composition en base de la séquence est aussi importante pour le recrutement d'AID : des structures appelées G-quadruplex, formées par l'association de plusieurs guanines de la séquence cible, fournissent à l'enzyme un point d'ancrage plus solide qu'un brin linéaire. Ces structures sont essentiellement présentes dans les régions ciblées lors de la CSR : les régions switchs (S). Un modèle murin portant une mutation au sein du domaine de reconnaissance des G-quadruplex mime le phénotype retrouvé chez des patients atteints d'un syndrome hyper IgM de type 2 qui présentent une absence d'hypermutation somatique ainsi qu'une restriction d'Ig à l'isotype M (Yewdell et al., 2020).

D'autres facteurs peuvent contribuer à la modulation du mécanisme d'hypermutation ; c'est notamment le cas pour le facteur d'épissage SRF1-3 (Serine/Arginine splicing factor 1-3) qui accroît le taux d'hypermutation notamment en favorisant l'accumulation d'AID au sein du noyau (Kanehiro et al., 2012; Kawaguchi et al., 2017). Ce facteur d'épissage est également capable de recruter la topoisomérase I (Kumar Singh et al., 2019). Cette enzyme a pour fonction de lever les contraintes des torsions de l'ADN qui lui sont imposées par l'avancement du complexe de transcription de l'ARN polymérase II. Elle génère donc des cassures transitoires sur un des deux brins de l'hélice et favoriserait l'accès du simple brin à AID (Baranello et al., 2013). Cependant, l'inhibition de cette enzyme résulte à



**Figure 8 : Prise en charge de l'uracile généré par AID lors du processus de SHM.**

- 1- Création d'un mésappariement au sein du brin d'ADN suite à la déamination d'une cytosine en uracile par l'enzyme AID.
- 2- Réplication classique générant une mutation de type transition.
- 3- Prise en charge par les voies de réparation fidèle au brin matrice de l'ADN.
- 4- Prise en charge du mésappariement par la voie BER initiée par UNG et faisant intervenir des polymérase infidèles telles que REV1 ou la pol $\beta$ .
- 5- Long patch repair : retrait d'une dizaine de nucléotides possible.
- 6- short patch repair : retrait d'un seul nucléotide.
- 7- Réparation par la voie MMR infidèle initiée par le complexe MSH2/MH6.

une augmentation de la fréquence de mutation induite par SHM. L'absence des coupures médiées par la topoisomérase I pourrait entraîner une accumulation des torsions de l'ADN, diminuant progressivement l'efficacité d'avancement de l'ARN polymérase II, rendant ainsi accessible plus longtemps l'ADN simple brin libéré lors de la transcription (Maul et al., 2015; Pommier et al., 2016).

L'étape de maturation des régions variables par la SHM est un indispensable à l'efficacité de fonctionnement des Ig, mais elle doit être contrôlée afin d'assurer un champ limité d'action à AID pour protéger les gènes à proximité. Des études initialement réalisées dans des tumeurs ont corrélé l'expression de AID et la présence d'un taux élevé de mutations dans plusieurs gènes dont certains oncogènes (Gordon et al., 2003; Müschen et al., 2000; Pasqualucci et al., 1998, 2001). En 2008, l'analyse d'une centaine de gènes réalisée chez la souris, dans des fonds génétiques présentant la perte de l'expression d'AID et des voies de réparations associées telles que le sont les voies BER (Base excision repair) et MMR (mismatch repair), a confirmé que de nombreux gènes subissent le mécanisme de SHM (bien qu'ils présentent un taux de mutations dix fois moindre que celui des gènes d'Ig). L'ensemble de ces gènes, appelés « AID Off target », est classé en trois groupes distincts établis en partie par l'efficacité de la réparation sous-jacente à l'action d'AID (Liu et al., 2008). Un premier groupe rassemble les gènes dont le mésappariement de base est pris en charge par des voies de réparations infidèles au brin matrice et générant des erreurs. Parmi ces gènes, sont répertoriés, entre autre, des acteurs indispensables au développement tels que *Bcl6*, *Pim1* ou *Cd83*. Le second groupe réunit les gènes hautement surveillés et réparés comme *Pax5* et *Cmyc*. Et, enfin, le dernier groupe se compose de gènes dont les mutations sont indétectables (Liu et al., 2008).

En effet, après l'intervention d'AID, la cytosine alors déaminée en uracile crée un mésappariement avec la guanine du brin opposé : cette structure sera prise en charge par les voies de réparation spécifiques des mésappariements de l'ADN lors de la deuxième phase du processus d'hypermutation. Dans un cas d'altération de l'ADN classique, les voies de réparation mises en place ont pour but de copier de façon la plus fidèle possible le brin matrice non touché par cette altération afin d'éviter l'introduction de mutation délétère. Dans le cadre de la SHM, la cellule B a besoin de générer des mutations qui ne soient pas fidèlement réparées ; on parle alors de mécanismes de réparation sujet aux erreurs ou « error-prone ».

Le mésappariement généré par l'intervention de AID peut subir trois sorts différents (**Figure 8**) :

- 1) Il n'est pas détecté et l'ADN polymérase considèrera l'uracile comme une thymine lors de la réplication, générant ainsi une mutation de type transition C vers T (Neuberger et al., 2005).
- 2) Il est pris en charge par les voies de réparation BER et MMR classiques faisant appel à des acteurs qui conduisent à une réparation fidèle, ne générant ainsi aucune mutation.



- 3) Il peut être pris en charge par les voies de réparation MMR et BER dites « non conventionnelles » car elles font appel à des acteurs de réparation qui génèrent des erreurs dans la copie du brin matrice.

Dans la voie de réparation MMR, le mésappariement uracile-guanine généré par AID est détecté par un hétérodimère protéique formé de MSH2 et MSH6 (MutS Homolog) (Jiricny, 2006). Ces deux protéines interagissent avec MLH1 (MutL Homologie 1) et PSM2 (Postmeiotic segregation 2) qui recrutent l'exonucléase EXO1 associée à PCNA (proliferating cell nuclear antigen) (Saribasak and Gearhart, 2012). Ce dernier est un acteur des voies de réparation classiques, qui agit dans les voies de réparation error-prone lorsqu'il est mono-ubiquitinylé sur la lysine 164. Cette modification permet alors le recrutement de polymérase infidèles (Langerak et al., 2007; Roa et al., 2008). EXO1 peut exciser entre une dizaine et une centaine de bases, qui seront remplacées par l'ADN polymérase  $\eta$  infidèle, suivie par l'action de l'ADN ligase I (**Figure 8**). Cette voie de réparation permet de générer des mutations majoritairement sur les bases A/T (Rada et al., 2004).

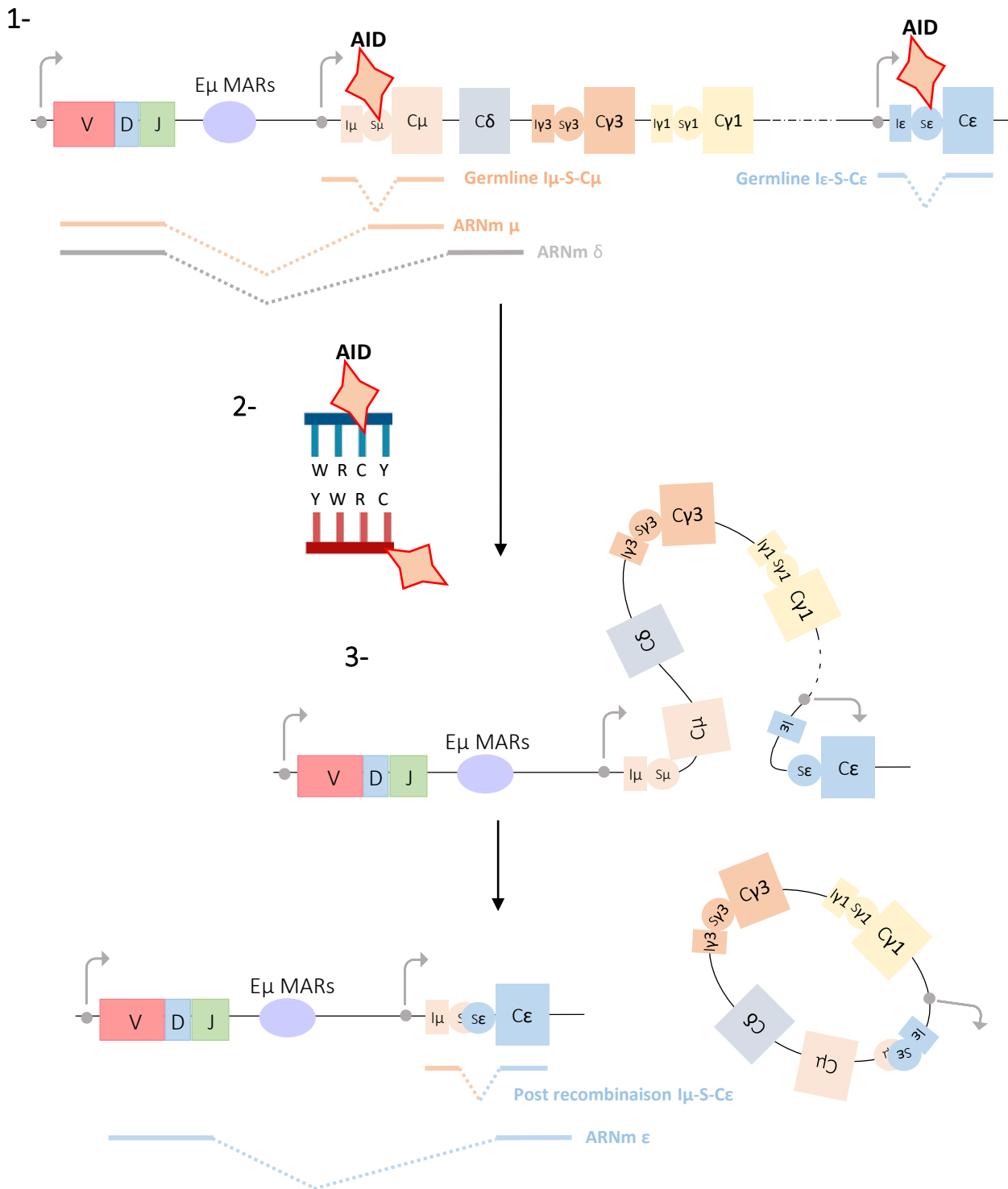
La voie BER, quant à elle, est initiée par l'action de l'enzyme UNG (Uracyl-DNA-Glycosylase) qui a la capacité d'exciser l'uracile générée par AID créant ainsi un site abasique (Di Noia and Neuberger, 2002). Il semblerait que SMUG1, une seconde ADN-glycosylase, soutiendrait l'action d'UNG en contribuant aux mutations aux bases A-T (Di Noia et al., 2006; Girelli Zubani et al., 2017).

La production de ce site abasique permet le recrutement d'endonucléase comme APE1 et APE2 (apuriniques/apyrimidiques endonuclease) qui clivent le brin coté 5' du site de la cytosine originale pour créer une coupure simple brin (Stavnezer et al., 2014). Cette coupure entraîne le recrutement de différents acteurs agissant en cascade pour réparer l'excision. Selon les facteurs recrutés, deux modèles de réparation peuvent être initiés. Le short patch repair entraîne l'excision d'un seul nucléotide qui sera remplacé via l'action des ADN polymérase  $\beta$  et  $\lambda$  suivies par l'ADN ligase III (**Figure 8**) (Almeida and Sobol, 2007). La délétion de ces deux ADN polymérase, intervenant également dans les voies classiques de réparation, montrent une faible mais significative diminution du taux d'hypermutation.

Une troisième polymérase pourrait intervenir dans la prise en charge de l'excision de l'uracile par la voie BER ; il s'agirait de l'ADN polymérase REV1 qui favoriserait l'introduction de mutation de type transversion d'une cytosine vers une guanine (Jansen et al., 2006).

Le second modèle de réparation ou long patch repair, comprend l'intervention de FEN1 (flap endonuclease 1), une endonucléase qui peut retirer entre deux et une dizaine de nucléotides qui seront remplacés par l'ADN polymérase  $\eta$  dont le recrutement serait également médié par PCNA (Fortini and Dogliotti, 2007; Krokan and Bjoras, 2013; Robertson et al., 2009). Enfin, le processus de long patch repair est clôturé par l'intervention de l'ADN ligase I, souvent associée à un co-facteur





**Figure 9 : Exemple du processus de la commutation de classe entre C $\mu$  et C $\epsilon$ (CSR).**

1- Représentation du locus IgH qui reprend la composition des gènes constants (exon I, région switch (s) et gène constant (C), les sites d'initiation de la transcription (représentés par des flèches) et les différents transcrits germinaux offrant l'accessibilité des régions à AID.

2- Séquence consensus ciblée par AID lors du processus de CSR.

3- Les mésappariements générés dans les régions switch donneuses et acceptrices par l'intervention d'AID induisent une cassure double brin. Les régions switch rapprochées sont ensuite liguées, libérant ainsi la portion du locus initialement située entre ces deux régions sous forme d'épisome.

XRCC1 (**Figure 8**). Il semblerait que, dans le cadre de la réparation error-prone, XRCC1 soit limitant dans le processus de SHM : sa délétion à l'état hétérozygote entrainerait une augmentation du taux de la fréquence d'hypermutation par rapport à un contexte sauvage (Saribasak et al., 2011).

L'inhibition de la voie BER, créée par la délétion de UNG, met en avant que cette voie de réparation contribuerait en grande partie à l'introduction des mutations de types transversions G/C (Di Noia and Neuberger, 2002). Chez l'Homme, une mutation de cette enzyme entrainerait un syndrome d'immunodéficience d'Hyper-IgM de type 5 qui serait, cependant, majoritairement dû à un biais dans le processus de commutation de classe.

Si l'hypermutation somatique permet de générer 40 % de mutation sur les bases G/C et 60 % sur les bases A/T, la commutation de classe (CSR), quant à elle, est un mécanisme qui permet de faire varier la portion constante de l'immunoglobuline afin d'apporter à l'Ig une fonction effectrice différente (**Figure 9**).

Lorsque le centroblaste réduit sa prolifération, qu'il migre vers la zone claire et devient un centrocyte, il est soumis à un nouveau type de sélection qui consiste à conserver les cellules B dont les gènes d'Ig mutés reconnaissent l'Ag avec une affinité importante. Dans la zone claire, les cellules B nouvellement mutées rencontrent un Ag présenté par les cellules dendritiques folliculaires. Cette rencontre permet à la cellule B d'exprimer à la surface de ces molécules du CMH des portions de l'Ag qui seront reconnues par un lymphocyte T folliculaire helper ou T<sub>FH</sub> (Victoria and Nussenzweig, 2012; Victoria et al., 2010). Ce nouvel échange T-B ne sera permis que si la cellule B est capable de reconnaître l'Ag avec une affinité conséquente. Si cette reconnaissance n'est pas permise, la cellule entrera en apoptose et sera digérée par des macrophages. En revanche, si l'affinité est assez importante, le T<sub>FH</sub> délivrera des signaux de survie à la cellule B, en permettant notamment d'initier la production de la protéine anti-apoptotique BCL-XL (Fang et al., 1997). L'interaction CD40-CD40L, récepteurs respectivement exprimés à la surface des B et des cellules T, contribue à la CSR. En effet, la déficience d'expression du récepteur CD40 à la surface des cellules B est aussi une caractéristique du syndrome d'immunodéficience à hyper-IgM (Revy et al., 2000). L'altération du récepteur bloque la commutation de classe induisant une absence des autres isotypes et conduisant à l'apparition d'infections répétées causées par des bactéries. D'autres signaux, délivrés sous forme de cytokines, participeront à la mise en place du mécanisme de commutation de classe. Une libération d'interleukine 4 (IL4), par exemple, induira une commutation vers un type  $\gamma 1$  ou  $\epsilon$  alors que l'IL5 favorisera une production d'IgA.

### **B. La commutation de classe.**

Cette ultime recombinaison génique, permettant le changement de classe isotypique des immunoglobulines, est également initiée par AID. Chaque gène constant (à l'exception de C $\delta$ ) est composé de trois parties distinctes (**Figure 9**) :

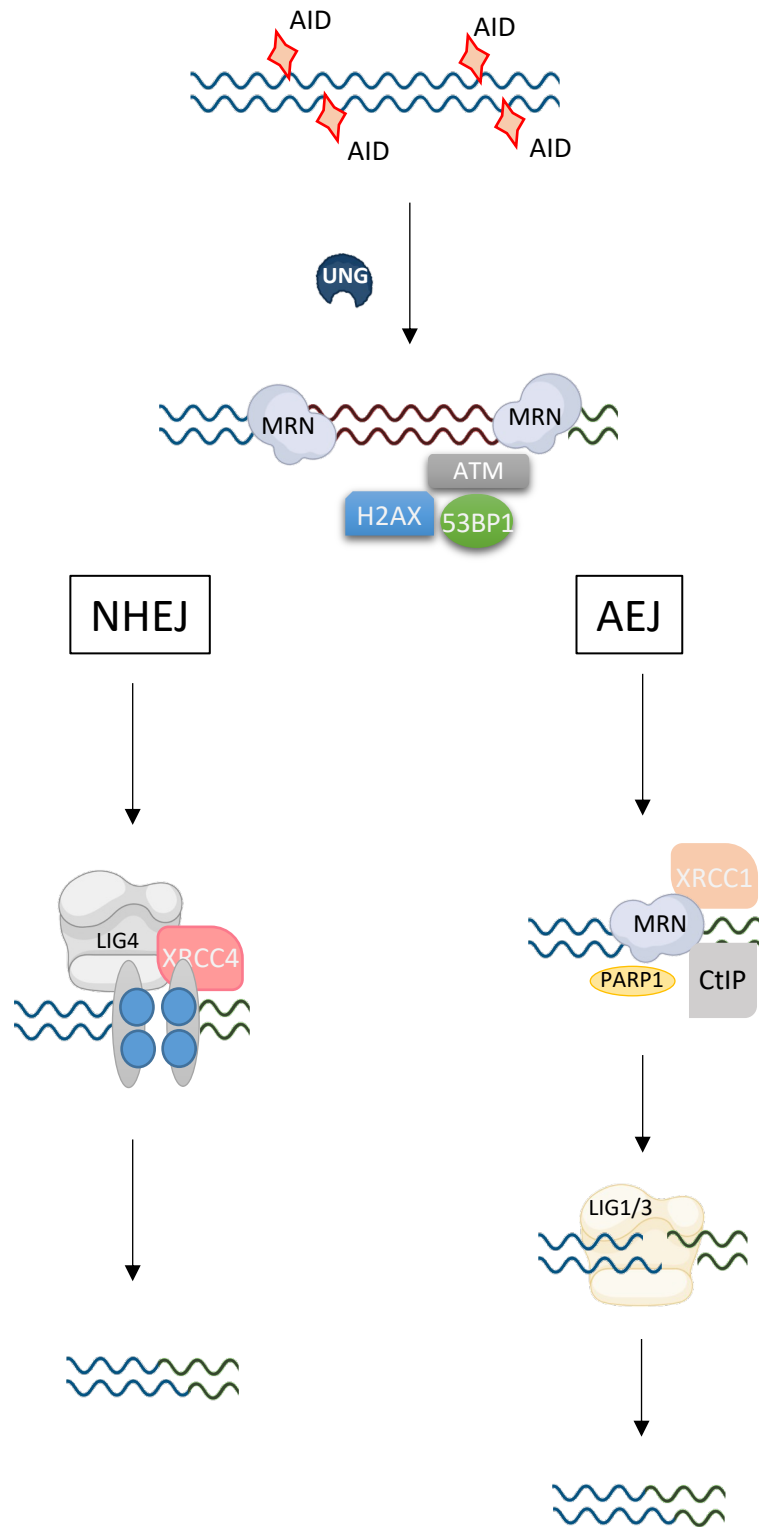


- Un exon I non codant appelé exon « intervening » précédé d'un promoteur I
- Une région switch (S) : région dont la taille varie entre 1 à 14 Kb et qui se compose de motifs répétés avec une séquence riche en G de type GAGCT et GGGGT.
- Le gène constant C

La région  $S_{\mu}$  (en amont de  $C_{\mu}$ ) est définie comme la région donneuse et sera recombinaisonnée avec une région switch  $S_{\epsilon}$  (dans le cas d'un switch vers IgE) appelée région acceptrice. Tout comme pour le mécanisme d'hypermutation, AID intervient sur un brin d'ADN monocaténaire rendu accessible par la machinerie de transcription. La composition des gènes constants, et notamment la présence des promoteurs  $I_{\mu}$  (pour la région donneuse) et  $I_{\epsilon}$  (pour la région acceptrice  $\epsilon$ ), permet d'initier la formation d'un transcrite germinale à chaque région. Si  $S_{\mu}$  subit une transcription active, la région acceptrice sera transcrite en réponse à une stimulation. La libération d'IL4 (exemple cité préalablement) entrainera l'activation d'une voie de signalisation favorisant la production du facteur de transcription STAT6 qui initiera la transcription germinale de la région  $S_{\epsilon}$  (Bacharier and Geha, 2000).

De nouveau, les modifications épigénétiques jouent un rôle important dans l'initiation de la transcription puisque des marques favorisant la transcription, telles que l'acétylation de l'histone H3K9 et H3K27, sont fortement présentes au niveau de la région  $S_{\mu}$  (Wang et al., 2009a; Wu et al., 2018). La région S du transcrite sera épissée comme un intron, le transcrite germinale épissé ne donnera, toutefois, lieu à aucune protéine. Bien qu'il ne soit ni traduit ni productif, ce transcrite est tout de même indispensable au déroulement de la CSR. Par son association au brin matrice et sa composition importante en base guanine, il permet la formation d'une structure triplex ADN-ARN, appelée R-loop, qui favorisera l'accès du brin non matrice à AID (Yu et al., 2003; Zhang et al., 2014b). Cette accessibilité est aussi permise par la composition en séquences répétées des régions switch qui entraîne une pause de l'ARN pol II, complexée à SPT5 (Pavri et al., 2010), dans sa progression sur le brin matrice. D'autres structures de type G-quadruplex pourraient être adoptées par la région S épissée. Ces derniers pourraient jouer également le rôle d'ARN guide pour AID (Pucella and Chaudhuri, 2017; Qiao et al., 2017).

L'action de cette enzyme au sein de la CSR consiste, tout comme pour l'hypermutation somatique, à déaminer une base cytosine en uracile afin d'entraîner son excision et de générer une cassure dans le brin d'ADN (**Figure 9**). L'enjeu pour AID dans la CSR est d'initier une cassure sur chaque brin (matrice et non matrice) afin de créer une cassure double brin (DSB) (Stavnezer and Schrader, 2014). L'accessibilité du brin matrice est favorisée, comme au sein du processus de la SHM, par l'action du RNA exosome qui contribue au recrutement d'AID (Basu et al., 2011; Laffleur et al., 2017). Les régions switch sont également très enrichies en séquences dites « hotspot » d'AID, telles que AGCT, qui



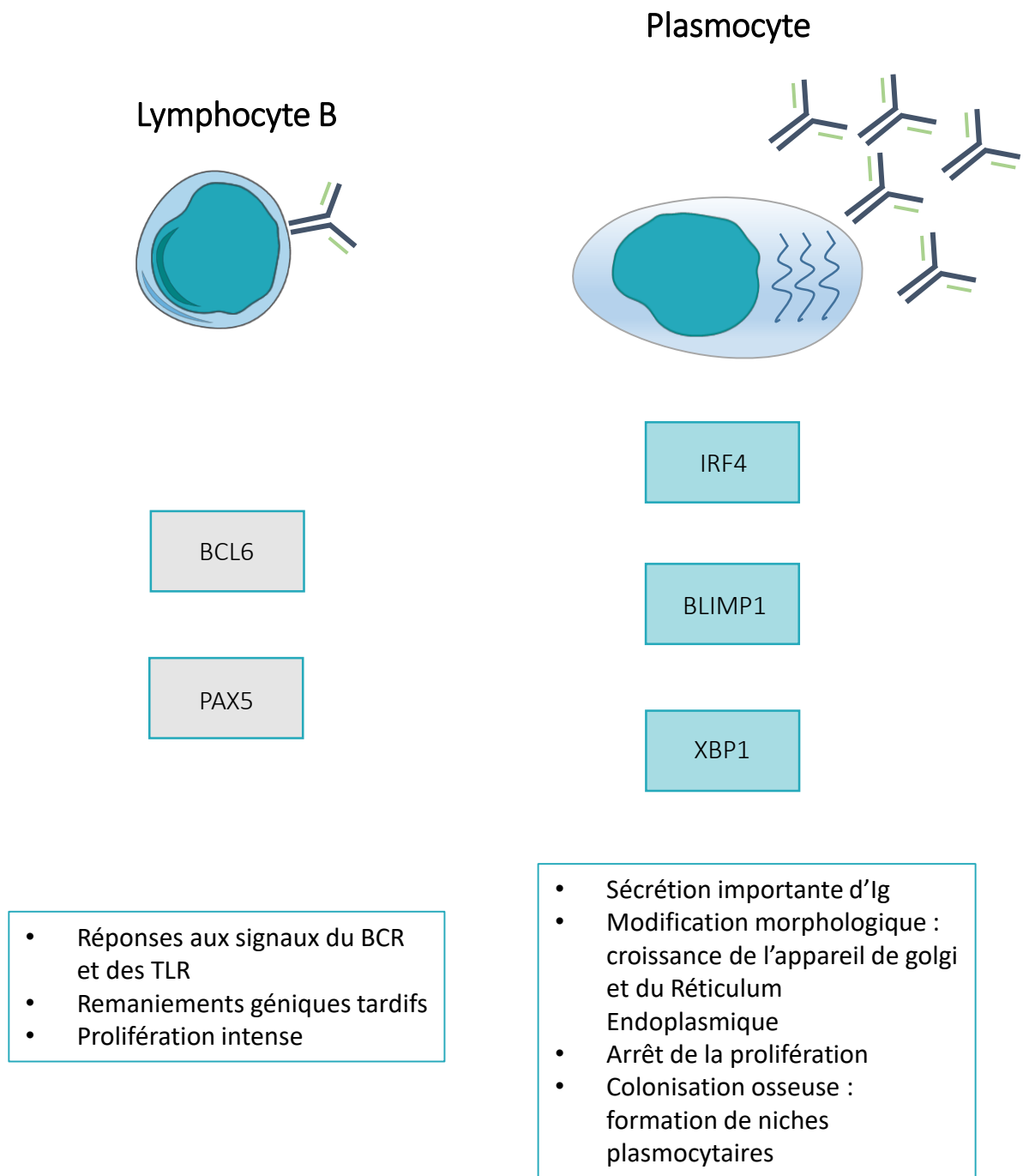
**Figure 10 : Voies de réparations initiées pendant le processus de CSR.**

Après l'intervention d'AID, la création de cassures double brin est médiée majoritairement par l'action d'UNG et des endonucléases APE. Ces cassures sont stabilisées par le complexe MRN et la voie de réparation NHEJ (Non homologous end joining) est majoritairement recrutée, soutenue par la voie AEJ (Alternative end joining).

seraient reconnues par les protéines stabilisatrices 14.3.3 participant au recrutement d'AID (Lam et al., 2013; Xu et al., 2010). La présence de répétitions multiples de ces séquences permet des déaminations très proches sur chacun des brins, générant ainsi des cassures simple brin opposées. Ces cassures sont favorisées par l'intervention d'UNG qui excise l'uracile et permet le recrutement d'endonucléase APE. Ces DSB sont ensuite reconnues par les protéines MRE11, RAD50 et NBS1 qui forment le complexe MRN (**Figure 10**). Celui-ci va activer les voies de réparation spécifiques des coupures double brin, notamment en recrutant la protéine ATM qui induira la phosphorylation de l'histone 2AX (histone 2A family member X) et de 53BP1 (Chaudhuri and Alt, 2004; Lee and Paull, 2007; Manis et al., 2004). Ces deux évènements vont participer à la formation d'un environnement stable à l'extrémité des coupures. Les régions switch donneuse et acceptrice vont ensuite être physiquement rapprochées puis liées, en priorité, par la voie de réparation NHEJ. Tout comme pour les recombinaisons VDJ, les protéines Ku70/Ku80 maintiennent les DSB et participent au recrutement de l'ADN-PKc qui phosphoryle différents acteurs dont l'endonucléase ARTEMIS. Enfin, XRCC4 et l'ADN ligase 4 lient les extrémités (Lieber, 2010). Différents modèles de délétions ont permis de mettre en avant l'implication d'une autre voie de réparation (Boboila et al., 2010) : la voie A-EJ pour Alternative End Joining qui fait intervenir différents acteurs tels que XRCC1 (X-Ray Repair Cross Complementing protein 1), CtIP (Ct-interacting protein) et la Poly-ADP Ribose polymerase 1 (PARP-1) ainsi que les ADN ligase I et III (Boboila et al., 2012). La commutation de classe et l'hypermutation somatique sont des évènements irréversibles permettant de finaliser la maturation des immunoglobulines en améliorant l'affinité antigénique et en leur apportant une fonction plus précise. Suite à ces deux processus, les lymphocytes B activés vont entrer dans la phase terminale de maturation afin de devenir un plasmocyte sécréteur d'immunoglobuline. Cette dernière étape de différenciation est mise en place par la modification du programme de transcription spécifique de la cellule B. Dans un premier temps, la cellule cesse d'exprimer CXCR4, CXCR5 et certaines molécules de rétention telles que les intégrines  $\alpha_4$  et  $\beta_1$  afin de quitter la zone claire et de sortir définitivement du centre germinatif.

### **1.2.5. Différenciation plasmocytaire : vers une usine à immunoglobulines**

La différenciation d'une cellule B, issue du centre germinatif, en plasmocyte est en grande partie régit par le facteur transcriptionnel IRF4. Son expression, faible et très transitoire dans les cellules du centre germinatif, stimule l'expression de BCL6 et PAX5, facteurs indispensables à la formation du CG. Ces deux protéines comptabilisent, parmi leurs différentes fonctions, un rôle de répresseur de la différenciation plasmocytaire, en inhibant (comme décrit précédemment) l'expression de nombreux gènes plasmocytaires (Nera and Lassila, 2006; Tunyaplin et al., 2004).



**Figure 11 : Facteurs indispensables à la différenciation plasmocytaire et modifications morphologiques du plasmocyte.**

La différenciation plasmocytaire est régie par un changement transcriptionnel important qui contribue à l'élaboration de la fonction de sécrétion massive d'immunoglobuline du plasmocyte, mais également au changement morphologique inhérents à cette fonction. Les facteurs transcriptionnels indispensables à cette différenciation sont répertoriés sur cette figure. Adapté d'après Shapiro-Shelef M., et al 2005

Une expression plus importante et prolongée d'IRF4 favorisera, en revanche, la transcription de l'ensemble des gènes permettant la différenciation en plasmocyte (Low et al., 2019). Un modèle murin de délétion d'IRF4 présente des cellules B incapables de finaliser leur différenciation et bloquer au sein du centre germinatif. Ce facteur s'associe également à PU.1 afin d'activer les enhanceurs des gènes de chaînes lourdes et légères d'immunoglobuline, participant ainsi à la fonction effectrice du plasmocyte (Sciammas et al., 2006).

IRF4 module aussi l'expression de *Prdm1* qui permet la production de BLIMP1 (Angelin-Duclos et al., 2000). Ce régulateur transcriptionnel inhibe à son tour l'expression de *c-myc*, *Bcl6* et *Pax5*, favorisant ainsi la répression d'un phénotype de cellule B du CG. BLIMP1 permet aussi l'expression de XBP1, un facteur qui jouera plus un rôle dans la fonction des plasmocytes en contribuant à la sécrétion des immunoglobulines via la régulation de gènes impliqués dans l'adressage des protéines au réticulum endoplasmique. Ces changements d'expressions géniques permettent à la cellule B de modifier sa morphologie pour devenir, dans un premier temps, un plasmablaste (**Figure 11**). Cette étape intermédiaire entre le lymphocyte B et le plasmocyte est caractérisée par l'expression à la fois de marqueurs de cellule B (BCR/CD86) et de marqueurs plasmocytaires (CD138) (Oracki et al., 2010; Smith et al., 1996). Progressivement, le plasmablaste réduit sa prolifération, amorce une légère sécrétion d'Ig, perd l'expression de marqueurs B, comme B220 et CD19, et subit des modifications morphologiques conséquentes telles que la croissance d'un Réticulum endoplasmique et d'un appareil de golgi. Ces modifications morphologiques, agrandissant la taille de la cellule, sont inhérentes à la fonction de sécrétion d'immunoglobulines (Nutt et al., 2015).

Les plasmocytes issus de la différenciation d'une cellule B d'un centre germinatif sont des plasmocytes dit à longue durée de vie (en opposition au plasmocytes à courte durée de vie issus des cellules de la zone marginale) et se situent dans la circulation sanguine et les organes lymphoïdes secondaires. Le transcriptome des plasmocytes inclut des gènes anti-apoptotiques régulés de façon positive tels que *Mcl1* et *Bim* et des gènes pro-apoptotiques réprimés comme *Bax* (Carrington et al., 2010; Tellier and Nutt, 2019; Vikstrom et al., 2010; Wang and Bhattacharya, 2014). Les plasmocytes peuvent éventuellement se trouver dans la circulation avant de retourner se loger dans la moelle osseuse ; on parlera alors de cellules B recirculantes. Cette migration est favorisée par la réexpression de CXCR4 qui lie CXCL12 présent à la surface des cellules stromales. Ces dernières favoriseront aussi la formation d'une niche plasmocytaire en délivrant des signaux de survie supplémentaires au plasmocyte. Le facteur XBP1 semble également jouer un rôle dans la colonisation osseuse (Hu et al., 2009).

Les lymphocytes du centre germinatif ont également la capacité de se différencier en cellules B mémoires (**Figure 6**).





### **I.2.6. Les cellules B mémoires**

Les cellules B mémoires sont sollicitées lors d'une seconde infection par un pathogène. Différenciées à partir de B activées dans un centre germinatif, elles se caractérisent par une prolifération lente et une expression d'immunoglobuline de surface de type variable (en fonction du switch subit lors de leur activation) mais elles n'ont pas la capacité de sécréter des Ig (ou très faiblement). Elles ont subi la maturation d'affinité et sont donc appelées B mémoire pour leur capacité à reconnaître rapidement l'Ag à l'origine de leur activation (Dogan et al., 2009; Good-Jacobson and Shlomchik, 2010). Ces cellules circulent dans le sang et stagnent dans les follicules de la rate et des ganglions. Lors d'une nouvelle infection par le pathogène d'origine, leur rapidité de reconnaissance de l'Ag leur permet de le présenter via leur molécule de CMHII au T<sub>FH</sub>, cellules avec lesquelles elles interagissent également via leur marqueur B7.1 (qui lie le CD28 présent à la surface des T). Cette interaction favorise leur différenciation rapide en plasmocyte et, de ce fait, la production d'immunoglobuline (Bergmann et al., 2013). Certaines cellules réintègrent aussi le centre germinatif et subissent de nouveau le processus de SHM. Ces lymphocytes B mémoire sont donc des acteurs d'une réponse adaptative persistante de façon indépendante à la présence du pathogène.

Cette seconde étape (maturation et différenciation finale des cellules B), se déroulant dans les organes lymphoïdes secondaires, se base sur une coopération entre les deux lignées lymphocytaires qui échangent différents signaux afin de se stimuler l'une et l'autre. Cependant, certains antigènes sont maintenant connus pour faciliter l'activation des cellules B de façon indépendante à cette coopération T-B ; on les qualifie alors d'antigènes thymo-indépendants (TI) de classe 1 ou 2. Les TI-1 induisent une expansion polyclonale des cellules B médiée principalement par la liaison de l'antigène au TLR, comme c'est le cas du LPS (lipopolysaccharide) (Bekeredjian-Ding and Jengo, 2009). Ces antigènes thymo-indépendants n'induiront cependant pas la mise en place de la SHM et la différenciation en cellule B mémoire.

Les TI-2 rassemblent des antigènes ayant une structure répétitive qui induisent une interconnexion de plusieurs BCR. Les cellules environnantes telles que les macrophages ou les cellules dendritiques favorisent cette activation en délivrant des signaux co-stimulateurs (Balázs et al., 2002). Cette indépendance à la cellule T permet de développer une réponse humorale rapide, avec une production conséquente d'IgM. Bien que ces IgM soient peu spécifiques, elles sont indispensables à la mise en place d'une réponse efficace notamment contre des bactéries encapsulées qui échappent à la vigilance phagocytaire. Cette réponse est principalement exercée par les cellules B de la zone marginale mais également par un autre type de cellules appartenant à la lignée lymphocytaire B, appelées B1. En effet, on distingue deux types majoritaires de cellules B. Le premier, les B2, sont les cellules de la voie dite classique de développement : les B folliculaire et B de la zone marginale (préalablement décrite). Le



second type se compose des cellules B1 qui subissent un développement parallèle dont l'origine est encore beaucoup discutée.

### **I.2.7. Les cellules B1**

Les cellules B1 sont à l'origine de la sécrétion d'immunoglobuline de type M produite naturellement, indépendamment d'une présence antigénique et sont généralement associées à l'immunité innée. Comme expliqué précédemment, elles sont cependant capables de générer une production d'Ig plus conséquente suite à la détection d'un antigène et ce sans l'intervention des cellules T, mais elles ne sont jamais à l'origine d'une mémoire immunologique. Ces lymphocytes représentent, chez la souris, 5% des cellules B et se situent principalement dans les cavités pleurale et péritonéale. Deux sous-groupes sont répertoriés : les B1a et les B1b qui se différencient par l'expression du marqueur CD5. Ces cellules, dotées d'une capacité d'autorenouveaulement, sont légèrement auto-réactives. Elles sont aussi caractérisées par une faible diversité de répertoire et un taux de nucléotides N insérés par la TdT limité (Baumgarth, 2011; Montecino-Rodriguez and Dorshkind, 2006; Prieto and Felipe, 2017).

Les lymphocytes B1 sont les premières cellules B à être produites au cours du développement. Elles proviennent du foie fœtal et sont générées en quantité importante durant la vie embryonnaire. Au stade adulte, leur origine est controversée et deux possibilités sont à ce jour répertoriées :

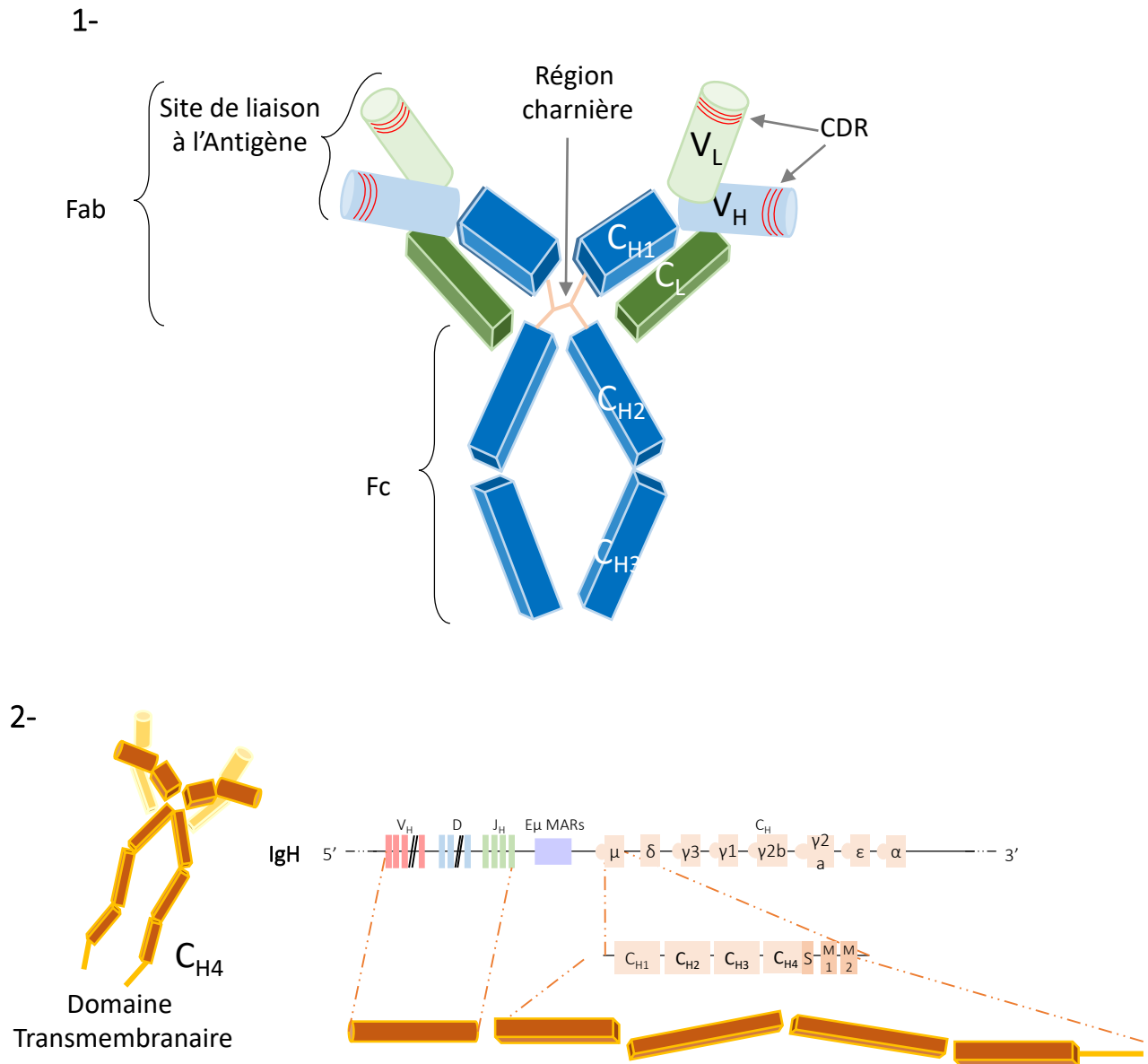
- 1) Le modèle « Induced differentiation » propose que les cellules B1 et les B2 proviennent du même précurseur et que le choix de différenciation dépend de l'intensité des signaux perçus par le BCR (Casola et al., 2004).
- 2) Le second modèle « Layered immune system » suggère en revanche une différenciation à partir de précurseurs différents (Barber et al., 2011; Esplin et al., 2009).

Certaines études démontrent que les deux modèles ne sont pas exclusifs et qu'il serait possible à la fois que certains précurseurs favorisent un type cellulaire et que la signalisation du BCR soit impliquée (Hardy and Hayakawa, 1991).

Les lymphocytes B1 se différencient alors des B2 de par leur fonction et leur localisation mais également de par leur indépendance à certaines cytokines, comme l'IL7, ou à certains facteurs tels que PU.1. D'autres acteurs sont cependant indispensables à leur différenciation ; c'est notamment le cas de la protéine de liaison à l'ARN Lin28b qui participe, entre autre, à la sélection positive des cellules B1 auto réactives.



La première étape de différenciation du lymphocyte B repose sur la production de son récepteur spécifique et la seconde consiste à son activation qui aboutit à la différenciation en plasmocyte sécréteur d'immunoglobulines. Les mécanismes détaillés dans ce premier chapitre illustrent la richesse du répertoire d'immunoglobulines. Cette dernière est liée à la capacité spécifique du lymphocyte B à réarranger les gènes à l'origine de la production des Ig. Dans le second chapitre, seront répertoriées les différentes classes d'immunoglobulines et détaillées les compositions des trois loci des gènes à l'origine de ce répertoire très diversifié.



**Figure 12 : Structure détaillée d'une immunoglobuline.**

1- Structure d'une IgG : positionnement des fragments Fab (Fragment antigen binding) et Fc (fragment cristallisable), de la région charnière et des CDR (complementary determining region).

2- Représentation d'une IgM transmembranaire et du locus IgH (détail des exons du gène C<sub>μ</sub>).

### II.1. Les Immunoglobulines

#### II.1.1. Structure des immunoglobulines

Les immunoglobulines sont des glycoprotéines d'environ 150 kDa composées de quatre chaînes peptidiques : 2 chaînes lourdes identiques (H : Heavy) de 50 kDa et 2 chaînes légères identiques (L : Light) de 25kDa (**Figure 12**). Cette structure enchâssée dans la membrane et associée à deux chaînes supplémentaires,  $Ig\alpha$  et  $Ig\beta$ , forme le BCR. Celui-ci lie les antigènes et les chaînes  $Ig\alpha$  et  $Ig\beta$  participent à la transmission du signal vers le compartiment intracellulaire. Les immunoglobulines sont aussi sécrétées sous forme soluble ; on parle alors d'anticorps.

Chaque chaîne d'Ig est constituée d'un domaine Variable (V), cible notamment du mécanisme d'hypermutation et d'une portion constante (C) dont la variation, induite par la CSR, confère la fonction effectrice de l'immunoglobuline (Edelman, 1973).

Les deux chaînes lourdes sont associées entre elles par des ponts disulfures établis entre des résidus cystéines. Chacune des chaînes lourdes est également liée, toujours via des ponts disulfures, à une chaîne légère. Les immunoglobulines ont alors une structure en Y. Il existe 5 chaînes lourdes différentes : IgM, IgD, IgG, IgA et IgE qui peuvent s'associer à 2 chaînes légères :  $\kappa$  ou  $\lambda$ .

#### A. *Structure globale*

De façon générale, les chaînes lourdes sont constituées de 3 régions constantes : les deux premiers forment la base du Y et portent la fonction effectrice et la troisième incluant la portion variable, forme « les bras du Y » et permet la reconnaissance antigénique. Les chaînes légères, quant à elles, constituent également les bras du Y et se composent d'une unique région constante, dont le rôle semble seulement structural, ainsi que d'une région variable participant aussi à la reconnaissance de l'antigène. Chaque bras est alors appelé fragment Fab (Fragment antigen binding) et la base du Y est le Fragment cristallisable (Fc). La liaison des 3 fragments est établie par une région charnière qui apporte une certaine flexibilité à la molécule (**Figure 12**) (Porter, 1973).





### **B. Région Variable**

Les régions variables portent des régions peptidiques courtes dites Hypervariables nommées HV1, HV2 et HV3 (la plus variable) qui sont à l'origine de l'affinité et la reconnaissance de l'antigène. On les décrit aussi sous le nom CDR1 2 ou 3 pour complementarity determining region. Ces 3 régions sont enchâssées entre des régions cadres FR1, FR2, FR3 et FR4 (Framework) dont la variabilité est moindre.

### **C. Régions Constantes**

Les régions C des chaînes légères ne semblent pas porter de fonction effectrice et ont seulement une fonction structurale pour l'Ig. En revanche les régions constantes des chaînes lourdes déterminent l'isotype de l'immunoglobuline qui définit sa fonction. Selon le type de chaîne lourde qui compose l'Ig, la protéine s'organise en domaines qui se différencient par leur nombre, la longueur de la région charnière ainsi que par le nombre et l'emplacement des ponts disulfures. Les immunoglobulines de types M et E ne possèdent pas de région charnière mais conservent tout de même une certaine flexibilité, en partie permise par la présence d'un domaine constant supplémentaire (Padlan, 1994).

La forme membranaire de chaque Ig est composée d'une région supplémentaire de 25 résidus qui permet l'ancrage à la membrane. La production, par la cellule B, à la fois de la forme sécrétée et de la forme membranaire est réalisée à partir d'un même transcrite qui sera épissé différemment afin de produire les deux formes (**Figure 12**).

Les Ig sécrétées, quant à elles, seront majoritairement sous forme monomérique à l'exception de deux isotypes : l'IgM peut être trouvée sous forme pentamérique (c'est notamment le cas dans le plasma) et l'IgA qui s'associe sous forme dimérique. Ces polymères sont formés par une petite protéine de 15 kDa appelée Chaîne J qui établit des liaisons entre les cystéines présentes à la base de la région Fc. Cette polymérisation a deux fonctions : faciliter le transport à travers les épithéliums et acquérir un composant sécrétoire dans le cas de l'IgA (Hendrickson et al., 1995) et améliorer l'avidité de l'IgM (force de liaison de l'Ig à l'Ag qui est établie par le nombre de liaison et non par la spécificité) (Niles et al., 1995).

Ces diversités au sein des régions constantes permettent de moduler la fonction de l'Immunoglobuline.

#### **II.1.2. Fonctions des Immunoglobulines.**

La fonction de reconnaissance de molécules antigéniques est portée par la région Fab (domaine variable) des Ig. Dans certains cas cette simple reconnaissance suffit, c'est notamment le cas lors de la fixation d'une toxine par l'Ig à la surface d'un virus, empêchant ainsi la liaison virus-cellule hôte. Cependant, dans la majorité des cas, les fonctions effectrices de l'anticorps sont portées par la région



Fc et sont indispensables (Davies and Metzger, 1983). Certaines cellules de l'immunité expriment des récepteurs capables de fixer ces régions. Ainsi une cellule présentant des antigènes à sa surface sera marquée par les anticorps IgG1 et 3 qui seront à leur tour reconnus par des macrophages, des neutrophiles ou des cellules NK facilitant ainsi le processus de phagocytose de ces cellules. Les IgE sont, quant à eux, reconnues par les mastocytes, éosinophiles et basophiles et leur fixation induit la sécrétion d'agents inflammatoires.

Un second rôle consiste à l'activation du système du complément. Ce système est composé d'une trentaine de protéines agissant en cascade pour permettre la dégradation des agents pathogènes de façon directe par attaque membranaire ou en induisant une inflammation et la phagocytose du pathogène.

Chaque isotype porte des caractéristiques propres notamment en terme de localisation et de demi-vie. Les IgA, par exemple, sont sécrétées dans les muqueuses sous forme de dimères complexés au composant sécrétoire et se trouvent dans le plasma sous diverses formes incluant des monomères. Les IgG1, 2 et 3 sont les activateurs du complément qui participent aux réponses anti-bactériennes et virales mais elles sont aussi capables de transfert à travers le placenta. L'IgE est, pour sa part, impliquée dans les mécanismes de réponses aux parasites et dans l'allergie (Schroeder and Cavacini, 2010).

La diversité du répertoire d'immunoglobuline est générée à partir de trois gènes distincts qui subissent des remaniements géniques au cours du développement et de la maturation de la cellule B comme précédemment décrit au chapitre 1. Dans la partie suivante, seront détaillées les structures germinales de ces trois gènes chez la souris.

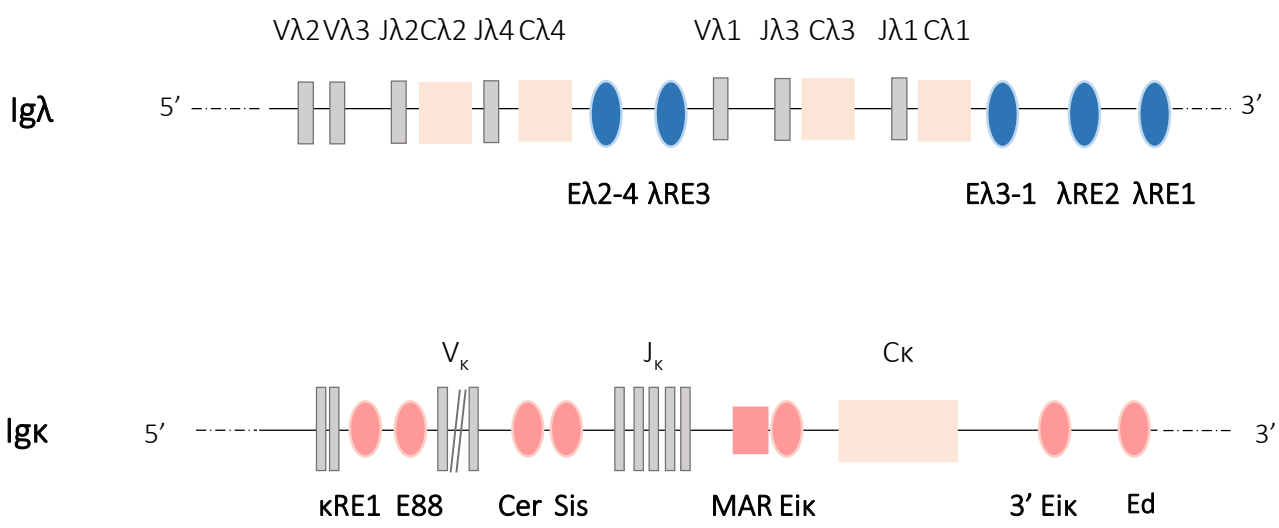
## **II.2. Origine des Immunoglobulines**

Les chaînes légères ont pour origine deux gènes distincts :  $Ig\lambda$  et  $Ig\kappa$ . Les mécanismes de l'exclusion isotypique et allélique permettent à la cellule B de ne générer qu'un seul type de chaîne légère mais la présence de ces deux gènes apporte une diversité supplémentaire au répertoire.

### **II.2.1. Le locus $Ig\lambda$**

Le locus  $Ig\lambda$  (**Figure 1**) est localisé sur le chromosome 16 chez la souris et 22 chez l'homme. Chez l'espèce murine, il occupe 200 kb et se compose de 3 gènes Variable (V), 4 gènes de Jonction (J) codant pour la région variable de la chaîne légère et 4 gènes C codant la région constante.  $Ig\lambda$  s'organise de façon très distinctes des deux autres loci. Ces gènes sont répartis selon deux groupes :

- $V\lambda_2 V\lambda_3 J\lambda_2 C\lambda_2 J\lambda_4 C\lambda_4$
- $V\lambda_1 J\lambda_3 C\lambda_3 J\lambda_1 C\lambda_1$



**Figure 13 : Position des enhancers sur les loci des gènes des chaînes légères.**

Bien souvent, la recombinaison permettant l'expression d'une chaîne légère lambda complète (région variable + région constante) a lieu entre une région variable et le segment J le plus proche soit  $V\lambda_2$  ou  $V\lambda_3$  est recombinaison avec  $J\lambda_2$   $C\lambda_2$  ( $C\lambda_4$  étant non fonctionnel) et  $V\lambda_1$  est associé à  $J\lambda_3$   $C\lambda_3$  (Carson and Wu, 1989; Storb et al., 1989). Les deux groupes de gènes sont séparés par la présence de deux éléments : un enhancer nommé  $E_{\lambda 2-4}$  dont l'activité est optimisée par la présence d'une séquence régulatrice de son côté 3',  $\lambda RE3$ . Trois autres éléments situés en aval du deuxième groupe de gènes (après  $C\lambda_1$ ) participent également à la régulation du locus ; il s'agit de l'enhancer  $E_{\lambda 3-1}$  suivi de deux éléments régulateurs  $\lambda RE2$   $\lambda RE1$  (**Figure 13**) (Hagman et al., 1990; Predeus et al., 2014).

Chez la souris, la grande majorité des cellules B produisent une chaîne légère Kappa. On retrouve seulement 5% de cellules B exprimant une chaîne lambda (et 40% chez l'humain) (Arakawa et al., 1996).

### II.2.2. Le locus Igk

Bien qu'Igk (**Figure 1**), porté par le chromosome 6 chez la souris (2 chez l'homme), ait une fonction redondante de celle du locus lambda, sa composition diffère quelque peu de ce dernier. Étendu sur 3 Mb, il est constitué de 176 gènes  $V_k$  dont 90 sont non fonctionnels, 5 gènes  $J_k$  (1 non fonctionnel) et 1 gène constant, répartis selon cet ordre dans le sens 5' – 3'. Ce locus comporte également des éléments régulateurs (**Figure 13**). Les premiers, du côté 5', sont insérés au milieu des segments V : il s'agit des enhancers  $\kappa RE1$ , décrit comme un activateur spécifique du type cellulaire et du stade du développement, ainsi que de  $E88$  dont la fonction est liée au positionnement du locus dans la structure nucléaire. Ces deux enhancers contribuent, via leurs fonctions respectives à établir la diversité du répertoire (Barajas-Mora et al., 2019; Predeus et al., 2014). Les segments V et J permettent la production de la portion variable de la chaîne kappa et sont séparés par deux autres régions régulatrices présentant des sites d'hypersensibilités à la DNase : la séquence CER (Contracting Element for Recombination) ainsi que l'élément SIS (Silencer in the intervening sequence) qui seraient, tous deux, impliqués dans la conformation et la localisation du locus au sein du noyau (Xiang et al., 2011, 2013). Les gènes  $J_k$  sont, eux aussi, isolés du gène  $C_k$  par un élément intronique, nommé Eik qui exerce une fonction primordiale d'enhancer en favorisant les réarrangements de la chaîne légère  $\kappa$  (Xu et al., 1996). Ce dernier est associé à une région de liaison à la matrice nucléaire MAR (Matrix Attachment Region) d'environ 430 pb située en 5' de Eik (Emorine et al., 1983).

Les MAR sont des régions de 200 à 1000 pb riches en bases A/T et réparties sur divers endroits du génome. Elles sont décrites pour être localisées de préférence en région télomérique, dans des introns proches de sites activateurs de la transcription tels que des enhancers ou des promoteurs et elles

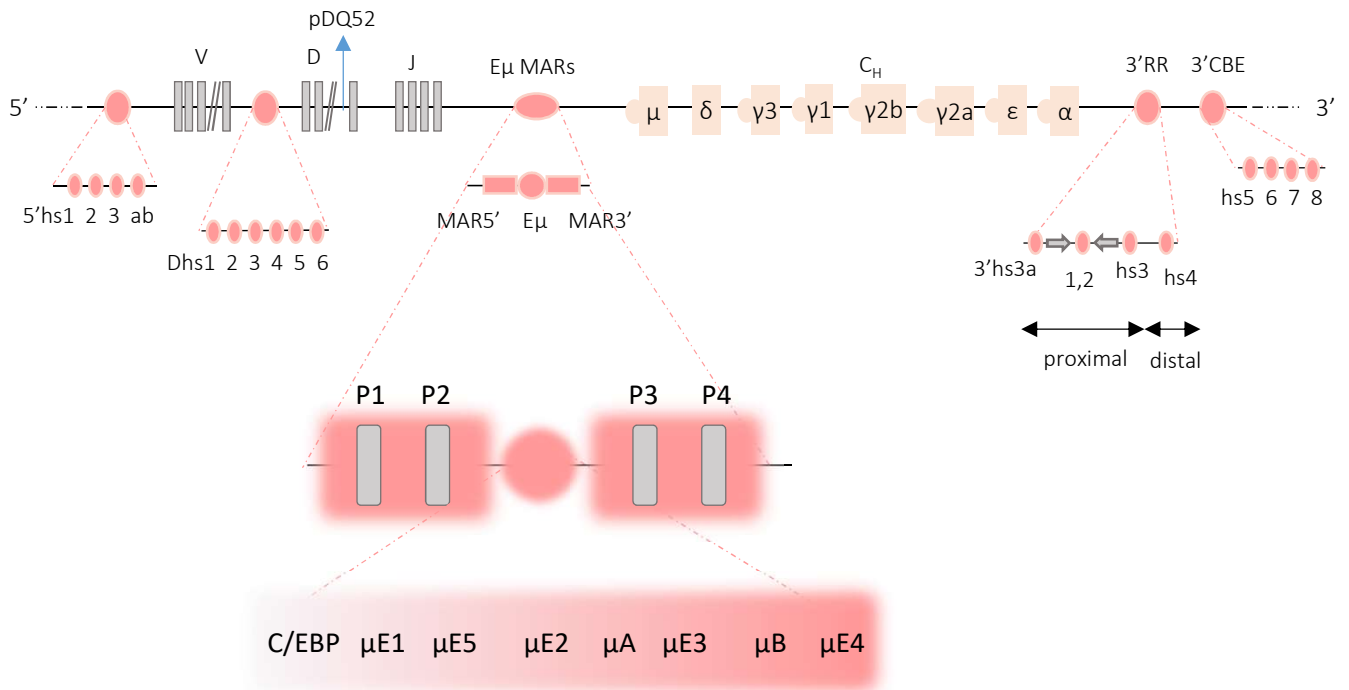


contiennent souvent de nombreuses régions de fixations pour des facteurs modulant l'expression des gènes cibles (Kisseljova et al., 2014). Parmi ces facteurs, on note notamment les topoisomérases dont la fonction de soulagement des torsions imposées à l'ADN lors de la transcription est décrite précédemment. De par leur capacité de liaison aux protéines de la matrice nucléaire, telles que les ribonucléoprotéines, les matrines ou les lamines, on distingue deux types de régions MAR différentes (Hakes and Berezney, 1991; Ludérus et al., 1992; Nakayasu and Berezney, 1991). Les régions MAR attachées de façon constante à la matrice nucléaire sont dites constitutives et sont sensibles à l'action de la DNase dans l'ensemble des types cellulaires. En revanche, les MAR dites facultatives exercent des liaisons non constantes donc dynamiques et réversibles et leur sensibilité à la DNase sera dépendante du type cellulaire et de leur liaison à la matrice (Gluch et al., 2008). Cette fonction de liaison à la matrice nucléaire permet aux MAR de contribuer à l'organisation du noyau en formant des boucles de chromatine au sein de la structure nucléaire via l'attachement des gènes à la matrice.

Dans les années 1990, des modèles murins, transgéniques et de délétion, ont permis de souligner une potentielle implication de la MARk dans les mécanismes de recombinaison VDJ et d'hypermutation somatique. La recombinaison d'un V $\kappa$  avec un J $\kappa$  a lieu de façon anticipée dans un contexte MARk KO et l'on observe aussi une baisse de la fréquence des mutations sur les régions cibles d'AID (Yi et al., 1999).

Deux autres séquences régulatrices clôturent le locus. L'enhancer 3'E $\kappa$ , se situant 9 kb en aval du gène C $\kappa$  est décrit comme sept fois plus fort que E $\kappa$ . Indispensable à la transcription optimale de la chaîne légère k, 3'E $\kappa$  contribue aussi à l'hypermutation somatique (Inlay et al., 2006). Cet enhancer est suivi en 3' par un ultime enhancer, Ed, qui présente des sites de liaisons aux facteurs de transcriptions NF $\kappa$ B et E2A (**Figure 13**) (Liu et al., 2002). Cet élément participe activement à la transcription au stade précoce du développement. Une dernière séquence particulière est à noter. Il s'agit de la séquence RS (recombination sequence) qui peut être sollicitée lors du mécanisme de la recombinaison VDJ. Un réarrangement généré à partir d'une séquence heptamère, située dans l'intron entre les segments J $\kappa$  et C $\kappa$ , et la séquence RS, entraîne la délétion du gène constant et des enhancers 3'E $\kappa$  et Ed. Ce réarrangement, conduisant à l'inactivation du locus Ig $\kappa$ , est retrouvé sur un ou deux allèles dans les cellules B exprimant une chaîne légère  $\lambda$ . Il servirait alors de signal pour initier la recombinaison sur Ig $\lambda$  (Durdik et al., 1984).





**Figure 14 : Positions des enhancers sur IgH.**

Le locus IgH est composé de nombreux enhancers (rouge) qui contribuent à la régulation de son expression et aux divers mécanismes de remaniements des gènes (la flèche bleue représente le promoteur pDQ52 contribuant à la recombinaison VDJ, et les flèches grises représentent les séquences IRIS formant la séquence palindromique). Cette figure reprend également la composition de l'enhancer cEμ et la position des sites de liaisons des MAR BP aux régions MAR.






### II.2.3. Le locus des chaînes lourdes IgH





Le locus IgH (**Figure 1**) est organisé de façon similaire au locus Igk et se localise sur le chromosome 12 chez la souris et 14 chez l'homme. Réparti sur 3 mégabases, il se constitue principalement de 160 gènes V, suivis de 20 gènes de diversité (D), de 4 gènes de jonction puis de 8 gènes codant chacun un isotype formant la portion constante de la chaîne lourde (Giudicelli, 2004; Johnston et al., 2006).

Ce locus, dans sa conformation germinale, présente également des séquences spécifiques permettant sa régulation. Si l'on parcourt le locus du côté 5' vers le côté 3', on retrouve tout d'abord, en amont des premiers gènes V, une région isolatrice appelée Région 5' qui est définie par la présence de 4 sites d'hypersensibilité à la DNase (hs) (**Figure 14**) (Pawlitzyk et al., 2006). Sa délétion n'entraînant pas de biais de développement du lymphocyte B, sa fonction n'est pas clairement définie.

Suivent les gènes V, D puis J qui s'associent pour coder la région variable. Les segments V codent, plus précisément, les portions FR1-CDR1-FR2-CDR2 et FR3 citées en début de chapitre (Lara-Ochoa et al., 1995). Ces gènes V sont classés selon leur position sur le locus, on parlera alors de gènes distaux pour les segments les plus en 5' puis de gènes intermédiaires et proximaux pour ceux en 3' (les plus proches des gènes D). Les segments V sont classés en 15 familles annotées *IgHV1* à 15, la famille la plus représentée étant la première. Parmi les 160 gènes V seulement une centaine sont fonctionnels et chaque V fonctionnel possède son propre promoteur qui se situe 250 pb en amont du gène (Buchanan et al., 1997). Ces promoteurs se composent d'une séquence octamérique ATGCAAAT, hautement conservée, servant de site de liaison pour des facteurs indispensables à la transcription tel que Oct2 ; d'une TATA box, en amont du site d'initiation de la transcription qui est elle-même chevauchée par un élément noté *Inr* qui participe aussi à l'initiation de la transcription (Johnston et al., 2006; Parslow et al., 1984; Roy et al., 2011). Certains promoteurs V (les plus actifs) possèdent des séquences supplémentaires comme les séquences DICE (Downstream Immunoglobulin Control Element) et PAIR (Pax5 Activated Intergenic Repeat) qui permettent le recrutement de facteurs tels que PAX5 ou E2A (Ebert et al., 2011; Tantin et al., 2004). Les promoteurs V participent au contrôle du mécanisme de la recombinaison VDJ via l'initiation de la transcription qui rend accessible ces gènes dans un intervalle de temps bien défini. La production d'une Ig fonctionnelle sera également permise par l'activation de la transcription initiée sur le promoteur du V réarrangé dans les stades cellulaires plus tardifs. En amont de chaque gène V, se trouve des exons de petites tailles (environ 90 nucléotides) annotés L, pour Leader, et qui permettent la production d'un peptide signal. Celui-ci facilitera le passage des précurseurs protéiques des Ig au sein des différents organites de la cellule (lors des étapes de maturation) et sera par la suite clivé par une peptidase.

Les segments V sont séparés des gènes de diversité par la présence d'une région intergénique  $V_H$ -D qui est également composé de 6 sites hs. Cette région permet d'établir un contrôle de la transcription germinale afin d'orchestrer séquentiellement les réarrangements de D vers les segments  $J_H$  puis de  $V_H$

Modèles chimériques	VDJ	SHM	CSR	PUBLICATION
Eμfull KO 	↘	-	-	Sakai et al., 1999
cEμ KO 	↘	-	-	
MAREμ5' KO 	✓	-	-	
MAREμ3' KO 	✓	-	-	
MARsEμ KO 	✓	-	-	

Modèles KO	VDJ	SHM	CSR	PUBLICATION
Eμfull KO 	↘	↘	✓	Marquet et al., 2014
cEμ KO 	↘	✓	✓	Perlot et al., 2005
MAREμ5' cEμ KO 	↘	-	-	Afshar et al., 2006
MARsEμ KO 	✓	↘	✓	Martin et al., soumis (annexe 1)

**Figure 15 : Tableau récapitulatif des modèles murins de délétion de la région EμMAR et les phénotypes associés (✓ normal ↘diminution).**

vers D-J<sub>H</sub>. La région intergénique limite alors la transcription anti-sens établie dans la portion du génome qui inclue les segments D et J ; le but étant d'éviter de rendre accessible les segments V et par conséquent d'éviter une recombinaison prématurée vers un V (Alessandrini and Desiderio, 1991; Guo et al., 2011).

Les gènes D sont répartis en 4 familles (DFL16, DSP2, DST4 et DQ52) et seuls 14 gènes sur 20 sont fonctionnels. La portion CDR3 de la région variable d'une Ig inclue dans sa séquence le segment D et l'ensemble des nucléotides N et P ajoutés de part et d'autre lors de la recombinaison VDJ. Les gènes de diversité sont, eux aussi, précédés par des promoteurs mais ces derniers sont inactifs. Seul le promoteur pDQ52 associé au dernier segment DQ52 a, quant à lui, une fonction indispensable dans l'initiation du mécanisme de la recombinaison VDJ. En effet, pDQ52 permet la production du transcrit stérile  $\mu_0$  qui s'étend de pDQ52 jusqu'à C $\mu$  (Guo et al., 2011; Kottmann et al., 1994) ; la suppression de cet élément induit un biais de recombinaison vers certains segments J. Enfin, les gènes de jonction J, qui sont tous fonctionnels, codent la dernière région cadre ou Framework FR4.

Tout comme sur le locus Igk, les gènes permettant la production de la région variable d'une immunoglobuline sont isolés des gènes codant la portion constante par un enhancer flanqué de deux régions d'attachement à la matrice. Sur le locus IgH, cet enhancer, core enhancer E $\mu$  (cE $\mu$ ), de 220 pb, est encadré par deux régions MAR de 310 pb pour la MAR du coté 5' et de 350 pb pour celle en 3'. Ces régions MAR sont composées chacune de deux sites de fixation pour des protéines regroupées sous le nom de protéine de liaison aux MAR (ou MAR binding protein) qui participent à de nombreux mécanisme cellulaire.

L'enhancer cE $\mu$  rassemble de nombreux sites de fixation, pour diverses protéines, répartis selon cet ordre, dans le sens 5' – 3' : C/EBP,  $\mu$ E1,  $\mu$ E5,  $\mu$ E2,  $\mu$ A,  $\mu$ E3,  $\mu$ B,  $\mu$ E4 et se clôture par une séquence octamérique (**Figure 14**). Ces sites permettent entre autre la liaison de facteurs tels que E2A, PU.1, les protéines de la famille OCT ou encore la protéine ZEB. Cette dernière contribue à établir la spécificité de lignage pour l'expression d'IgH ; elle entraîne la répression du gène en bloquant l'activité de E $\mu$  via sa fixation aux sites  $\mu$ E4 et  $\mu$ E5 dans les cellules non B (Genetta et al., 1994).

L'enhancer E $\mu$  est associé au mécanisme de la recombinaison VDJ et au développement précoce du lymphocyte B (**Figure 15**). Différents modèles ont mis en avant son rôle au cours de l'ontogénèse B. Un premier modèle consistant à l'insertion de l'enhancer dans un transgène a démontré qu'il pouvait initier la recombinaison d'un gène D vers un segment J (Ferrier et al., 1990). Par la suite, son implication dans la mise en place des recombinaisons a été confirmée au sein de modèles où l'enhancer est remplacé par un oligonucléotide ou par une cassette Neo<sup>r</sup>-pGK montrant respectivement une diminution conséquente des réarrangements DJ ou un blocage complet de la recombinaison (Chen et al., 1993; Perlot et al., 2005; Sakai et al., 1999; Serwe and Sablitzky, 1993). Des modèles d'animaux KO ont par la suite montré des altérations des recombinaisons D vers J ainsi qu'un blocage partiel du



développement B au stade pro-B (Perlot et al., 2005). Cette implication dans les réarrangements des gènes de la région variable s'explique par la contribution de l'enhancer dans l'initiation de la transcription germinale qui fournit l'accessibilité des gènes aux enzymes du mécanisme. cE $\mu$  favorise l'émergence des marques actives de la transcription telles que l'acétylation de la lysine 9 et la triméthylation de la lysine 4 de l'histone 3 aux niveaux des gènes D (Chowdhury, 2001). La délétion de l'enhancer seul entraîne une diminution importante des transcrits  $\mu_0$  et I $\mu$  et sa délétion complète (délétion cE $\mu$  et MAR) induit une réduction des transcrits anti-sens initiés aux gènes D (Perlot et al., 2005). Ces constats sont corrélés avec une étude démontrant aussi une diminution du recrutement de l'ARN polymérase II au niveau de ces gènes (Bolland et al., 2004; Chakraborty et al., 2007).

Si le rôle de cette région est bien certifié au sein du mécanisme de la recombinaison VDJ, sa contribution pour la CSR et la SHM n'est pas clairement établie. Les modèles de délétions de cette région ne présentent pas de biais dans la commutation de classe (Perlot et al., 2005). En revanche, les données sur la fréquence de mutation sont controversées. Des modèles d'hybridomes présentant la délétion complète de l'enhancer ne montrent pas de biais dans le mécanisme de SHM. A contrario, la délétion du core E $\mu$ , ou uniquement des MAR, toujours au sein d'hybridomes, entraîne la réduction de la fréquence des mutations (Ronai et al., 2005). Ces données s'opposent à celles obtenues dans les modèles animaux cE $\mu$  KO ou E $\mu$ Full KO (présentant la délétion totale des MAR et de l'enhancer) qui ne présentent aucun biais dans la fréquence d'hypermutation (Li et al., 2010a; Perlot et al., 2005). Cet enhancer serait donc majoritairement indispensable aux étapes précoces du développement de la cellule B. Cependant l'étude de la délétion des deux régions MAR dans un modèle murin, réalisée au sein de notre laboratoire, explique en partie la contribution de ces régions dans le mécanisme de SHM. (Martin et al., 2022) **(Annexe 1)**

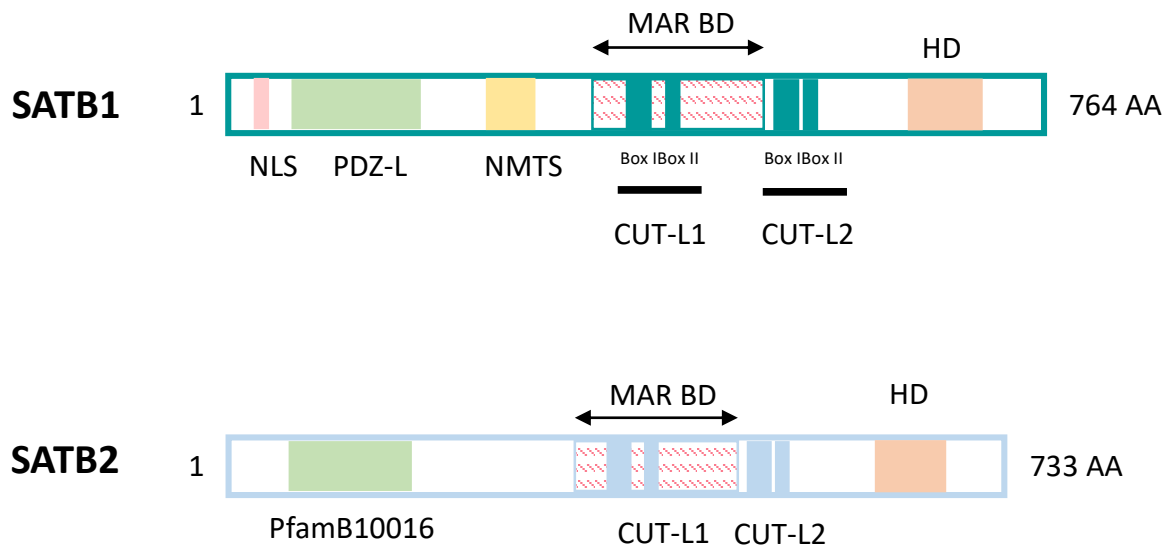
Cette région complexe est suivie par les gènes codant les différentes régions constantes d'une immunoglobuline. Chez la souris, on dénombre 8 gènes répartis selon cet ordre sur le locus :  $\mu$ ,  $\delta$ ,  $\gamma_3$ ,  $\gamma_1$ ,  $\gamma_2b$ ,  $\gamma_2a$ ,  $\epsilon$  et  $\alpha$ . Ces gènes sont composés de plusieurs exons, certains codant un domaine constant de l'immunoglobuline (le dernier exon présente également une séquence S, secretion-coding, codant la portion carboxyterminale contribuant à la sécrétion), et d'autres codant la région charnière (selon les Ig) et d'autres les domaines transmembranaires M1 ou M2 **(Figure 12)** (Johnston et al., 2006). Chaque gène constant, à l'exception de  $\delta$ , est précédé d'un exon et d'un promoteur I « intervening ». La production d'un IgD sera, quant à elle, permise par un épissage alternatif réalisé sur un transcrit long comprenant l'intégralité des régions constantes de C $\mu$  à C $\delta$ . La protéine ZFP138, possédant une structure apparentée à celle de la protéine U1 associée au spliceosome, pourrait contribuer à cet épissage ; en effet, sa délétion au sein d'un modèle murin entraîne une absence de la production d'IgD et une augmentation de la production d'IgM (Enders et al., 2014).



La production simultanée d'une Ig membranaire et d'une Ig sécrétée est autorisée par la présence de deux sites de polyadénylation : pA1, situé avant les exons membranaires, et pA<sub>m</sub> situé après ces exons. La terminaison d'un transcrit au premier site permet de produire une forme sécrétée, tandis qu'une terminaison plus lointaine associée à un épissage du dernier exon codant un domaine d'Ig sur l'exon M1 (entraînant la délétion de la séquence de sécrétion) permet la production de la forme transmembranaire.

En aval du dernier gène constant, se trouvent les deux dernières régions régulatrices du locus IgH : la région 3'RR et la région 3'CBE (**Figure 14**). La 3'RR est la région régulatrice la plus conséquente du locus, elle s'étale sur 30kB et se compose de plusieurs enhancers hs3a – hs1-2 – hs3b et hs4 séparés par des séquences particulières (Lieberson et al., 1991). Une séquence palindromique participe à la structure de la 3'RR, centrée sur l'élément hs1-2 ; elle se compose de séquences répétées et inversées bordées par les enhancers hs3a et hs3b. Ces séquences, nommées IRIS (inverted repeated intervening sequences), forment, associées aux 3 premiers enhancers, la partie proximale de la 3'RR, le dernier enhancer, hs4, étant la partie distale (Chauveau and Cogné, 1996; Madisen and Groudine, 1994). Au sein de cette région, se trouvent également des séquences « like swith » (LS) similaires aux séquences switch qui composent les gènes constants du locus et permettent la commutation de classe. Il est d'ailleurs observé, dans certaines cellules B, un phénomène de recombinaison suicide ou LSR (locus suicid recombination) qui consiste à une commutation de la région switch  $\mu$  vers une séquence LS de la 3'RR, entraînant ainsi la délétion de tous les gènes constants (Péron et al., 2012). De façon globale, les enhancers hs3a hs1-2 et hs3b ne sont pas indispensables aux stades précoces du développement et ont une activité redondante. En revanche, l'enhancer hs4 est actif dès les premiers stades et sa délétion influe sur la production de la chaîne lourde et limite l'expression du BCR. La séquence palindromique joue, quant à elle, un rôle dans le mécanisme d'hypermutation : son absence conduit en effet à une diminution importante de la fréquence de mutation. La délétion complète de la région conduit à une perturbation de l'ensemble des mécanismes cités ci-dessus : une diminution de l'expression de la chaîne lourde et du BCR, une quasi abrogation de la commutation de classe reflétée par une absence de tous les isotypes à l'exception d'IgM ainsi qu'une fréquence d'hypermutation quasi nulle (Bruzeau et al., 2021). Enfin, le locus se clôture par la région 3'CBE composée des sites hs5, hs6, hs7 et hs8 qui a une fonction isolatrice contribuant à l'organisation nucléaire (Zhang et al., 2021). L'ensemble des éléments régulateurs jalonnant les loci des gènes d'Ig permettent un contrôle fin des mécanismes de remaniement et de l'expression des chaînes H et L, par l'intermédiaire de nombreux facteurs protéiques dont les fonctions demeurent à ce jour partiellement décrites. Mes travaux de thèse se sont alors concentrés sur l'étude des fonctions des protéines capables de lier les régions MAR.





**Figure 16 : Représentation schématique des structures protéiques de SATB1 et SATB2.**

NLS : nuclear localization sequence / PDZ-L : domaine semblable au domaine PDZ (domaine de dimérisation) / NMTS : nuclear matrix targeting sequence (domaine d'association à la matrice nucléaire) / CUT-L : domaine semblable au domaine CUT qui permet, avec l'homéodomaine (HD) la liaison à l'ADN.

## PARTIE III. LES PROTEINES DE LIAISON AUX REGIONS MAR

---

Les régions MAR, introduites au chapitre précédent, se présentent sous la forme de véritable plancher d'ancrage pour de nombreuses protéines, regroupées sous le nom de MAR binding proteins. Cette fonction leur confère un rôle soit répresseur soit activateur selon les acteurs recrutés. Parmi ces derniers, on retrouve des protéines dont l'expression, dite tissu spécifique, semble finement régulée au cours du développement B ; c'est le cas pour les protéines des familles SATB et ARID, ainsi que la protéine CUX1. Ce chapitre présente leurs structures, leurs fonctions et les différents processus dans lesquels ces protéines semblent être impliquées.

### III.1. Famille de protéine SATB (Special A-T rich Binding)

#### III.1.1. SATB1

##### A. Structure génomique et protéique de SATB1

*Satb1* est un gène composé de 11 exons, situé sur le chromosome 3 chez l'Homme et 17 chez la souris. Son expression est contrôlée par 4 promoteurs distincts P1, P2, P3 et P4 (Khare et al., 2019; Patta et al., 2020). Ce gène code une protéine de 86 kDa constituée de 764 acides aminés (aa), découverte en 1992 par sa liaison aux régions régulatrices MAR du locus IgH (Dickinson et al., 1992). Cette protéine est capable de se lier sur 3 sites I, II, et III situés dans la MAR 5' de  $E\mu$  ; et 3 sites IV, V et VI dans la MAR 3' de  $E\mu$ , avec une affinité plus accrue pour le site IV. Au sein de ses sites, la fixation s'effectue sur des régions dites A-T riches, plus spécifiquement appelées séquences ATC. Ces séquences sont composées de bases A et T et d'une base C, les bases G étant exclues sur le brin de fixation. Outre la nécessité de cette séquence spécifique pour sa fixation, SATB1 ne peut établir de liaison que dans une région où les brins sont dissociés, comme les régions BUR (base unpairing region). L'ensemble de ces éléments situés dans le sillon mineur de l'ADN en conformation B permet l'interaction physique SATB1-ADN (Dickinson et al., 1992).

Cette fixation très spécifique est sous-jacente à une structure protéique complexe composée d'un domaine de fixation aux MAR de 150 aa, le MAR-binding domain (MAR-BD) (**Figure 16**). Une étude de ce dernier montre qu'il est constitué de deux boîtes. La boîte I de 34 aa se situe en N-terminal du MAR-BD et est suivie de la boîte II de 16 aa en C-term. C'est l'analyse des séquences nucléiques et protéiques chez la souris et chez l'être humain, ayant révélées respectivement 92 et 98% d'homologie, qui a identifié la présence de ces deux boîtes (Nakagomi et al., 1994). Ces régions sont répétées en C-term



de la structure protéique et forment des domaines semblables aux domaines CUT propres à certains facteurs de transcription de la drosophile et des mammifères. Pour SATB1 de souris, ces régions se composent de 73 aa et sont appelées CUT-like domain 1 et 2 (Dickinson et al., 1997). Ces domaines sont constitués de 5 hélices  $\alpha$ , et bien que SATB1 se fixe sur le sillon mineur de l'ADN, la troisième hélice du domaine CUT entre dans le sillon majeur et permet la stabilisation de l'ADN dans un état désapparié (Yamaguchi et al., 2006; Yamasaki et al., 2007).

SATB1 possède également une séquence homologue aux homéodomaines. Constituée de 60 acides aminés localisés en région C-terminal, cette séquence se différencie des homéodomaines classiques par l'insertion d'un acide aminé à la fin de la première hélice, et la substitution d'un motif WFQ, très conservé, par un motif FFQ. Bien que les homéodomaines classiques aient la capacité de lier des séquences d'ADN, celui de SATB1 semble ne pas être suffisant pour sa fixation. En revanche, il augmente l'affinité de celle-ci pour sa séquence cible. En utilisant comme cible la région MAR en 3' de c $\mu$ , la présence de l'homéodomaine accroît jusqu'à 6 fois la liaison de SATB1. L'homéodomaine reconnaît le motif C/A TAATA et la mutation de ses résidus arginine 5 et arginine 3 entraîne une baisse de l'affinité de SATB1 pour sa cible (Dickinson et al., 1997).

SATB1 possède également une séquence de liaison à la matrice nucléaire NMTS (Nuclear Matrix Targeting Sequence). La délétion de sa séquence protéique entre les résidus 224 et 278 induit une baisse de son interaction avec la matrice, altérant ainsi sa fonction de facteur de transcription (Seo et al., 2005).

Sa localisation est donc principalement nucléaire et médiée par la présence d'une région N-term hautement conservée de 20 résidus. La substitution d'une arginine en alanine entraîne une altération de son transport intranucléaire. Toutefois, bien que cette région soit conservée, elle ne constitue pas un NLS (Nuclear Localisation Signal) classique et son transport semble être encadré par un mécanisme différent de celui de l'importine  $\beta$  (Nakayama et al., 2005).

Un dernier domaine est indispensable à la fonctionnalité de la protéine SATB1, le PDZ-like domain, homologue au domaine PDZ. Celui-ci, formé de deux feuillets  $\beta$  antiparallèles, encadrés par deux hélices  $\alpha$ , a la capacité de former des interactions protéine-protéine et permet à SATB1 de former des dimères (Purbey et al., 2008). Une étude lui attribue des caractéristiques communes avec les domaines d'ubiquitines et montre qu'il permet, au-delà de la dimérisation, de former des tétramères (Wang et al., 2012). Le clivage du domaine PDZ-like par la caspase 6 entraîne une perte de la capacité d'oligomérisation et a pour conséquence la dissociation protéine-ADN (Galande et al., 2001).



## **B. Régulation de l'expression et de l'activité de SATB1**

L'expression de SATB1 est spécifique de certains tissus. Découverte par sa fixation à *IgH* dans les lymphocytes B, elle est retrouvée exprimée majoritairement dans les lymphocytes T et plus faiblement dans les testicules et le cerveau (Dickinson et al., 1992). Bien qu'elle soit restreinte à certains tissus, nous verrons par la suite que son domaine d'expression est plus vaste que celui initialement décrit.

Sa liaison à la chromatine induit de multiples effets. Des séquences d'ADN, isolées par leur capacité de liaison à la matrice nucléaire, ont montré que 30 % d'entre elles étaient liées à SATB1 (De Belle et al., 1998). Cette interaction avec la matrice offre à SATB1 la possibilité d'organiser des boucles de chromatine dont l'expression pourrait être finement régulée (De Belle et al., 1998; Galande et al., 2007; Seo et al., 2005). Au sein des cellules T, SATB1 va former des structures semblables à des cages appelées « cage-like » qui encadreront l'hétérochromatine. Une étude récente utilisant la microscopie électronique de super-résolution indique que SATB1 formerait des nanosphères co-localisées à des sites de transcription active (Zelenka et al., 2021b). Ainsi, avec cette conformation, SATB1 est capable d'influencer l'expression de gènes situés jusqu'à 50Kb de son site de fixation (Cai et al., 2003). C'est sa capacité de tétramérisation qui lui permettra de rassembler des gènes plus ou moins distants (Wang et al., 2012).

Cette modulation de l'expression des gènes est exercée via le recrutement de nombreux partenaires, tels que P300 et CUX1 (Cut like homeobox 1), pour lesquels SATB1 sert de protéine d'ancrage à la chromatine (Fujii et al., 2003). Son interaction avec la protéine PML1 (Promyelocytic leukemia protein), lui permet notamment de réguler de façon négative l'expression des gènes du locus *CMH1* (Galande et al., 2007; Kumar P. et al., 2007). Au locus *IL2R $\alpha$* , SATB1 recrute ISW1 et ACF1 (ATP-dépendant chromatin assembly factor 1) qui modifie la chromatine de façon dépendante à l'ATP, afin de réprimer l'expression de l'*ILR $\alpha$*  (Yasui et al., 2002). L'interleukine 2 (*IL2*) est également réprimée par la fixation de SATB1 sur les sites P1 et P2 du promoteur *IL2*, induisant alors le recrutement de l'histone déacétylase 1 (HDAC1) qui maintient le gène dans un état transcriptionnellement inactif (Kumar et al., 2005). La liaison SATB1-HDAC1 est la cible de TAT (protéine rétrovirale transactivatrice) qui, lors d'une infection d'un lymphocyte T par le VIH (virus immunodéficience humaine), entre en compétition avec HDAC1. TAT rompt alors la fixation SATB1-HDAC1, pour se lier elle-même à SATB1. L'acétylation du promoteur, engendrée par ce déplacement de HDAC1, conduit à une expression anormale de l'*IL2*, illustrant ainsi le rôle fondamental de SATB1 comme facteur de transcription (Kumar et al., 2005).

Initialement, SATB1 est décrit comme un facteur inhibiteur de l'expression génique (Kohwi-Shigematsu et al., 1997), mais son rôle n'est pas toujours restreint à l'inhibition. Une analyse de protéines, se fixant à l'élément L2a du promoteur *CD8a*, révèle que la protéine activatrice de la transcription du gène n'est autre que SATB1 (Banan et al., 1997). Il est aussi montré que le renouvellement de l'expression du



corécepteur CD8, permettant la maturation d'un LT double positif CD4<sup>+</sup>CD8<sup>+</sup> à un simple positif CD8<sup>+</sup>, est permise par l'intervention de SATB1 (Nie et al., 2008).

SATB1 s'attache également sur une région en amont du gène codant la protéine C-MYC. Au sein d'une cellule T, l'expression de *c-myc* est drastiquement augmentée après stimulation de la cellule avec un traitement ionomycine/PMA (Phorbol, Myristate, Acetate). Lorsque SATB1 n'est plus exprimée, la stimulation du lymphocyte T n'influe plus sur l'expression de *c-myc*. SATB1 favoriserait sa transcription. Cependant, le taux basal d'expression de *c-myc* est plus important dans un LT non activé et dépourvu de l'expression de SATB1, que dans un LT non activé, sauvage. Dans ce cas, SATB1 participerait à la répression de *c-myc* (Cai et al., 2003). Cet exemple illustre l'ambivalence du rôle de SATB1, qui peut être activateur ou inhibiteur de la transcription pour un seul et même gène cible. Son action est finement régulée et dépendante du contexte environnemental.

Cette régulation s'établit de plusieurs façons. (i) Dans un premier temps par le contrôle de l'expression de son gène qui dépend de 4 promoteurs, dont l'activation s'effectue de façon alternative et selon le type cellulaire (Khare et al., 2019; Patta et al., 2020). L'expression de SATB1 est aussi directement contrôlée par la variation de marques épigénétiques sur son locus. Une lignée cellulaire, établie à partir de cellules pulmonaires cancéreuses, expose une augmentation de l'expression de SATB1, après un traitement composé d'agent déméthylant et d'inhibiteur d'histone déacétylase (Selinger et al., 2011). (ii) Des microARN sont aussi impliqués dans le contrôle de SATB1. La phosphorylation de la protéine STAT5 (Signal transducer and activator of transcription 5), médiée par la kinase JAK3 (Janus Kinase 3), permet son activation et entraîne la production du microARN 155. Celui-ci inhibe l'expression de SATB1 dans un lymphome T cutané (Fredholm et al., 2018). Dans un autre exemple, la régulation négative du microARN 448 par un traitement chimiothérapeutique, entraîne une surexpression de SATB1 dans le cadre de l'émergence d'un cancer du sein (Li et al., 2011).

(iii) Différents acteurs permettent, non pas de contrôler l'expression de SATB1, mais de contrôler son activation. Lorsque la protéine kinase C (PKC) phosphoryle le domaine PDZ-like de SATB1, elle favorise le recrutement de HDAC1 par SATB1 et augmente l'affinité SATB1-ADN. Ce recrutement entraîne une déacétylation de la chromatine sur 10 kb à partir de son site de fixation et permet son maintien dans un état transcriptionnel inactif (Cai et al., 2003; Pavan Kumar et al., 2006). SATB1 agit alors comme un inhibiteur de la transcription. C'est notamment le cas lorsque la protéine est associée à CtBP1 (C-terminal binding protein). Ce complexe induit un effet répresseur qui est optimisé par son association à HDAC1 (Purbey et al., 2009).

En revanche, lorsque la MAR-BP interagit avec l'acétyl transférase PCAF (P300/CBP-associated factor), SATB1 est acétylée sur le résidu K136 de son domaine PDZ-like, entraînant ainsi sa dissociation de l'ADN et la levée de l'inhibition du gène cible (Pavan Kumar et al., 2006).





Les différentes régulations et les nombreuses associations impliquant SATB1, illustrent son statut de protéine à rôles multiples et expliquent son implication dans plusieurs processus moléculaires et cellulaires.

La protéine est, entre autre, impliquée dans un mécanisme de réparation de l'ADN. Les domaines CUT-like de SATB1 lui permettent d'interagir avec OGG1 (8-oxo-guanine glycosylase 1). Cette glycosylase reconnaît les lésions de l'ADN, et clive le N-glycose de la base altérée. Ce clivage permet d'activer la voie BER de réparation de l'ADN. Un knock down de SATB1 montre un retard dans la mise en place du mécanisme de BER, expliqué par la diminution de l'activité de OGG1. A l'inverse, une surexpression de SATB1 est associée à l'augmentation de l'expression de gènes impliqués dans les voies de réparation de l'ADN (Kaur et al., 2016).

En lien étroit avec son rôle dans les mécanismes de réparation de l'ADN, SATB1 promeut aussi un effet anti-apoptotique. Dans les lignées tumorales du sein, une forte expression de la protéine est associée à la diminution de l'activité de caspases et un knock down de SATB1, entraîne une augmentation de l'activité des caspases 3 et 8 (Li et al., 2010b). Cette activité est également modifiée dans les cellules tumorales du foie présentant une forte expression de SATB1. La protéine adaptatrice FADD (Fas Associates protein with Death Domain) participant à la voie d'activation des caspases est, quant à elle, réprimée dans ces mêmes cellules (Tu et al., 2012).

Un modèle contraire prête à SATB1 un effet pro-apoptotique. Les cellules Sézary, qui sont des cellules témoins d'une atteinte leucémique, lors de la mise en place d'un lymphome T cutané, n'expriment pas SATB1. Ce fait est corrélé à une faible expression du récepteur pro-apoptotique Fas-L (Fas-ligand). La surexpression de SATB1 dans ces cellules permet de remettre en place le mécanisme de mort spontanée (Wang et al., 2011). D'autres études attribuent à SATB1 un certain contrôle de l'expression de la protéine pro-apoptotique BCL2 (Gong et al., 2010; Ramakrishnan et al., 2000).

Ces exemples mettent une fois de plus en évidence l'ambivalence du rôle de SATB1 dans les mécanismes de mort cellulaire ; ces différences peuvent notamment s'expliquer par l'environnement moléculaire de la protéine. Une étude réalisée sur des fibroblastes d'embryon murin, met en corrélation l'expression de SATB1 et celle de la protéine p16. Celle-ci participe à la régulation du cycle cellulaire en induisant un blocage potentiel du cycle en phase G0/G1. L'association de l'expression de SATB1 et p16 dans les fibroblastes tendrait vers un arrêt du cycle avec une entrée de la cellule en état de quiescence. En revanche, une forte expression de SATB1, en l'absence de p16, induirait un dysfonctionnement de la mise en place du mécanisme apoptotique et engendrerait une prolifération non contrôlée des cellules (Agrelo et al., 2013). Le rôle de la MAR-BP dans le processus apoptotique semble dépendant des acteurs avec lesquels celle-ci s'associe.

L'implication de SATB1 dans les différents mécanismes moléculaires entraîne des effets à l'échelle cellulaire et sur l'organisme. SATB1 est notamment impliquée dans la différenciation des cellules



souches embryonnaires. Elle favorise l'expression de gènes impliqués dans ce processus et réprime l'expression de ceux permettant l'auto-renouveauement tel que *Nanog* (Savarese et al., 2009).

Un modèle murin de délétion de SATB1 a mis en évidence son rôle au sein de différents tissus. Les souris dépourvues de l'expression de SATB1, sont de tailles plus petites, vivent 3 à 4 semaines et présentent des caractéristiques de déficience neurologique (Alvarez et al., 2000). Une étude sur le développement post-natal des neurones pyramidaux corticaux révèle une fonction importante de la protéine. Ces neurones, chez la souris déficiente pour SATB1, présentent une réduction du nombre d'épines dendritiques. Cet aspect est associé à une altération de l'expression de gènes clés dans la formation de ces épines dont les IEGs (immediate early genes) tel que Fos, et EGR1 (Early Growth Response 1). L'altération du développement des épines peut entraîner une perte de la transmission synaptique et engendrer des dysfonctionnements du développement (Balamotis et al., 2012).

Le phénotype léthal résultant du KO de SATB1 peut être expliqué par ce dysfonctionnement cérébral, mais les souris SATB1<sup>-/-</sup> présentent également une altération du développement des organes liés au système immunitaire. Elles ont une rate, un thymus et des ganglions lymphatiques de petites tailles (Alvarez et al., 2000).

### **C. Implication de SATB1 dans le système immunitaire**

Bien que la protéine SATB1 joue un rôle clé dans le développement cérébral, son rôle est principalement répertorié dans les cellules du système immunitaire.

Les cellules souches hématopoïétiques sont à l'origine de la formation de nombreuses cellules du système immunitaire. SATB1 est impliquée dans leur capacité d'auto-renouveauement. Contrairement aux cellules souches embryonnaires, au sein desquelles le rôle de la protéine est décrit précédemment, SATB1 va favoriser l'auto-renouveauement des cellules CSH. Dans les CSH du foie fœtal issues d'embryon WT et SATB1<sup>-/-</sup>, l'absence de SATB1 induit une augmentation de l'expression de *c-myc* et des gènes de la voie de signalisation NOTCH, impliqués dans la différenciation cellulaire. En mesurant le taux d'expression du facteur NUMB, on peut aussi déterminer si les divisions de ces cellules sont des divisions témoins d'un auto-renouveauement, ou des divisions permettant la différenciation cellulaire. Les cellules SATB1<sup>-/-</sup> montrent une capacité d'auto-renouveauement 6 fois moindre à celle des CSH WT (Will et al., 2013). SATB1 permet de réprimer les gènes de différenciation dans les CSH et de favoriser l'auto-renouveauement de ces cellules.

Une fois de plus, le rôle de SATB1 n'est pas restreint à un seul aspect et nous verrons que cette protéine est tout de même impliquée dans la différenciation des cellules CSH. Celle-ci passe par plusieurs étapes (détaillées au chapitre 1) dont la formation de progéniteurs multipotents lymphoïdes-primaires qui donnent lieu à des progéniteurs précoces spécialisés dans la lignée lymphocytaire (ELP) (Adolfsson et



al., 2005; Lai and Kondo, 2008). Au cours de ces étapes de différenciation, l'expression de SATB1 augmente et ce plus particulièrement dans les progéniteurs spécifiques des lignées lymphocytaires ; cet indicateur plaide en faveur de l'implication de la MAR-BP dans la mise en place d'une différenciation préférentiellement lymphocytaire. Une culture de CSH, en présence de cellules stromales exprimant un ligand de NOTCH, Delta-like 1 (DL1) permet d'induire la différenciation de ces progéniteurs en cellules T. Les cellules CSH, dépourvues de la protéine SATB1, arborent une altération dans leur différenciation en cellules T dans ces conditions et restent bloquées au stade T double négatif CD4<sup>-</sup> CD8<sup>-</sup> (Satoh et al., 2013). Il en est de même lorsque les CSH sont cultivées de façon à promouvoir une différenciation B ou myéloïde : les cellules SATB1<sup>-/-</sup> optent pour une différenciation majoritairement myéloïde et on observe une réduction du nombre de cellules différenciées en lymphocytes B. Des analyses complémentaires dans des modèles de surexpression de SATB1, confirment ces résultats et révèlent une augmentation de l'expression de nombreux gènes liés à la différenciation lymphocytaire (Satoh et al., 2013).

Bien qu'initialement décrit comme favorisant l'auto-renouvellement des CSH, SATB1 apparaît également comme un acteur déterminant dans le choix de différenciation des cellules vers une lignée lymphocytaire.

Néanmoins, son champ d'action ne se limite pas aux lymphocytes. Cette MAR-BP participe aussi à la régulation de gènes impliqués dans la maturation des cellules myéloïdes et des érythrocytes. Au sein des érythrocytes, elle forme un complexe avec la protéine CBP (CREB Binding Protein) et contrôle les marques épigénétiques du locus de la  $\beta$ -globine, favorisant ainsi l'expression de  $\epsilon$ -globine (gène du stade embryonnaire) et réprimant le gène  $\gamma$ -globine (exprimé au stade fœtal) (Wen et al., 2005).

Dans les cellules myéloïdes, son activité s'exerce en association au complexe CBP et p300. Ce dernier augmente l'affinité de liaison entre SATB1 et le gène *gp91<sup>phox</sup>*, qui code une sous-unité du complexe NADPH oxydase. SATB1 semble réprimer l'expression de ce gène (Fujii et al., 2003; Hawkins et al., 2001).

Le large champ d'action de SATB1 est en lien avec son profil d'expression spécifique : SATB1 agit dès le stade de cellule souche en favorisant les lignées lymphocytaires, puis au stade des précurseurs myéloïdes et érythrocytaires, en participant à la répression de gènes dans le but de limiter leur maturation. La protéine est, cependant, plus largement étudiée au sein de la lignée T dans laquelle SATB1 participe au développement, à la sélection, la maturation et la fonction de la cellule.

Le développement du lymphocyte T débute avec une cellule dite double négative CD4<sup>-</sup> CD8<sup>-</sup> (DN) qui subira des étapes de maturation pour devenir double positive CD8<sup>+</sup> CD4<sup>+</sup>(DP). Suite à une interaction avec une molécule du complexe majeur d'histocompatibilité, la cellule optera pour une conservation soit du CD4 (CMH II) et deviendra un T helper (ou auxiliaire) ; soit du CD8 (CMH I) et deviendra alors un T cytotoxique. Le modèle murin établi pour étudier la délétion complète de la protéine SATB1 dans



la lignée germinale décrit une altération du développement des lymphocytes T. Les souris KO présentent un nombre de cellules DN réduit et sont bloquées au stade double positif (Alvarez et al., 2000). Parmi les cellules doubles négatives, on retrouve 4 populations établies selon leur co-expression des marqueurs CD44 et CD25. Entre ces 4 populations, l'expression de la protéine est variable, et plus élevée dans les cellules CD44<sup>-</sup> CD25<sup>-</sup>. Son expression est aussi plus importante dans les cellules DP et CD4<sup>+</sup> immatures (Gottimukkala et al., 2016). Cette variation de l'expression de la protéine, commune à plusieurs lignées du système immunitaire, suggère une implication de SATB1 dans le développement T.

La fonction de SATB1 dans la formation des différentes sous-populations de T n'est pas totalement élucidée, mais l'on sait que sa délétion restreinte au lignage T entraîne une perturbation dans le choix de maturation en cellules T auxiliaires ou cytotoxiques. SATB1 permet l'expression de gènes régulant cette spécialisation, tels que *Thpok*, indispensable à la maturation CD4, ou *Runx3* indispensable à la maturation en CD8. De plus, la MAR-BP est aussi en mesure de réguler directement l'expression de *Cd4* et *Cd8* (Kakugawa et al., 2017). Son importance dans les cellules CD8 a été précédemment évoquée, notamment dans le mécanisme de renouvellement de l'expression du corécepteur CD8 (Nie et al., 2008).

En plus de son implication dans la polarisation CD4/CD8, SATB1 influe également sur la maturation des sous-populations de lymphocytes T helpers Th1 ou Th2. Son expression est engendrée suite à l'activation de la sous-population Th 2 (Cai et al., 2006). Dans ces cellules, ce sont les promoteurs P2 et P3 du gène *Satb1* qui sont activés, permettant ainsi d'exprimer de façon plus importante la protéine, que par l'activation du promoteur P1 (Khare et al., 2019).

Une fois exprimée, SATB1 participe à la formation de boucle de chromatine transcriptionnellement active et permet donc l'expression des cytokines spécifiques de la population Th2 (Ahlfors et al., 2010; Cai et al., 2006). La lignée cellulaire OT II (CD4<sup>+</sup> CD8<sup>-</sup>), cultivée en condition favorisant la maturation en Th2, produit des cytokines liées à une maturation Th1 lorsque les cellules sont dépourvues de SATB1. Alors que cette MAR BP semble favoriser la polarisation en Th2, la délétion de SATB1 restreinte au lignage T entraîne la perte de l'expression de CD40L, marqueur d'une activation Th1 (Kakugawa et al., 2017). Cette nouvelle controverse indique, tout comme pour la polarisation CD4/CD8, que la fonction de SATB1 n'est pas totalement claire pour la maturation Th1/Th2. Outre sa fonction au cours du développement T, SATB1 participe aussi à la fonction de ces cellules.

Tout comme la cellule B, la cellule T a elle aussi recours à la recombinaison VDJ pour réarranger les gènes qui codent son récepteur spécifique : le TCR. Dans les cellules T CD4<sup>+</sup> CD8<sup>+</sup>, dépourvues de l'expression de SATB1, les réarrangements des gènes *Rag1* et *Rag2* sont altérés et leur expression est diminuée. Une région régulatrice, appelée ASE (anti-silencer element), a pour fonction d'induire l'expression des gènes *Rag*, notamment en déjouant un élément silencer qui inhibe cette expression.





SATB1 se lie sur l'ASE ainsi que sur une séquence des promoteurs des deux gènes *Rag* afin d'arranger la chromatine en structure favorable à la transcription. Par la formation de cette structure, SATB1 recrute l'ARN polymérase II, favorisant de surcroît la transcription des gènes. En absence de la MAR-BP, cette structure est modifiée, elle n'intègre plus le gène *Rag2* et limite son expression (Hao et al., 2015).

SATB1 est donc essentielle à l'acquisition de la fonction des lymphocytes T par sa contribution au réarrangement du TCR. Par la suite, la signalisation du TCR engendrera à son tour l'expression de SATB1. Des populations de thymocytes, établies sur leur niveau d'expression du marqueur d'activation CD69, présentent des niveaux d'expression de la MAR-BP différents, l'expression la plus élevée se trouvant dans la population de cellules T activées. L'expression de SATB1 est aussi corrélée à celle de NUR77, une protéine exprimée de façon précoce et en réponse à la signalisation du TCR. De plus, des cellules T cultivées en présence d'un inhibiteur de PI3K, une protéine participant à la voie de signalisation en aval du TCR, présentent une réduction de leur expression de SATB1 (Gottimukkala et al., 2016). L'activation du TCR permettrait d'induire l'expression de SATB1 et ce, de façon dépendante de I/ la voie de signalisation NF- $\kappa$ B (Khare et al., 2019) et de II/ de la protéine GATA3, un facteur de transcription indispensable au développement T (Gottimukkala et al., 2016).

En dehors de ce lien direct avec le TCR, SATB1 participe aussi à la production de nombreux gènes importants pour l'établissement de la fonction des LT dont les molécules du CMH I (Kumar P. et al., 2007), des facteurs régulant l'expression de l'IL4 et IL12 (Ahlfors et al., 2010), et certaines cytokines et récepteurs cytokiniques tels que IL2, IL2R $\alpha$  (Kumar et al., 2005) IL5, IL4 et IL13 (Cai et al., 2006).

SATB1 régule aussi l'expression de la protéine PD1 (Programmed Death protéine 1) dans les cellules T tumorales. La MAR-BP se fixe aux régions régulatrices du promoteur du gène *Pdcd* et recrute le complexe NURD (Nucleosomal Remodeling Deacetylase) afin de réprimer son expression. Le facteur TGF $\beta$  (Tumor Growth Factor  $\beta$ ), quant à lui, inhibe l'expression de SATB1 et restaure celle de PD1 (Stephen et al., 2017). Cette régulation de PD1 suggère que SATB1 favorise l'état actif de la cellule, car la protéine PD1 fixe le ligand PD-L1 présent à la surface de certains types cellulaires, afin d'induire l'inactivation des cellules T dans le but d'éviter l'emballement du système immunitaire (Freeman et al., 2000). La protéine PD1 est un des moyens de contrôle de l'activation des cellules T parmi lesquels on retrouve également les lymphocytes T régulateurs (Treg).

Le rôle de ces cellules est d'établir une surveillance du système immunitaire et d'éviter une activation aberrante pouvant conduire à un phénotype parfois auto-immun. En se liant au super-enhancer des précurseurs des Treg (Treg-SE), la MAR-BP active la transcription de gènes spécifiques des Treg dont le facteur FoxP3 (Kitagawa et al., 2017). Une altération du développement des Treg est observable chez les souris dans lesquelles l'expression de SATB1 est supprimée de façon spécifique dans les cellules hématopoïétiques (Kondo et al., 2016). Lorsque SATB1 est éteint seulement dans les cellules T, la



fonction des Treg est directement perturbée (Kitagawa et al., 2017). Il semble cependant que, comme au sein des précurseurs hématopoïétiques, la fenêtre d'expression de SATB1 soit finement établie. Le taux d'ARN messager de la protéine est plus élevé dans les cellules précurseurs des Treg que dans les Treg eux-mêmes (Kitagawa et al., 2017). L'expression de SATB1 est réprimée une fois le développement Treg engagé, et ce, par le facteur FoxP3 (Beyer et al., 2011).

La dérégulation de l'expression de la protéine dans les précurseurs Treg et les Treg eux-mêmes participerait à la mise en place de maladie auto-immune. SATB1 participe aussi à la régulation du pattern d'expression des gènes des cellules pathogènes Th17 d'un modèle murin, qui est utilisé pour l'étude de la sclérose en plaque (Yasuda et al., 2019). Les souris ayant des cellules hématopoïétiques dépourvues de *SATB1* constitueraient aussi un modèle de choix pour permettre l'étude de la maladie auto-immune du syndrome Sjögren (Tanaka et al., 2017). Ces trois exemples soulignent l'implication potentielle de SATB1 dans les processus auto-immun.

La compréhension de l'implication de SATB1 dans les mécanismes sous-jacents au développement et aux fonctions des cellules du système immunitaire reste complexe, de par ses modalités d'expression variables selon le stade du développement cellulaire ; et par son mode d'action ambivalent selon l'environnement moléculaire.

#### ***D. Lien SATB1-pathologie / développement cancéreux***

L'implication de SATB1 dans l'émergence des cellules cancéreuses fut montrée pour la première fois dans les tumeurs du sein en 2008. Des analyses de lignées cellulaires et de cellules saines et tumorales de patients ont corrélé l'augmentation de l'expression de la protéine avec le caractère agressif de la tumeur. Des knock down de SATB1 ont aussi montré une réduction du potentiel métastatique des cellules tumorales (Han et al., 2008). A l'inverse, dans une lignée non agressive greffée sur des souris immunodéprimés, une surexpression de la protéine favorise la croissance tumorale, la vascularisation, le phénotype indifférencié des cellules, leur entrée dans le système vasculaire ainsi que la formation de nodule pulmonaire (Han et al., 2008). Dans cette étude, les auteurs ont montré la dérégulation de nombreux gènes entre une lignée agressive exprimant SATB1 et la même lignée dépourvue de l'expression de la protéine. Les gènes dérégulés *métastatin*, *VEGF* (Vascular Endothelial Growth Factor), *TGFβ1* (Transforming Growth Factor) et *CTGF* (Connective Tissue Growth Factor), sont tous impliqués dans le mécanisme de formation de métastases ou d'ancrage des cellules (Han et al., 2008). SATB1 participe aussi à établir une résistance aux traitements thérapeutiques (Li et al., 2010b).

Des carcinomes ovariens témoignent aussi du rôle potentiellement oncogène de la protéine SATB1. La classification FIGO (Fédération Internationale de Gynécologie et d'Obstétrique), permet de trier les types de carcinomes ovariens selon le stade de gravité de la maladie. On observe une forte expression



de SATB1 dans les cancers de stades les plus avancés de cette classification. Les carcinomes les plus malins, avec invasions métastatiques, présentent pour la plupart une forte concentration de la protéine au niveau nucléaire (Nodin et al., 2012a; Xi, 2012). Cette corrélation SATB1-promotion métastatique est retrouvée également dans les carcinomes de la prostate (Mao et al., 2013).

En dehors du système reproducteur, SATB1 est également impliquée dans plusieurs tumeurs de la voie digestive. Avec une expression 2 à 7 fois plus importante dans les cellules tumorales, elle est associée à un stade avancé dans le développement de carcinomes du larynx (Zhao et al., 2010).

On l'identifie aussi au sein de tumeurs œsophagiennes, où elle promeut l'expression de gènes nécessaires à la prolifération (Song et al., 2017). Dans les atteintes gastriques, elle est marqueur d'un pronostic vital faible (Tong, 2010), notamment de par la corrélation de son expression et de celle de l'HEPARANASE, une enzyme favorisant la dégradation de la matrice extracellulaire des tissus (Cheng et al., 2010) impliquée dans la formation de métastases (Xiao et al., 2017).

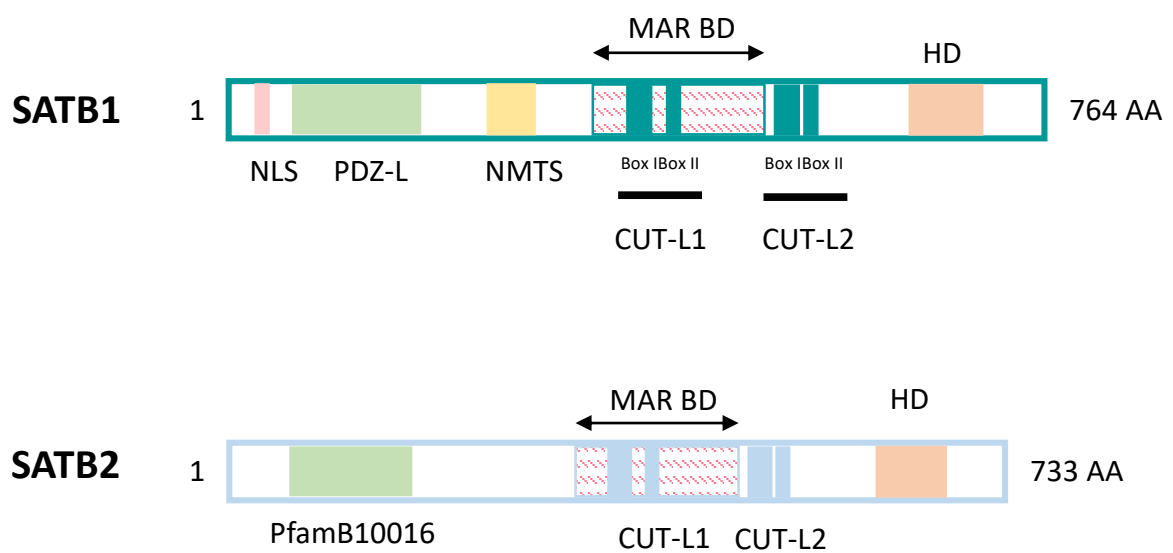
En 2012, la protéine est pour la première fois associée au développement de tumeurs colorectales (Meng et al., 2012), avec invasion métastatique (Zhang et al., 2014a). On soulève un peu plus tard un potentiel rôle antagoniste entre SATB1, qui serait marqueur de pronostic vital faible et SATB2, qui compenserait l'effet néfaste de son homologue (Nodin et al., 2012b). On émet aussi une hypothèse d'une altération des mécanismes post-transcriptionnels régulant l'expression de SATB1 dans les cellules tumorales (Kowalczyk et al., 2015).

Des organes annexes au système digestif sont également concernés par l'implication de SATB1 dans le développement tumoral, comme le foie (Tu et al., 2012) et le rein (Cheng et al., 2014). En revanche, la perte d'expression de SATB1 serait un marqueur de faible pronostic vital dans les tumeurs pulmonaires (Selinger et al., 2011).

Une autre étude, réalisée sur des tissus de patients et une lignée cellulaire, relie SATB1 à l'invasion métastatique, l'inhibition de l'apoptose, l'angiogenèse, et la croissance des cellules tumorales d'un gliome (Chu et al., 2012).

Etant donné la forte implication de la protéine SATB1 dans les mécanismes tumoraux et le système immunitaire, il n'est pas étonnant que cette MAR BP soit associée à certains lymphomes T. Cependant, dans ce cas, son rôle reste très controversé. Si la déméthylation du promoteur de *SATB1*, entraînant sa surexpression, semble être associée au développement d'un lymphome T cutané caractérisé par la présence de cellules CD30<sup>+</sup> (Wang et al., 2014) ; plusieurs études montrent qu'en réalité la fonction de la protéine n'est pas toujours oncogénique. Selon le type de lymphome T étudié, SATB1 n'induirait pas les mêmes effets.

Les lymphomes T cutanés (CTCL : cutaneous T cell lymphoma) comprennent un sous type appelé lymphoproliférations primitives cutanées CD30<sup>+</sup> (LPD) qui regroupe la papulose lymphomatoïde (LyP) et les lymphomes anaplasiques à grande cellule T (LAGC). L'analyse de l'expression de SATB1 dans ces



**Figure 16 : Représentation schématique des structures protéiques de SATB1 et SATB2.**

NLS : nuclear localization sequence / PDZ-L : domaine semblable au domaine PDZ (domaine de dimérisation) / NMTS : nuclear matrix targeting sequence (domaine d'association à la matrice nucléaire) / CUT-L : domaine semblable au domaine CUT qui permet, avec l'homéodomaine (HD) la liaison à l'ADN.

différents sous types montre que la protéine est exprimée dans 91.7% des LyP et 38 % des LACG, LyP étant le lymphome le plus agressif. L'ensemble des cas exprimant SATB1 présentent une polarisation des cellules T en Th17, qui facilite la vascularisation tumorale. En revanche, l'expression de SATB1 est corrélée à une meilleure réponse aux traitements thérapeutiques (Sun et al., 2018).

Un autre type de lymphomes T cutanés, le Syndrome de Sézary, arbore un profil cytokinique Th2 avec des cellules ayant perdu l'expression de SATB1 (Grzanka et al., 2012; Wang et al., 2011). Ce qui souligne ainsi l'hétérogénéité de la fonction de la protéine SATB1 dans la polarisation Th.

Dans un dernier type de lymphome T cutané, les Mycosis fongoïdes, la forte expression de SATB1 est corrélée à celle de MTI//II une métalloprotéine, et le marqueur Ki67, impliqués dans la prolifération cellulaire et la résistance apoptotique (Jankowska-Konsur et al., 2016).

Dans le cadre des pathologies cancéreuses, SATB1 est majoritairement oncogène, mais sa fonction sera dépendante de son niveau d'expression, de l'environnement physiologique, du stade de développement de la maladie et des partenaires associés tels que SATB2.

### III.1.2. SATB2

#### A. *Structure génomique et protéique de SATB2*

La protéine SATB2 partage 61% d'homologie avec SATB1. Elle est identifiée pour la première fois en 2003, tout comme SATB1, par sa capacité de liaison aux régions MAR du locus IgH (Dobrev et al., 2003).

Le gène *SATB2*, de 191 kb, est situé sur le chromosome 2 du génome humain et encode une protéine de 82.5 kDa composée de 733 acides aminés (FitzPatrick, 2003). Elle se compose dans sa partie C-terminale d'un homéodomaine, de deux domaines CUT, et d'un domaine N-terminal Pfam B10016 correspondant au domaine PDZ-like de SATB1 (**Figure 16**). L'ensemble de ces différentes structures est fortement conservé entre les deux protéines (Dobrev et al., 2003; FitzPatrick, 2003).

Néanmoins, on remarque des modifications post-transcriptionnelles différentes, sous-jacentes aux variations de séquences entre les deux protéines. A l'inverse de SATB1, on ne répertorie à ce jour pas de site de phosphorylation et d'acétylation sur la protéine SATB2. En revanche, deux lysines en position 233 et 350 de la protéine SATB2, sont les cibles d'un complexe enzymatique formé par 3 protéines. Ce complexe, constitué en partie par l'enzyme PIAS5, permet de fixer un motif SUMO (Small Ubiquitin related Modifier) aux lysines. Cette sumoylation accroît la régulation de la transcription des gènes cibles de SATB2 par optimisation de sa fixation à l'ADN. Elle semble aussi modifier la localisation intranucléaire de la protéine. Lorsque SATB2 est sumoylée, la protéine se situe en périphérie du noyau ; cette localisation est modifiée si la sumoylation est perturbée (Dobrev et al., 2003).





La sensibilité à la caspase 6 de la protéine SATB1 est une caractéristique qui diffère aussi pour SATB2. Bien que celle-ci conserve le domaine PDZ-like et semble capable de former des dimères, elle ne possède plus le site de clivage de la caspase 6 (Dobrevá et al., 2003).

### **B. Régulation de l'expression et de l'activité de SATB2**

La fonction de SATB2 est associée à sa fixation aux régions MAR, démontrée tout d'abord au locus IgH (Dobrevá et al., 2003), puis au locus IgL (Singh et al., 2020) et enfin sur les MAR de gènes impliqués dans le développement neuronal (Britanova et al., 2005).

Son expression est spécifique de certains tissus. On la retrouve, entre autre, exprimée dans les cellules pré-B, le rein et les tissus cérébraux (Dobrevá et al., 2003). A l'échelle embryonnaire, son expression est déjà hétérogène et différente de l'expression de la protéine SATB1 (Britanova et al., 2005). Peu d'études décrivent la régulation de l'expression du gène *Satb2* mais certaines montrent un contrôle via la production de micro ARN. Les micro ARN 23a et 27a contrôlent l'expression de SATB2 au cours de la différenciation des ostéoblastes. Ils fixent la région 3'UTR de l'ARN messager de *Satb2* pour inhiber sa traduction protéique et bloquer la différenciation des ostéoblastes (Hassan et al., 2010).

Au locus IgH, SATB2 joue le rôle de facteur favorisant la transcription (Dobrevá et al., 2003). Les gènes impliqués dans la formation osseuse tel que *Runx2* sont aussi activés par la présence de SATB2 (Gong et al., 2014). La protéine peut cependant induire des effets contraires et avoir le rôle de répresseur sur certains gènes tels que *Ctip2*, qui code un facteur de transcription impliqué dans la formation de projection neuronale des neurones corticaux (Alcamo et al., 2008).

Pour jouer son rôle de contrôleur de la transcription, SATB2 a recours au recrutement et à l'interaction avec différents partenaires, dont certaines enzymes spécialisées dans la modification de la structure de la chromatine. Dans le mécanisme de la répression de *Ctip2*, la protéine SATB2 se lie sur les régions régulatrices du gène pour induire des modifications dans la structure de la chromatine, telles que la méthylation et l'acétylation (Alcamo et al., 2008). Ces phénomènes sont permis par la capacité d'interaction de SATB2 avec l'histone déacétylase HDAC1, MTA2 (metastasis associated protein 2) et le complexe NURD, tous impliqués dans le remodelage de la chromatine (Gyorgy et al., 2008). Lorsque l'activité de HDAC1 est inhibée, l'effet répresseur de SATB2 est levé (Gyorgy et al., 2008).

Le mode d'action de SATB2, très similaire à celui de SATB1, lui permet d'induire des effets variés sur la cellule en participant, entre autre, à la régulation du cycle cellulaire et de l'apoptose. L'expression de la protéine, au cours des différentes étapes du cycle cellulaire, est modulée et influe sur l'expression de *Cdc25*, *Cdc6* et *Cdk2* (Dowrey et al., 2019). La délétion de SATB2, entraîne aussi la déformation de la membrane nucléaire, processus qui déstructure le noyau lors de l'apoptose (Dowrey et al., 2019).



Ces fonctions attribuées à SATB2 expliquent son implication dans différents processus à l'échelle de l'organisme.

### ***C. Implication de SATB2 dans l'organisme***

Surexprimée dans les lignées de cellules pré-B, SATB2 se lie aux régions MARsEμ et favorise la transcription de transgènes (Dobrev et al., 2003). Une étude plus récente suspecte une implication de la protéine dans le processus de la SHM. La protéine SFSR1-3, un variant de facteur d'épissage participant au mécanisme de l'hypermutation somatique des gènes d'Ig (Kanehiro et al., 2012), promeut l'expression de SATB2 et UBN1, une protéine modulatrice de la chromatine. SFSR1-3 favorise également leur localisation au locus IgL, ainsi que la transcription de ce dernier (Singh et al., 2020).

SATB2 est aussi impliquée dans la calcification des cellules musculaires lisses (Hao et al., 2016) et plus largement dans le développement neuronal. La cartographie de son expression au stade embryonnaire permet de comprendre son rôle dans la différenciation de sous types neuronaux (Britanova et al., 2005), en particulier les neurones corticaux pour lesquels SATB2 régule l'expression de gènes (Szemes et al., 2006).

SATB2 participe également au développement cranio-facial, ainsi que celui de la mâchoire (FitzPatrick, 2003) et à la formation osseuse via la différenciation des ostéoblastes (Dobrev et al., 2006). L'altération de son expression par délétion du gène, ou mutation hétérozygote, conduit à des perturbations du développement (Leoyklang et al., 2007; Zarate et al., 2019).

Le syndrome associé à SATB2 (SAS) est une maladie présentant plusieurs symptômes tels que des retards mentaux, une dysmorphie cranio-facial, une sensibilité épileptique, une faible densité osseuse et une déformation de la mâchoire (Zarate and Fish, 2017; Zarate et al., 2019).

En 2019, une étude sur des patients atteints de la maladie neurodégénérative de Parkinson suspecte également un rôle neuro-protecteur pour la protéine (Guo et al., 2019).

### ***D. Lien SATB2-pathologie / développement cancéreux***

Des études récentes attribuent une fonction oncogénique à SATB2. Dans le cadre d'un cancer buccal, SATB2 favoriserait la croissance cellulaire, l'invasion, la prolifération et la formation de métastase (Ge et al., 2020). Dans les cellules squameuses de carcinome de la tête, la présence de la MAR BP est corrélée à une résistance permet d'établir une résistance à la chimiothérapie (Chung et al., 2010).

A contrario, SATB2 a des effets de suppresseur de tumeur dans le cancer colorectal. L'altération de son expression fait place à un pronostic vital faible (Wang et al., 2009b). Différentes études le réfèrent comme un marqueur biologique de bon pronostic, notamment lorsque son expression est conjointe à



celle d'autres marqueurs, tels que CDX2 dont la perte de l'expression est un marqueur de pronostic engagé (Eberhard et al., 2012; Ma et al., 2019).

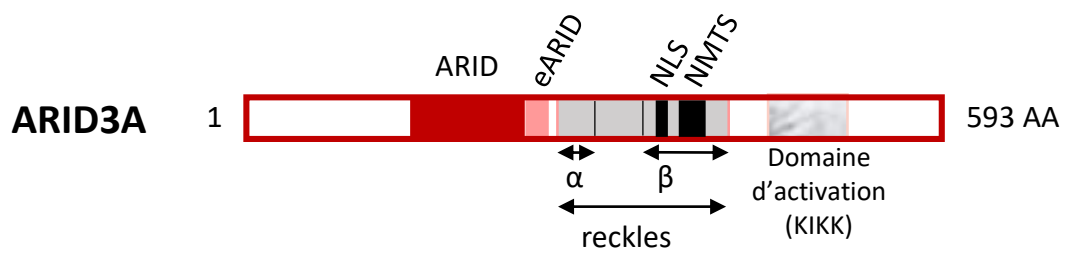
Dans le cancer colorectal, la fonction protectrice de SATB2 contrecarre la fonction pro-tumorale portée par SATB1 (Mansour et al., 2015; Nodin et al., 2012b). Certaines études mettent en évidence un rôle parfois antagoniste entre les deux protéines, l'une favorisant l'expression d'un gène quand l'autre la réprime : c'est le cas pour l'expression du gène *Nanog* dans les cellules souches embryonnaires (Savarese et al., 2009). Dans les cellules érythroïdes, lorsque SATB1 permet l'expression des gènes  $\epsilon$ -globine (Wen et al., 2005), SATB2 se fixe sur les régions MAR des gènes  $\gamma$ -globine pour favoriser leur expression (Zhou et al., 2012). La recherche de partenaires physiques de SATB2 révèle que les deux protéines de la famille SATB ont la capacité de se lier (Zhou et al., 2012). Cette caractéristique soulève différentes questions sur les mécanismes de coopération et de régulation des fonctions des deux protéines.

Contrairement à d'autres MAR BP telles que ARID3A, leurs fonctions multiples et leurs liens avec le système immunitaire, la famille des protéines SATB demeure à ce jour très peu étudiée au cours du développement et de la maturation des lymphocytes B.

### **III.2. Famille de protéine A-T Rich Interaction Domain (ARID)**

Le groupe de protéines « A-T Rich Interaction Domain » est la famille la plus conséquente parmi les protéines de liaison à l'ADN. Elle contient 15 membres répartis en 7 sous-familles : ARID1 à ARID5, JARID1 (Jumonji-ARID1) et JARID2 (Jumonji-ARID2). Parmi ces différentes sous-familles, seules les protéines des groupes ARID3 et ARID5 sont capables de fixer les régions MAR de façon spécifique. Les membres des familles ARID1, 2, 4, JARID1 et 2 réalisent, quant à eux, des liaisons non spécifiques sur l'ADN (Wilsker et al., 2005).

Les ARID3 sont les protéines les plus étudiées chez l'homme. Ce groupe est composé de 3 membres : ARID3A, ARID3B et ARID3C. Ces trois protéines sont hautement conservées chez les mammifères et au sein de diverses espèces comme la drosophile et la levure. Elles partagent 80% d'homologie de séquence au sein de leur domaine ARID et ont la particularité, contrairement aux autres membres de la famille ARID, de posséder un domaine supplémentaire appelé extended-ARID (eARID) ou domaine d'extension ARID qui participerait à la liaison ADN-protéine (Kortschak et al., 1998; Wilsker et al., 2005). Le membre le plus important de cette famille est la protéine ARID3A impliqué au sein de nombreux processus indispensables au développement chez l'être humain et la souris.



**Figure 17 : Représentation schématique de la structure protéique de ARID3A.**

ARID : domaine de liaison à l'ADN / NLS : nuclear localization sequence / NMTS : nuclear matrix targeting sequence (domaine d'association à la matrice nucléaire) / reekles : domaine de localisation et de dimérisation de la protéine.

### III.2.1. ARID3A

#### A. Structure génomique et protéique de ARID3A

La protéine ARID3A de 75 kDa est codée par le gène du même nom composé de 9 exons et situé sur le chromosome 19 chez l'homme et 10 chez la souris. Elle contient 593 acides aminés et se compose de différents domaines : un domaine aminoacide, dont la fonction reste à déterminer, est suivi du domaine ARID et de son extension (eARID) (Kortschak et al., 1998; Nixon et al., 2004a). En position C-terminale de eARID se trouve une région supposée d'activation ainsi qu'un domaine supplémentaire nommé Protein Interaction Domain qui permet la multimérisation d'ARID3A avec différents partenaires. La structure protéique se conclue par une courte portion C-terminale (**Figure 17**).

Les fonctions de ces composantes ont été mises en évidence à travers la production de sept protéines mutantes. Si la modification des domaines aminoacides et de la portion C-terminale de la protéine se révèle peu informative concernant leurs rôles au sein de la structure protéique, les mutations engendrées dans le domaine ARID mettent en évidence sa fonction majeure : chacune des mutations effectuées dans ce domaine induit une perte de liaison à l'ADN (Herrscher et al., 1995). En revanche, ni la délétion de la séquence consensus KIKK du domaine d'activation, ni les mutations créées dans le domaine d'interaction protéique n'entraînent aucun biais de liaison à l'ADN, suggérant que ce rôle serait entièrement dirigé par le domaine ARID. La séquence KIKK située au cœur de la région putative d'activation jouerait un rôle dans la localisation de la protéine. Au sein du domaine d'interaction protéique, il semblerait que la séquence consensus nommée *reckles* porte la capacité de former des multimères (Herrscher et al., 1995). Le domaine REKLES est spécifique à la famille ARID3 et se divise en deux séquences : une région dite *rekles α* séparée de la région *rekles β* par une séquence nommée *rekles spacer*. Les régions  $\alpha$  et  $\beta$  furent initialement décrites comme régulant, respectivement, l'importation et l'exportation de la protéine via un trafic CRM1- exportine (Kim and Tucker, 2006). Par la suite, le segment  $\beta$ , hautement conservé entre espèce, est décrit comme multifonctionnel car il participerait à la dimérisation de la protéine mais également à sa liaison avec la matrice nucléaire. *Rekles β* est composé d'une séquence NLS qui permet le transport et d'une séquence non consensus NMTS qui permet l'interaction d'ARID3A avec les protéines de la matrice nucléaire (Kim et al., 2007).

#### B. Mode d'action et régulation de ARID3A

ARID3A est une protéine découverte par sa participation à la formation d'un complexe capable de se lier aux régions MARsE $\mu$  du locus *IgH* (Webb et al., 1991). L'utilisation de la Distamycine A comme un inhibiteur de liaison protéines - MAR démontre que la fixation d'ARID3A sur ces régions à lieu dans le sillon mineur sur une séquence AT/ATC d'une taille minimum requise de 13 paires de base (Herrscher et al., 1995). La protéine se fixe sur les sites P1, P2, P3 et P4 des régions MARsE $\mu$  (**Figure 14**) sous forme





de tétramère afin de renforcer son effet de modulateur de la transcription. L'absence de l'un de ces 4 sites de fixation induit une diminution de son effet activateur transcriptionnel, suggérant une réponse dose-dépendante de la protéine (Kaplan et al., 2001).

A travers cet ancrage aux MAR, ARID3A joue différents rôles. Elle est notamment capable de moduler l'activation de la transcription de gènes cibles en créant des distorsions d'ADN afin de rendre la chromatine plus accessible (Kaplan et al., 2001; Lin et al., 2007). Cette modulation de la chromatine modifie également la sensibilité à la DNASE I des séquences proches des sites de fixation de ARID3A (Kaplan et al., 2001).

L'activité de la MAR BP est régulée par sa capacité à se lier avec de nombreux partenaires. Elle participe, par exemple, à la formation d'un complexe favorisant la transcription du gène des chaînes lourdes des immunoglobulines (Webb et al., 1991). Au sein de ce complexe, ARID3A est liée à la Bruton's Tyrosine Kinase (Btk) et son substrat : le facteur d'initiation à la transcription I (TFII/I). Btk est composée d'un domaine PH (pleckstrin homologie domain) qui lui permet d'interagir avec différents partenaires dont ARID3A. Cette double interaction, établie par ARID3A, permet i/ d'ancrer le complexe Btk-ARID3A-TFII/I sur l'ADN et ii/ de rapprocher la kinase et son substrat TFII/I qui, une fois phosphorylé, sera capable d'initier la transcription du gène ciblé (Rajaiya et al., 2006; Webb et al., 2000).

ARID3A a aussi la capacité de se lier à ARID3B. Des expériences de co-transfection dans une lignée B démontrent leur interaction physique. Cette liaison inter-ARID3 potentialise l'effet transactivateur d'un transgène rapporteur lorsque ce dernier est contrôlé par un promoteur présentant une région de fixation pour ces deux MAR-BP, il est démontré que la co-transfection des deux ARID augmente la production de la luciférase en comparaison à une simple transfection de l'une ou l'autre des deux protéines (Kurkewich et al., 2016). Il semblerait que ARID3A soit également capable d'interagir avec le troisième membre de la famille ARID3, nommé ARID3C ou Bright-like (Tidwell et al., 2011).

L'interaction physique entre ARID3A et les facteurs E2F1 et Oct2 illustre aussi son implication dans la modulation des boucles de chromatine. ARID3A se fixe au promoteur Cp du virus EBV, contrôlant l'expression des protéines virales dans les cellules B. Via cette fixation et son interaction avec E2F1 et Oct2 qui, eux, se fixent sur l'Origine de réplication virale (OriP), ARID3A permet un rapprochement de régions distantes et autorise la transcription initiée par Cp (Boreström et al., 2012).

Autre exemple de partenariat d'ARID3A : son interaction avec l'autoantigène Sp100 (ou LYSp100). Cette interaction se fait au sein des corps nucléaires PML (promyelocytic leukemia), qui sont des structures auto-assemblées, associées à la matrice nucléaire et qui participent à l'intégrité du génome. Le rôle d'ARID3A dans ces structures n'est pas entièrement définie bien qu'elle semble tout de même contribuer à leur dissociation (Fukuyo et al., 2004; Zong et al., 2000). A travers cette fixation à Sp100, la MAR BP stabiliserait l'intégrité protéique de P53 ; la coopération entre ARID3A et P53 provoquerait



l'arrêt de la croissance cellulaire à la suite de dommages sur l'ADN induits par les UV ou la doxorubicine. (Lestari et al., 2012).

Par ailleurs, à la suite de l'induction de ces dommages à l'ADN, p53 favorise la transcription de *Arid3a* en se fixant sur le deuxième intron de son gène (Ma et al., 2003). L'expression de *Arid3a* dans les CSH est également induite par la voie des TLR, tels que les TLR9, 4, 3 et 7 (Ratliff et al., 2020). Cet effet est également montré dans les lymphocytes B par l'activation du TLR9 via le CpG (Ward et al., 2016). Le gène *Arid3a* fait aussi partie des gènes connus, répertoriés pour être déméthylés au cours du vieillissement ce qui suppose une expression évoluant avec l'âge de l'individu (Acevedo et al., 2015). Le contrôle d'ARID3A s'effectue également au niveau de la traduction, notamment via l'axe Lin28b-let7. Cet axe est particulièrement étudié dans le cadre de la génération de cellule B1a par le foie fœtal et les précurseurs B chez l'adulte. Le facteur Lin28b induit l'expression d'*Arid3a* et inhibe, en parallèle, l'expression du micro ARN let7 qui a pour fonction d'induire la dégradation de l'ARN messager de *Arid3A*. Des modèles murins ou cellulaires de surexpression de Lin28b ou let7 démontrent le contrôle de la production d'ARID3A via cet axe (Hayakawa et al., 2019; Zhou et al., 2015). D'autres micro ARN ont la capacité de se lier à l'ARN de la MAR BP afin d'inhiber sa traduction, et participent à la régulation de l'expression de ARID3A, c'est le cas pour mir125b et mir30 dans les lymphocytes B (Kassambara et al., 2017; Puissegur et al., 2012), ainsi que pour mir574-5 qui favorise la différenciation des fibroblastes (Cui et al., 2020).

Outre son expression, les modifications post-traductionnelles semblent jouer un rôle important pour les fonctions de la protéine. Bien qu'elle soit décrite comme un facteur nucléaire, des études visant à comprendre la fonctionnalité de chaque domaine structurant la protéine révèlent qu'une portion non négligeable de sa production demeure intra-cytoplasmique (Nixon et al., 2004a). En effet, entre un à dix pourcent de la production d'ARID3A, dans un lymphocyte B mature, se trouve palmitoylé. Cette modification permet un transport dans les rafts lipidiques de la membrane plasmique où ARID3A participe à la signalisation du BCR. Proche de ce dernier, elle interagit avec des enzymes de sumoylation telles que SUMO1, UBC9 et PIAS1. Cette interaction entraîne une sumoylation d'ARID3A qui a pour effet son éviction en dehors des rafts lipidiques et une potentialisation de la signalisation du BCR (Schmidt et al., 2009). La fonction de limitation des signaux du BCR par ARID3A dans les rafts lipidiques est un exemple de sa contribution dans le fonctionnement de l'immunité.

### ***C. Implication de ARID3A dans le système immunitaire***

La contribution d'ARID3A dans l'immunité débute dès la mise en place du système immunitaire, au sein des cellules souches hématopoïétiques. En 2010, Carroll Webb démontre l'importance de la protéine pour le développement de l'organisme par la création d'un modèle murin ARID3A KO. La délétion de la

Modèle murin ARID3A	Phénotype	Publication
KO germinal	Létalité importante au stade embryonnaire E13,5 Altération de la différenciation des CSH	Webb et al., 2011
Dominant négatif exprimé par le promoteur humain <i>CD19</i>	↘ IgM ↗ B1 totaux du péritoine	Nixon et al., 2008
Transgène sous le contrôle du promoteur <i>CD19</i> humain	↗ MZ ↘ FO ↗ B1a Production auto-anticorps	Oldham et al., 2011
KO conditionnel dès les stades précoces du lignage B par Cd79a-Cre recombinase	↗ populations B de la moelle ↗ transitionnels, FO et MZ ↘ Ig sérique et Ig en réponse à PC-KLH	Habir et al., 2017
Transgène sous le contrôle de <i>Rag2</i>	↗ B1a dans le péritoine	Hayakawa et al., 2019
KO conditionnel dans la moelle osseuse adulte par Cd2-Cre recombinase	↘ B1a dans le péritoine	Hayakawa et al., 2019

Figure 18 : Tableau récapitulatif des modèles murins de délétion de la protéine ARID3A et les conséquences sur le développement B.

protéine engendre un phénotype léthal : les embryons, pour la majorité, ne survivent pas au-delà du stade E13.5 et présentent une réduction du nombre d'érythrocytes et des progéniteurs lymphoïdes et myéloïdes. L'absence de ARID3A perturberait la capacité de différenciation des CSH (Webb et al., 2011). Par la suite, des expériences de knock down et de surexpression d'ARID3A confirment son implication dans la différenciation des cellules souches hématopoïétiques (issues de cordon ombilical humain) qui présentent une forte expression de la protéine. Sa délétion entraîne une perte de la différenciation en cellules B, tandis que sa surexpression favorise une différenciation myéloïde suggérant ainsi un rôle fonctionnel dose-dépendant dans la mise en place de la différenciation cellulaire (Ratliff et al., 2016). Cependant, dans les cellules souches hématopoïétiques de l'adulte, l'expression d'ARID3A est plus faible, et son rôle semble moins décisif en ce qui concerne le choix de lignage des cellules (Ratliff et al., 2020). Cette implication de la MAR BP dans la différenciation des CSH est corrélée à l'expression de l'interféron alpha (IFN $\alpha$ ). La stimulation des cellules par l'activation des TLR via le CpG, entraîne de façon corrélée, l'augmentation de l'expression d'ARID3A ainsi que celle de l'IFN $\alpha$ . En revanche, lorsque l'expression d'ARID3A est inhibée, la stimulation des cellules ne module plus la production d'IFN $\alpha$  (Ratliff et al., 2020). Cette corrélation établie entre la MAR BP et l'IFN $\alpha$  est également retrouvée dans les neutrophiles et les cellules dendritiques plasmocytaires, ainsi qu'au sein des lymphocytes B (Ratliff et al., 2019; Ward et al., 2016).

Initialement, ARID3A fut décrite comme étant exprimée seulement par les populations B issues de la moelle osseuse (Nixon et al., 2004b; Webb et al., 1998). Plus tard, le pattern d'expression de la protéine sera affiné : ARID3A est finement régulée au cours du développement B avec une expression accrue au sein des cellules immatures de la moelle osseuse, puis diminuée dans les B de la zone marginale et les folliculaires, et de nouveau importante dans des cellules du centre germinatif et les plasmocytes (Habir et al., 2017; Nixon et al., 2004b). Cette modulation fine de l'expression d'ARID3A est sous-jacente à un rôle important de la protéine au cours du développement B.

A travers un modèle murin exprimant un dominant négatif de la protéine, il est montré que ARID3A joue un rôle essentiel dans les cellules de la cavité péritonéale car ce modèle entraîne une augmentation du nombre totale de cellules B1 dans ce compartiment (**Figure 18**) (Nixon et al., 2008). Cependant, dans un modèle conditionnel permettant la délétion d'ARID3A dans les cellules B de la moelle osseuse, on observe un blocage de la différenciation en cellule B1a dans le péritoine (Hayakawa et al., 2019). Cette observation est renforcée par des résultats obtenus sur un modèle transgénique de surexpression de la protéine qui présente, quant à lui, une augmentation de la proportion des cellules B1a au sein de la cavité péritonéale (Hayakawa et al., 2019). Il est donc envisageable que la participation d'ARID3A dans la différenciation en cellules B1a soit, entre autre, guidée par son implication dans l'axe Lin28B/let7 qui gouverne la différenciation de ces cellules aux stades fœtale et adulte (Xu et al., 2020; Zhou et al., 2015).



Dans un autre modèle transgénique (plaçant *Arid3a* sous le contrôle du promoteur *Cd19*), l'expression de cette MAR BP, restreinte au lignage B, augmente également la proportion des cellules B de la zone marginale au détriment des cellules B Folliculaires qui, moins nombreuses, présentent de surcroît des caractéristiques de cellules MZ (Oldham et al., 2011).

Les conclusions issues de ce modèle transgénique seront contredites plus tard par l'étude d'un modèle murin de KO conditionnel permettant, via un allèle *Cd79a-cre* (Hobeika et al., 2006), la délétion de ARID3A dès les stades précoces du développement B. Confirmant la nécessité d'ARID3A pour le développement des cellules B1, ce modèle montre aussi que l'ensemble des populations B de la rate, transitionnelles, folliculaires et de la zone marginale sont augmentées de façon significative (Habir et al., 2017). ARID3A favoriserait une différenciation des cellules B en type B1 et limiterait l'expansion des populations B2.

En dehors de sa contribution au développement des cellules B, ARID3A influence également la fonction de ces cellules. Découverte de par sa capacité de fixation aux MARsE $\mu$ , ARID3A, initialement nommée BRIGHT (B cell Regulator of *IgH* transcription), fut d'abord décrite comme un activateur de la transcription du gène des chaînes lourdes des immunoglobulines en réponse à une stimulation induite par l'interleukine 5 (Webb et al., 1991). Cette fonction nécessite cependant un contexte particulier. Plusieurs études mettent en évidence sa capacité d'activer la transcription d'*IgH* au sein de lignées B transfectées avec un vecteur portant le promoteur associé au segment VS107 et les régions MARsE $\mu$ . Il est aussi démontré que la fixation d'ARID3A aux régions MARsE $\mu$  et l'activation d'*IgH* ne sont optimales qu'en présence d'au moins 3 des 4 sites des MARsE $\mu$  fixant ARID3A (Webb et al., 1991). Ces premières études proposent que la fonction activatrice de ARID3A, dépend à la fois de la présence d'un promoteur VH capable de la fixer, et d'un contexte favorisant l'activation de E $\mu$ . Des expériences de transfection du promoteur VS107 et des MARsE $\mu$ , réalisées au sein de cellules T, ne montrent pas d'activation de la transcription d'*IgH* médiée par ARID3A (Herrscher et al., 1995). La MAR BP aurait alors la capacité de lier, d'une part, le promoteur VS107 et, d'autre part, les régions MARsE $\mu$  créant une interaction ARID3A-promoteur-enhancer afin de générer une distorsion de l'ADN et de permettre l'activation du gène (Kaplan et al., 2001). Plus récemment, il est également décrit que la MAR BP se lierait aux régions MARsE $\mu$  de façon dépendante au cycle cellulaire. Sa production, son transport intranucléaire et sa fixation sont augmentées lorsque la cellule est en phase de synthèse de l'ADN ; en revanche, la production d'ARID3A est diminuée en phase de quiescence (Kim et al., 2017).

D'autres facteurs pourraient également expliquer la modulation de la transcription d'*IgH* par ARID3A, notamment sa participation au recrutement du complexe BTK/TFII-I qui, comme décrit précédemment, favorise la phosphorylation de TFII-I et, par conséquent, la transcription d'*IgH* (Rajaiya et al., 2006; Webb et al., 2000).





Le modèle transgénique plaçant *Arid3a* sous le contrôle du promoteur du gène *Cd19* humain appuie les études *in vitro* précédentes et souligne le rôle d'ARID3A quant à la production d'Ig. Ces animaux présentent un taux d'IgM, d'IgG2a et d'IgG2b supérieur à celui des souris contrôles (Shankar et al., 2007). Néanmoins, cette étude confirme la nécessité d'un contexte particulier pour favoriser cette activation d'*IgH* via ARID3A. Des immunisations réalisées, d'une part, avec les antigènes PhosphatidylCholine-Keyhole Limpet Hemocyanine (PC-KLH) (induisant une réponse impliquant majoritairement VS107) et, d'autre part, avec du NitroPhenylacetyl-Chicken Gamma Globuline (NP-CGG) n'induisent pas la même réponse chez les sujets transgéniques. Si l'on observe une augmentation du taux d'IgM-PC (Phosphatidyl Choline), la sécrétion d'IgM-NP (Nitrophenylacetyl) ne semble pas varier, soulignant l'importance du promoteur V quant à l'activation d'*IgH* via ARID3A (Shankar et al., 2007). Les autres modèles murins de délétion conditionnelle ou exprimant un dominant négatif confirment ces données. L'absence d'une protéine fonctionnelle conduit à une diminution du taux d'IgM sérique suite à une immunisation PC-KLH (Habir et al., 2017; Nixon et al., 2008), en revanche, le NP-CGG n'induit pas de variation de ce taux d'Ig (Nixon et al., 2008).

Toutefois, il est intéressant de noter que les souris exprimant un dominant négatif d'ARID3A produisent cependant moins d'Ig sécrétée que d'Ig membranaire contrairement aux souris contrôles (Nixon et al., 2008). L'origine de cette inversion demeure inexplicée et pose la question de l'influence de ARID3A sur des mécanismes annexes tels que l'épissage ou la sécrétion des transcrits des chaînes lourdes d'Ig. (Schmidt et al., 2009).

En somme, l'implication d'ARID3A au sein des cellules B s'établit à différents niveaux. C'est un facteur qui participe au développement des cellules B1 en potentialisant leur différenciation en B1a. Dans les cellules B2, son rôle reste à confirmer : décrit parfois comme un inhibiteur de l'expansion B2 et d'autre fois comme un activateur. Sur le plan fonctionnel, il permet d'amorcer la transcription du locus *IgH* dans un environnement précis, et contribue à la signalisation du BCR.

On le retrouvera également impliqué dans certains contextes pathologiques.

#### **D. Lien ARID3A-pathologie / développement cancéreux**

ARID3A est associée à un phénomène d'auto-immunité. Le modèle murin transgénique, conférant une surexpression, présente une production d'anticorps anti-nucléaire ainsi qu'une protéinurie importante et la présence de dépôt d'immunoglobuline au niveau rénal (Oldham et al., 2011; Shankar et al., 2007). La production d'anticorps anti-nucléaire est une caractéristique du Lupus érythémateux disséminé (LED) dont la forme aggravée se traduit, entre autres, par une surproduction de l'interféron  $\alpha$ . Il semblerait que l'expression de cet interféron soit, en partie, corrélée à la présence d'ARID3A. Le profil d'expression génique des cellules de patients ayant une forte expression d'ARID3A apparaît différent



de celui issu des cellules de patients sains ou exprimant faiblement ARID3A (Ratliff et al., 2019). Une seconde étude démontre qu'une activation par le CpG des cellules de patients porteurs du lupus entraîne l'augmentation de l'expression d'*Arid3a*, une production de IFN $\alpha$ , ainsi que celles de gènes signature d'une réponse inflammatoire à l'IFN $\alpha$  (Ward et al., 2016).

On retrouve la MAR BP également impliquée dans différents processus tumoraux, notamment au sein des lymphocytes B. Chez la souris âgée, les cellules B1a sont sujettes à devenir malignes et prônes à l'émergence d'une leucémie lymphocytaire chronique (LLC). Un modèle murin exprimant l'oncogène T cell leukemia (TCL1) développe de façon spontanée une leucémie chronique. Le croisement de ces souris avec un modèle de délétion d'ARID3A entraîne un blocage dans l'émergence de la maladie suggérant ainsi que ARID3A participerait à la mécanistique de mise en place de la leucémie (Hayakawa et al., 2019). On trouve également *Arid3a* comme gène signature des Lymphomes diffus à grande cellule B de type B activé (Care et al., 2013).

Si dans ces exemples, ARID3A apparaît comme étant plutôt pro-tumorale, d'autres cas lui attribuent un rôle anti-tumoral. Le micro ARN 125b est décrit comme un proto-oncogène transloqué et surexprimé chez les patients atteints de leucémie aiguë lymphoblastique. Une fonction de mir125b consiste en la dégradation d'ARID3A au sein de lignées B leucémiques et de pre-B primaires murins (Puissegur et al., 2012) ; lorsqu'il est surexprimé, mir125b, conduit à l'extinction d'ARID3A et confère dans ces lignées un blocage de la différenciation ainsi qu'une protection contre l'apoptose (Puissegur et al., 2012). Ces données sont corrélées avec une étude démontrant que la délétion de la protéine confère un pouvoir de survie et de plasticité aux cellules (An et al., 2010).

En dehors des processus tumoraux liés à la cellule B, ARID3A est impliquée au sein de mécanismes cancéreux osseux. Elle favorise notamment la production d'un long ARN non codant, TP73 AS1, impliqué dans la formation d'ostéosarcome (Saadat et al., 2020).

Dans les cellules fibroblastiques murines, ARID3A a aussi la capacité de contourner la mise en place de la sénescence cellulaire et d'accroître la rapidité de prolifération des cellules (Peeper et al., 2002).

Enfin, la MAR BP est également soupçonnée de participer au développement de cellules cancéreuses ovariennes. Si dans un cas elle est décrite comme un facteur de bon pronostic (Zhang et al., 2020), dans un autre, elle induirait la dérégulation de nombreux gènes, tels que des gènes de pluripotence ou caractéristiques de cellules souches. Dans ce dernier cas, ARID3A est associée à son homologue ARID3B et il est proposé que les deux MAR BP régulent mutuellement leur expression (Dausinas et al., 2020).



### III.2.2. ARID3B

#### A. Structure génomique et protéique de ARID3B

*Arid3b* est un gène situé sur le chromosome 15 chez l'Homme, 9 chez la souris, se composant pour les deux espèces de 9 exons. Il permet la production d'une protéine de 61 KDa qui partage une forte homologie avec ARID3A (Numata et al., 1999). Elle se compose, en effet des mêmes domaines, à savoir : un domaine N-terminal portant une fonction indéterminée, un domaine de liaison à l'ADN suivi du domaine d'extension (eARID) ainsi qu'une portion C-terminale composée des 3 sous-parties du domaine Rekles :  $\alpha$ ,  $\beta$ , et la région spacer (Kim et al., 2007). Cette configuration lui permet un fonctionnement proche de celui de ARID3A.

#### B. Mode d'action et régulation de ARID3B

ARID3B est une protéine de liaison à l'ADN qui fixe le site consensus TGGGATTACAGG. La majorité des sites de fixation de la protéine se situe au sein de région proche de promoteurs ou d'enhancers (Bobbs et al., 2015). Par cette fixation, ARID3B joue un rôle de facteur de transcription majoritairement activateur des gènes qu'il cible. C'est le cas, par exemple, pour le gène tumeur suppresseur *p53*, dans lequel il est trouvé plusieurs sites putatifs de fixation des protéines ARID. Des études par siRNA ont démontré que ARID3B est non seulement un régulateur de l'expression du gène *p53* lui-même, mais également un initiateur de la transcription des gènes cibles de P53 dont *Puma* et *Pig3* (Pratama et al., 2015). D'autres cibles de la protéine sont répertoriées et décrites comme étant activées par ARID3B, notamment *Cdc2* (Saadat et al., 2021), *Egfr* (Bobbs et al., 2015), *Cmyc* ou *Hmg2a* (Liao et al., 2016).

Une fois fixée sur son site consensus au sein de son gène cible, la fonction d'activateur de la transcription de ARID3B est liée à sa capacité à recruter la Lysine déméthylase spécifique des histones KDM4C. Des expériences de surexpression d'ARID3B au sein de cellules de carcinome épidermoïde montrent une réduction du niveau de triméthylation des histones 3 sur la lysine 9 et la lysine 27 (Liao et al., 2016). Cette déméthylation induite par KDM4C modifie la conformation de la chromatine et permet d'initier la transcription du gène (Liao et al., 2016). L'inhibition pharmacologique de KDM4C confirme cet effet en modulant la transcription des gènes contrôlés par ARID3B (Liao et al., 2020).

Contrairement à ARID3A que l'on trouve à la fois dans le noyau et à la membrane plasmique, ARID3B est entièrement localisée au sein du noyau des cellules. Cette localisation permettrait à ARID3B de maintenir ARID3A au niveau nucléaire (Kim et al., 2007) ; les deux protéines ARID ayant en effet la capacité de former des hétérodimères via leurs domaines rekles ont la capacité de s'hétérodimériser. Des études de co-transfection des deux protéines dans des fibroblastes montrent une augmentation de la localisation nucléaire de ARID3A uniquement lorsque cette dernière est accompagnée avec ARID3B. L'interaction entre les deux protéines se fait via la portion  $\beta$  du domaine rekles permettant de



masquer le domaine d'exportation d'ARID3A et de favoriser sa rétention au noyau (Kim et al., 2007). Si ARID3B permet la rétention d'ARID3A au sein du noyau, son importation nucléaire, quant à elle, est médiée par l'importine 9. La régulation de l'expression de ARID3B est assez similaire à celle d'ARID3A. L'importine 9 ainsi que les deux protéines ARID sont toutes les 3 ciblées par le micro ARN let7 qui réprime leur expression alors que, Lin28b, quant à lui, les favorise. ARID3B est en retour capable de réguler de façon positive l'expression des gènes cibles réprimés par let7 (Liao et al., 2016). Un autre exemple de micro ARN ciblant les deux protéines ARID est le micro ARN 125b. Son expression induit une répression de ARID3B et modifie la capacité de mouvement des cellules, suggérant une potentielle implication de ARID3B dans la mobilité cellulaire (Akhavantabasi et al., 2012).

### **C. Implication de ARID3B dans l'organisme**

ARID3B est initialement découverte dans les cellules testiculaires puis son spectre d'expression a été décrit au niveau placentaire, embryonnaire, et au sein de diverses lignées cellulaires comme les leucocytes et les lymphocytes B (Numata et al., 1999).

Etant donné l'importance d'ARID3A pour le développement des lymphocytes B et la forte homologie entre les deux protéines ARID3, il apparaît normal que la délétion d'*Arid3b*, de façon conditionnelle au sein des lymphocytes B, induise un biais important dans le développement de ces cellules (Kurkewich et al., 2016). Les animaux issus de ce modèle conditionnel présentent une altération dans le nombre de progéniteurs lymphoïdes communs et de cellules pro B au sein de la moelle osseuse ainsi qu'une réduction du nombre de cellules B1b dans le péritoine, accompagnée d'une augmentation de cellules B1a (Kurkewich et al., 2016). Cette contribution au développement B est renforcée par la capacité de la protéine à se fixer sur les régions MAR du locus IgH (Numata et al., 1999). Cependant et contrairement à ARID3A qui se montre être indispensable pour les cellules souches hématopoïétiques et pour la différenciation en lymphocyte, ARID3B ne semble pas être nécessaire au développement des CSH et au choix du lignage B.

La délétion complète de la protéine à l'échelle d'un organisme confère un blocage du développement. Dans un modèle de souris KO, ARID3B est indispensable au développement embryonnaire. Les embryons dépourvus de ARID3B ne survivent pas au-delà du stade 12.5 jours et présentent des anomalies cardiaques, de la formation du crâne et de l'arc branchiale (Takebe et al., 2006; Uribe et al., 2014). Cette implication dans le développement embryonnaire illustre plusieurs fonctions de la protéine exercées au niveau cellulaire. ARID3B contribue en effet aux mécanismes de l'apoptose, de la différenciation, de la prolifération et de la mobilité cellulaire. L'inhibition de son expression bloque, entre autre, l'expression de gènes de différenciation des cardiomyocytes (Uribe et al., 2014), de progression du cycle cellulaire tels que la *Cycline E1* (Saadat et al., 2021), de mobilité des cellules





comme *Lims2* (Casanova et al., 2011; Uribe et al., 2014) et de la régulation de la mort cellulaire notamment via son lien avec *p53* (Pratama et al., 2015). Cette dernière fonction pro-apoptotique est controversée dans les études impliquant ARID3B dans le développement tumoral.

#### **D. Lien ARID3B-pathologie / développement cancéreux**

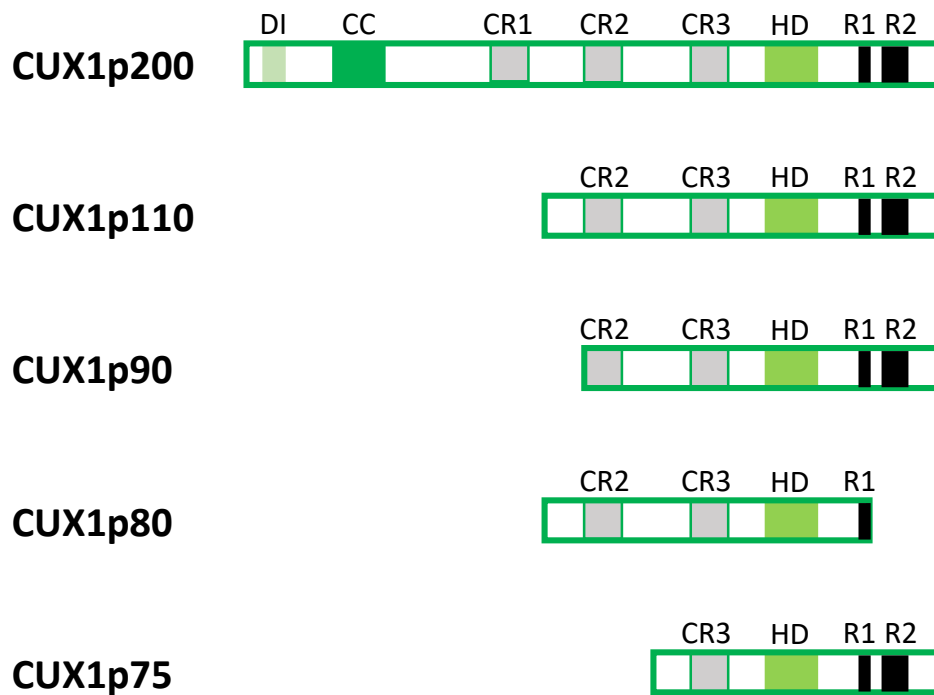
ARID3B est impliqué dans différents processus tumoraux et principalement dans le cancer du sein (Oguz Erdogan et al., 2014), les neuroblastomes (Kobayashi et al., 2013) et les cancers colorectaux (Liao et al., 2020). Son expression est souvent associée à une augmentation de la masse tumorale, d'une prolifération et d'une mobilité cellulaire accrue (Liao et al., 2016).

Ses effets sont majoritairement médiés par la capacité d'ARID3B à promouvoir l'expression de gènes caractéristiques des cellules souches et favorisant la formation de métastases tels que *Lgr5* et les facteurs de la voie NOTCH (Liao et al., 2020; Roy et al., 2014). Dans certains cas, son expression est également un facteur de mauvais pronostic, notamment lorsqu'elle est associée à l'expression de HES1 (Hairy and enhancer of split 1) comme décrit au sein de tumeurs colorectales (Liao et al., 2020). Une étude révèle aussi la présence d'une isoforme d'ARID3B dont le pouvoir oncogénique serait plus important que celui de la forme complète (Joseph et al., 2012).

### **III.3. CUX1 (Cut Like Homeobox 1)**

Parmi les différentes protéines de liaison aux régions d'attachement à la matrice nucléaire, on trouve également Cut-like homeobox 1 (CUX1) qui figure parmi les MAR BP les plus impliquées au sein de divers processus biologiques.

*Cux1* est un gène hautement conservé au sein de nombreuses espèces, situé sur le chromosome 7 chez l'Homme, 5 chez la souris ; il a initialement été découvert chez la drosophile. Composé de 23 à 24 exons selon l'espèce (respectivement Homme – souris), il permet la production d'une protéine de 200 kDa dans sa conformation pleine, nommée CDP (CAAT displacement protein) chez l'Homme et CUX1 P200 pour la protéine murine. Celle-ci se compose de plusieurs domaines : un domaine auto-inhibiteur en région N-terminale est suivi d'une région coiled-coil (CC), de 3 domaines Cut-repeat (CR), d'un Homéodomaine (HD) et se conclut par deux domaines de répression R1 et R2 (Liu et al., 2020; Mailly et al., 1996). Le domaine auto-inhibiteur permet d'inhiber la fonction de la protéine via son interaction avec d'autres domaines, et les domaines Cut-repeat ainsi que l'homéodomaine confèrent la fonction de fixation de l'ADN à CUX1.



**Figure 19 : Représentation schématique de la structure protéique de CUX1 et de ses différentes isoformes.**

DI : domaine inhibiteur / CC : coiled-coil / CR : cut-repeat / HD : homéodomaine / R : domaine répresseur

Plusieurs isoformes de la protéine sont décrites, les mieux étudiées sont nommées selon leurs poids P110, P90, P80 et P75 (**Figure 19**). Les isoformes P110 et P90 sont générées par un processus protéolytique médié par la CATHEPSIN1 qui retire la portion N-terminale de la protéine incluant les domaines inhibiteurs et le CR1. La forme P80 est également générée par la CATHEPSIN1 associée à une caspase qui retire une portion supplémentaire de la région en C-terminale de la protéine, supprimant ainsi le domaine R2 et une portion du domaine R1. L'isoforme P75, quant à elle, est générée à partir d'un transcrit plus court initié au niveau de l'exon 20 du gène. Cette forme se compose essentiellement du dernier domaine CR, de l'HD ainsi que des deux domaines de répression. Il existe une isoforme supplémentaire, P150, peu étudiée et dont la protéolyse n'est pas bien connue. Cette forme possède une portion tronquée (dans la région C-terminale) de son domaine HD qui ne lui permet plus de se lier à l'ADN. L'étude de cette isoforme souligne le rôle indispensable de ce domaine pour la fonction de la protéine (Hulea and Nepveu, 2012; Nepveu, 2001).

En effet, CUX1 est un facteur transcriptionnel fixant de nombreux sites au sein du génome souvent situés dans les régions MAR. CUX1 est également nommée CDP en raison de sa fixation au site CCAAT et de sa capacité de déplacer de la protéine CBP (CCAAT binding protein); au-delà de ce motif canonique, CUX1 se lie aussi aux séquences ATCAAT ou ATCGAT. La fixation de CUX1 sur ces motifs l'autorise à moduler la transcription de gènes à courte comme à longue distance, aussi bien en *cis* qu'en *trans* (Vadnais et al., 2013).

La protéine exerce majoritairement un effet répresseur de la transcription de ces gènes cibles qu'elle induit via différents mécanismes. Dans un premier temps, tout comme les MAR BP des familles ARID et SATB, CUX1 favorise l'organisation de la chromatine en boucles. Dans le cas de la répression du gène *Tcrβ*, l'effet de CUX1 est potentialisé par son association à la protéine SMAR1 (Scaffold/matrix-associated region-binding protein), soulignant ainsi sa capacité à s'associer à des protéines partenaires (Kaul-Ghanekar, 2004). Pour autre exemple, dans les cellules rénales, *via* l'intermédiaire de sa portion C-terminale (Li et al., 1999), CUX1 interagirait avec le facteur répresseur GRG4 et HDAC1 et 3 afin de réprimer le promoteur du gène de l'inhibiteur de la kinase P27 (Livingston et al., 2019).

CUX1 pourrait également réprimer un gène par simple compétition de liaison avec des facteurs activateurs tels que l'histone acétyltransférase GCN5 et le facteur de transcription ATF-1 (Li et al., 1999).

CUX1 fut également décrite comme un facteur activateur de la transcription : des études ont décrit que certains gènes d'histones et de kinases du cycle cellulaire, possédant des sites de fixation pour CUX1, ont une expression corrélée à la liaison de la MAR BP sur l'ADN (van Wijnen et al., 1996). CUX1 est également partenaire du complexe HINF-1 qui est aussi un facteur d'initiation de la transcription de gènes d'histones telles que l'Histone 4 (van Wijnen et al., 1996). Le rôle de CUX1 sur la transcription est donc dépendant, en partie, de l'environnement moléculaire du gène cible. Des expériences de co-



transfection démontrent que la surexpression de CUX1 seule n'a pas d'effet sur l'expression du gène rapporteur portant le promoteur de la tyrosine hydroxylase ; en revanche sa co-transfection avec un activateur de la transcription tel que ITF2 favorisera la transcription du gène cible (Yoon and Chikaraishi, 1994).

Des modifications post-traductionnelles de CUX1 permettent de réguler son activité transcriptionnelle en modulant sa capacité de liaison à l'ADN. La protéine kinase C et la Casein kinase II sont notamment mises à contribution pour phosphoryler CUX1 sur ses domaines CUT ; cette modification empêche toute capacité de fixation à l'ADN (Coqueret et al., 1996, 1998). En revanche, la phosphatase Cdc25A déphosphoryle l'homéodomaine, favorisant, entre autre, la liaison de CUX1 sur le promoteur *p21<sup>WAF</sup>* et induit sa répression (Coqueret, 1998).

La fonction régulatrice ambivalente de CUX1 et son expression ubiquitaire font de cette protéine un acteur pléiotrope majeur. On retrouve notamment cette MAR BP impliquée dans certains mécanismes de réparation des dommages subis par l'ADN. Elle favorise le recrutement et l'activité de la DNA-glycosylase 8-oxoguanine impliquée dans la voie BER (Ramdzan et al., 2015). CUX1 module aussi l'expression des facteurs ATM et ATR, respectivement impliqués dans la réparation de cassure double ou simple brin (Vadnais et al., 2012) et, enfin, elle participe au recrutement de l'histone méthyltransférase EHMT2 (Euchromatic Histone Lysine Methyltransferase 2) qui favorise la rétention d'ATM sur le site de cassure et la réparation de la lésion double brin (Imgruet et al., 2021).

CUX1 participe également à la progression dans le cycle cellulaire en contribuant à la répression des gènes d'inhibiteurs de cycline-kinases tels que P27 et P21 (Coqueret, 1998; Livingston et al., 2019). Le rôle déterminant de CUX1 lors de la réplication est illustré par la dynamique de son activité de liaison à l'ADN au cours des différentes phases du cycle : sa fixation est inexistante en phase de quiescence, elle augmente progressivement durant la phase G1, atteint son maximum en phase S (Coqueret, 1998) puis diminue de nouveau en phase G2 (Santaguida et al., 2001).

Sur un plan physiologique, CUX1 contribue au développement cérébral par sa participation à la formation des épines dendritiques et des synapses (Cubelos et al., 2010) ; elle permet aussi la différenciation des cellules épithéliales pulmonaires et des follicules pileux (Ellis et al., 2001).

Dans le système immunitaire, CUX1 intervient dans les cellules souches hématopoïétiques en limitant leur prolifération et en favorisant leur état de quiescence (An et al., 2018). Par la suite, plusieurs études suggèrent que CUX1 limiterait également l'expansion des cellules myéloïdes et favoriserait la survie des lignées lymphoïdes. Un modèle murin présentant la délétion de l'homéodomaine montre une réduction du nombre de cellules lymphoïdes Thymiques ainsi que des cellules B provenant de la moelle osseuse. Dans les deux types cellulaires, cette réduction serait liée à une augmentation de l'apoptose ; dans la lignée B plus spécifiquement au cours de la progression du stade pro-B au stade pre-B (Sinclair et al., 2001). Il est proposé qu'une répression de CUX1, par sa liaison aux régions MARsEμ du locus IgH,



s'effectue par compétition directe avec la fixation de ARID3A (Kim et al., 2017) ; toutefois, il faut noter que l'expression de la chaîne lourde et par conséquent, celle du BCR, ne semble pas impactée par l'absence de CUX1 (Sinclair et al., 2001). Ce modèle de délétion de l'homéodomaine présente également une hyperplasie myéloïde, suggérant un rôle important de CUX1 dans ce lignage (Sinclair et al., 2001). Cette hypothèse est soutenue par l'apparition de syndrome myélodysplasique, myéloprolifératif ou l'émergence de leucémie aigüe myéloïde lors de la modulation de l'expression de la MAR BP (An et al., 2018). Tout comme SATB1 et ARID3A, CUX1 est capable de fonction ambivalente dans l'émergence de tumeur : parfois oncogène et parfois suppresseur de tumeurs.

Plusieurs études mettent en évidence une fonction protectrice de la protéine. Sa délétion ou des mutations invalidantes sont notamment trouvées au sein de carcinomes colorectaux, pancréatiques, rénaux, pulmonaires ainsi que dans des leucémies myéloïdes et des cancers du sein (Liu et al., 2020). Cependant, de nombreux types de cancer présentent une forte expression de CUX1, suggérant un rôle plus oncogène ; c'est le cas de certains cancers du sein, des mélanomes, des gliomes, de cancers pancréatiques, et de myélomes multiples. Le groupe d'Halong Wu répertorie ces aspects et propose plusieurs explications quant à cette divergence de fonction pour la protéine. Le rôle de CUX1 serait dépendant de l'isoforme exprimée qui, selon le cas, peut modifier la reconnaissance et la dynamique de fixation sur les sites et les possibilités d'interaction avec les acteurs de la réparation des dommages à l'ADN (Liu et al., 2020).







## ***Objectifs***



## Objectifs

---

Le développement d'un lymphocyte B nécessite différentes étapes pour faire en sorte de garder sous contrôle la maturation des gènes d'Ig. Tout au long de son développement, la cellule est donc soumise à des phases de sélection permettant d'éliminer les cellules auto réactives et de sélectionner les cellules les plus spécialisées. C'est un moyen d'établir un premier contrôle à l'échelle cellulaire pour limiter des évènements indésirables tels que la production d'auto anticorps à l'origine de maladies auto immunes. Depuis plusieurs années, de nombreuses études mettent en avant la présence d'autres contrôles s'établissant, cette fois, à l'échelle moléculaire pour encadrer les remaniements géniques dont les débordements conduisent notamment à l'émergence de certains lymphomes. Ces contrôles moléculaires sont permis, en partie, par la composition elle-même des gènes d'Ig. Comme décrit précédemment, des régions régulatrices non codantes jalonnent notamment le locus des chaînes lourdes et permettent d'amorcer les remaniements géniques en recrutant les différents acteurs dans un intervalle de temps et une localisation très précise. Si l'implication de la région 3'RR est bien définie dans les mécanismes de SHM et de commutation de classe, le rôle de la région E $\mu$  et plus particulièrement la fonction des régions MAR qui l'encadrent, notamment au sein du mécanisme d'hypermutation somatique, est plus controversée.

Un modèle murin de délétion des régions MARsE $\mu$ , établi au sein de notre laboratoire, a fait l'objet d'une étude étendue qui a précisé l'implication de ces régions au cours de la SHM. Les MARsE $\mu$  ne semblent pas contribuer au développement des cellules B mais leur délétion perturbe le mécanisme d'hypermutation en cis (sur le locus *IgH*) (**Annexe 1**) (Martin et al., 2022), mais aussi en trans sur les loci des chaînes légères (*Ig $\kappa$* ) et également sur les gènes cibles illégitimes d'AID (*Bcl6*, *CD83*). Le locus des chaînes lourdes, dans ce modèle murin, présente une diminution de la fréquence de mutation dans la région située en amont de l'enhancer cE $\mu$ . En revanche, la région switch  $\mu$ , (également décrite comme légèrement mutée lors du mécanisme d'hypermutation en contexte sauvage) présente, dans le modèle MARsE $\mu$  KO, une fréquence de mutation plus élevée. Ces données dessinent alors les régions MARsE $\mu$  comme des barrières physiques permettant de délimiter partiellement la région ciblée par le processus d'hypermutation. Il semblerait alors cohérent que cette barrière soit établie par l'intermédiaire de l'association des régions MAR à la matrice nucléaire, phénomène dynamique et réversible, qui expliquerait le taux de mutations non négligeable survenant sur la région switch  $\mu$  dans un contexte sauvage. Des données complémentaires de FISH-3D montrent une relocalisation vers le centre du noyau des gènes d'Ig en absence des régions MARsE $\mu$ , soutenant cette hypothèse de barrière physique établie par l'association à la matrice nucléaire et mettant en lien la localisation du gène et son ciblage par AID lors du processus de mutation.



Le croisement du modèle MARsEμ KO avec des modèles de délétion des acteurs principaux des voies de réparation de l'ADN (UNG, MSH) a conduit à des modèles double ou triple mutants précisant le rôle des régions MAR dans le mécanisme d'hypermutation et illustrant son rôle de plateforme de recrutement pour les acteurs des voies de réparation (annexe article 1).

L'implication de ces régions dans ce processus, leurs fonctions de séquence d'arrimage pour diverses protéines et leur ambivalence transcriptionnelle selon les partenaires recrutés posent la question du rôle des MAR BP au sein des mécanismes de remaniements géniques et de leurs contrôles.

Mon travail de thèse se divise alors en deux axes, chacun visant à étudier le rôle d'une MAR BP (SATB1 ou ARID3A) dans les processus physiologiques gouvernant le développement des lymphocytes B. La délétion dès le stade germinatif dans des modèles animaux de l'une ou l'autre de ces deux protéines ne permet pas le développement normal des animaux. Ces délétions entraînent un phénotype létal dès le stade embryonnaire et/ou très tôt après la naissance et, par conséquent, empêchent toute étude chez l'animal adulte. Nos études ont donc été réalisées sur des modèles murins dont la délétion de *Satb1* ou *Arid3a* est restreinte au lignage B via l'expression de la recombinaison CRE placée sous le contrôle du promoteur *Cd79a*.

Le profil d'expression des protéines, variables au cours du développement B, ainsi que des données bibliographiques et préliminaires obtenues dans des modèles semblables suggèrent déjà l'implication de SATB1 et ARID3A dans divers processus propres à la cellule B.

L'expression de SATB1 est plus importante dans les stades précoces du développement avant d'être diminuée dans les stades plus tardifs dès lors que la cellule est activée. L'étude de notre modèle conditionnel a permis de mettre en lien cette variation d'expression et la fonction de la protéine au sein des cellules B. Nos données montrent une ambivalence fonctionnelle de la protéine qui contribue à la régulation des gènes d'immunoglobulines de façon dépendante du stade cellulaire. De plus, en lien avec l'étude du modèle MARsEμ KO, il n'est pas étonnant d'observer une variation de la fréquence d'hypermutation sur les régions cibles du processus sur les gènes d'Ig (article 1). La corrélation entre le contrôle de l'expression des gènes d'Ig et la modulation du processus d'hypermutation suggère, dans notre modèle d'inactivation de SATB1, que le biais du profil mutationnel soit induit uniquement par le contrôle transcriptionnel. Cependant, l'étude de l'hypermutation sur les gènes cibles illégitimes d'AID, tels que *Pim1* et *Bcl6*, montre également, en l'absence de SATB1, une modulation de la fréquence d'hypermutation qui, cette fois-ci, n'est pas corrélée à la transcription de ces oncogènes. Dans ce dernier cas, le mécanisme reliant SATB1 à la SHM de ces gènes s'avère plus complexe et nécessitera d'autres investigations.

A l'opposé de *Satb1*, l'expression d'*Arid3a* est plus importante dans les cellules activées du centre germinatif et les plasmocytes. L'étude de notre modèle confirme l'importance de la protéine pour ces deux populations lymphocytaires en mettant tout d'abord en avant sa contribution dans le mécanisme



d'hypermutation sur les gènes d'Ig. D'autres données ont révélé une fonction inattendue de la protéine dans la différenciation plasmocytaire.

En parallèle de ces deux études, j'ai pu contribuer à l'élaboration de modèles Knock down qui permettront d'étudier l'implication d'autres MAR BP, notamment SATB2 et CUX1, dans le processus d'hypermutation.

Un modèle de délétion de la région MAREik est également en cours de construction pour permettre d'étudier le contrôle de la SHM médié par cette région en Cis (Igκ) mais également en Trans (IgH, gènes AID illégitimes).







# ***Résultats***



## Résultats

---

### **PARTIE I. Etude du rôle de la protéine SATB1 dans le lignage B**

#### **Article 1: A dual function for the chromatin organizer Special A-T rich Binding protein 1 in mature B-lineage cells.**

Morgane Thomas<sup>1</sup>, Ophélie Alyssa Martin<sup>1,2\*</sup>, Charlotte Bruzeau<sup>1\*</sup>, Justine Pollet<sup>1</sup>, Sébastien Bender<sup>1</sup>, Claire Carrion<sup>1</sup>, Sandrine Le Noir<sup>1‡</sup> and Eric Pinaud<sup>1‡</sup>

<sup>1</sup> UMR CNRS7276, Inserm1262, Université de Limoges: Contrôle de la Réponse Immune B et des Lymphoproliférations, 2, rue du Dr. Marcland, 87025 Limoges, France, Team 2, B-NATION : B cell Nuclear Architecture, Immunoglobulin genes and Oncogenes

<sup>2</sup> current address: Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton BN1 9RQ, United Kingdom

\* equal contribution

L'implication complexe des régions MAR $\mu$  dans le mécanisme d'hypermutation somatique soulève la question d'une potentielle contribution des protéines de liaisons de ces régions au sein de ce processus. Le rôle de la protéine SATB1 est à ce jour très peu étudié dans le contexte du développement et des remaniements géniques spécifiques de la cellule B. Son importance pour la formation neuronale, et son rôle au sein des cellules souches hématopoïétiques conduisent à une létalité précoce dans un modèle murin dépourvue de l'expression de la protéine. Un modèle de délétion conditionnelle (cKO) dans le lignage B, établie par l'intermédiaire de l'expression d'une Cre recombinase, contrôlée par le promoteur *Cd79a* nous a permis d'étudier les fonctions de la protéine de façon spécifique au lymphocyte B.

L'analyse des deux phases précoces et tardives du développement B dans ce modèle ne révèle pas de contribution particulière de SATB1 pour la maturation de ces cellules. De plus, et malgré l'implication de  $\mu$  dans les premiers événements de recombinaison VDJ, les souris SATB1 cKO ne présentent pas de biais d'utilisation du répertoire des chaînes lourdes d'immunoglobulines. La production d'anticorps sériques n'est, elle aussi, pas impactée par la délétion de SATB1. En revanche, l'ensemble des populations médullaires et spléniques présente une proportion des IgM membranaires (composant le BCR) corrélée à une altération de la transcription d'*IgH* dans les cellules B matures non activées. Ces données caractérisent SATB1 comme un activateur de la transcription d'*IgH* dans les cellules B naïves et au repos. Cependant dans les cellules B activées des centres germinatifs et dans les plasmablastes, le niveau d'expression de l'IgM membranaire ainsi que la transcription d'*IgH* apparaissent augmentées chez les animaux SATB1 cKO, caractérisant ici, SATB1 comme un répresseur de la chaîne lourde. Bien que la délétion de la protéine induit une modulation de la transcription d'*IgH* dépendante du statut



d'activation de la cellule, elle n'a cependant pas d'effet direct sur la recombinaison de classe analysée dans des essais classiques d'activation *in vitro*. Néanmoins cet impact transcriptionnel conduit à une augmentation de la fréquence d'hypermutation évaluée dans la région switch  $\mu$  et dans l'intron en aval de l'exon VDJ réarrangé. Ces données sont confirmées par une analyse par RNAseq qui montre que la délétion de SATB1 n'influence pas drastiquement l'ensemble du transcriptome de la cellule B mais influe préférentiellement la transcription des gènes d'Ig. L'étude des voies d'activation dérégulées dans les sous populations B de nos modèles suggèrent, en accord avec nos données d'hypermutation somatique, que SATB1 occupe une fonction de répresseur de la maturation de la réponse anticorps.



# A dual function for the chromatin organizer Special A-T rich Binding Protein 1 in B-lineage cells

Morgane Thomas 1, Ophélie Alyssa Martin 1,2\*, Charlotte Bruzeau 1\*, Justine Pollet 1, Sébastien Bender 1, Claire Carrion 1, Sandrine Le Noir 1‡ and Eric Pinaud 1‡

1 UMR CNRS7276, Inserm1262, Université de Limoges: Contrôle de la Réponse Immune B et des Lymphoproliférations, 2, rue du Dr. Marcland, 87025 Limoges, France, Team 2, B-NATION : B cell Nuclear Architecture, Immunoglobulin genes and Oncogenes

2 current address: Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton BN1 9RQ, United Kingdom

\* equal contribution Paste the full affiliation list here.

‡ corresponding authors: Eric Pinaud and Sandrine Le Noir.

**Email:** eric.pinaud@unilim.fr and sandrine.le-noir@unilim

**Author Contributions:** MT, OM, CB, SB and CC performed experiments. MT analyzed the data. EP and SLN conceived and supervised the study. MT and OM developed the experimental model. JP performed the bio-informatic analysis. MT, EP and SLN wrote the manuscript.

**Competing Interest Statement:** The authors declare no competing financial interests.

**Classification:** Biological science / Immunology and Inflammation.

**Keywords:** B cells, nuclear factor, ambivalent, somatic hypermutation

## Abstract

SATB1 (Special A-T rich Binding protein 1) is a cell type specific factor involved in chromatin remodeling events that participate in regulation of the genetic network in developing T cells and neurons. In T cells, SATB1 is a key factor required for lineage commitment, VDJ recombination, development and maturation. In B cells, SATB1 is described as binding to the MARs-E $\mu$  regions of the IgH locus. Considering that its expression varies during differentiation, the involvement of this factor needed to be clarified in B cells. Using a KO mouse model deleting SATB1 from the pro-B cell stage, we were able to examine the consequences of SATB1 deletion in naive and activated B cell subsets. Our model indicates firstly that SATB1 is not essential for B cell development and the establishment of a broad IgH repertoire. Second, we show that this factor exhibits an ambivalent function in mature B cells, acting sequentially as a positive and negative regulator of Ig gene transcription in naive and activated cells, respectively. Third, our study indicates that the negative regulatory function of SATB1 in B cells extends to the germinal center response in which this factor



limits somatic hypermutation of Ig genes. This finding suggests that SATB1 may limit the introduction of unwanted mutations into B cells.

### **Significance Statement**

While SATB1 is essential for the development of T cells in which it participates in chromatin organization, its function during B-lineage cell maturation has been poorly described so far. Based on a conditional knock out mouse model, this study indicates that SATB1 exerts ambivalent regulatory functions on immunoglobulin expression at mature stages. Ig genes are sequentially activated by SATB1 in naive B cells and then repressed in activated cells where it exerts a limiting function on somatic hypermutation.

### **Main Text**

#### **Introduction**

The transcription factor Special A-T rich binding protein 1 (SATB1) is a factor able to bind to Matrix Attachment Regions (MAR) in the nucleus (1). This MAR-binding protein (MAR-BP) is indispensable for T lymphocyte development (2–4) through its implication in properly organizing nuclear architecture, especially chromatin folding (5–7). Murine SATB1, sharing more than 98% identity with its human homologue, can multimerize through an ubiquitin-like domain (8) and interacts with chromatin through its CUT-like and homeodomains respectively involved in DNA binding affinity and specificity (9, 10). One historic feature of SATB1 is its ability to tie nuclear matrix proteins (11, 12). Another essential feature, as a transcription factor, is its capacity to bind DNA for gene regulation. Strikingly, such interactions were found to occur in nucleosome-dense regions, preferentially at AT-rich sequences in the nucleosomal core (13), a feature that strongly supports its candidacy as a pioneer factor. Given its multivalent potency as a nuclear matrix scaffolding factor, a genome organizer involved in chromatin looping and a transcription factor, the exact mechanism linking this MAR-BP to gene regulation remains puzzling. The large body of literature related to SATB1 in T cells suggest that this factor is capable of flexible functions; either binding to euchromatin (5, 6) or nucleosome-dense regions (13), depending on the cell type and development stage assessed. Long distance interactions between promoters and enhancers can be promoted by SATB1 homotetramerization (8). A negative regulatory function for SATB1 was attributed to its capacity to recruit chromatin modifiers such as Histone deacetylase 1 (HDAC1) (14). Indeed, SATB1 is subject to posttranslational modification: phosphorylation of this MAR-BP modifies interactions with chromatin corepressor and coactivator complexes, leading to a switch in its transcriptional activity (15, 16); SATB1 acetylation disrupts its interaction with C-terminal binding protein 1 (CtBP1) (17). Moreover, regulation of *Satb1* gene is by itself tightly regulated in T cells by

interchanging promoter usage, resulting in fine tuning of protein expression (18, 19). These multiple levels of regulation suggest the need for meticulous control of SATB1 expression.

SATB1 deletion models in the mouse led to major alterations in neuronal, hematopoietic and immune systems (2, 20) and pointed out broad and critical functions for this protein in mouse development. SATB1 is a major player in early hematopoiesis since it promotes hematopoietic stem cell self-renewal (HSC) (21). Based on its relative expression level in differentiating HSC (22), SATB1 favors the lymphoid lineage (23). In erythrocyte and myeloid cells, this protein also modulates epigenetic marks by association with CBP (CREB binding protein) to control  $\beta$ -globin and NADPH oxidase gene expression (24–26). There is also growing evidence showing that SATB1 is involved in cancer and auto-immune diseases (6, 7, 27, 28) highlighting the necessity to decipher its regulation in immune cells.

Due to thymus alteration in *Satb1* KO mice (2), SATB1's functions were largely studied in T-cells. Initially described as a repressor factor (12, 29), SATB1 proved to be also a transcriptional activator (6, 30). As an example of ambivalence, SATB1 could repress *c-myc* in resting T-cells while stimulating its expression in activated T cells (3). It is now established that this MAR BP is a major regulator of thymocyte development throughout all differentiation stages from progenitors to regulatory subsets (28, 31, 32) (28)., One notable role of SATB1 in T cells is its activator function for *Rag* gene expression in order to promote VDJ recombination (33).

By contrast to the vast knowledge on SATB1 fine tuning and regulatory functions in T lymphocytes, little is known on the potency of SATB1 to regulate B cell development, beyond HSC fate decision (22, 23). Initial studies reported a discrete defect on B cell numbers in *Satb1* KO mice (2). Taking advantage of its elegant fluorescent reporter model, Yokota group's described fluctuations in *Satb1* expression throughout B cell development with higher expression in naive B cell subsets. In this study authors proposed that SATB1 is involved in BCR-mediated B cell survival (34). Many aspects of the literature point to a potential function for SATB1 during B cell development. Given its regulatory function for *Rag* gene expression in T cells (33), the potential implication of SATB1 in V(D)J recombination of Ig genes in B cells needs to be examined. Moreover, since SATB1 was discovered by its ability to bind the  $MAR_{E\mu}$  region of the *IgH* locus (1), we could postulate a role for SATB1 in Ig heavy chain expression.

In this study, we examined SATB1 implication in B-lineage cell development and depicted its critical function on immunoglobulin production using a conditional KO mouse model allowing deletion of this factor in B-lineage cells. Although we found SATB1 nonessential for B cell development, we showed that this factor displays an ambivalent function in late developing B cell subsets: acting sequentially as a positive and then negative regulator of Ig gene transcription. The negative

regulatory function of SATB1 extends to the germinal center reaction in which this factor limits Ig gene somatic hypermutation.

## **Results and discussion**

### ***Conditional deletion of *Satb1* in murine B lineage cells***

Since depletion of *Satb1* in ES cells and mouse embryos showed multiple dysfunctions and did not allow animals to survive beyond 1 month of age (2), we evaluated SATB1 contribution to B cell development in a conditional knock out model inducing SATB1 depletion in B-lineage cells from the pro-B cell stage. In our model, initially derived from *Satb1<sup>tm1a</sup>* allele, *Satb1* conditional allele (*Satb1<sup>flx</sup>*) contains exon 4 flanked with two *LoxP* sites that permit its specific deletion when coupled to one allele carrying a CRE recombinase insertion into *Cd79a* gene (*Cd79a<sup>cre</sup>*) (35) (Fig. 1A). Mice carrying the homozygous SATB1 deletion in B-lineage cells (*Satb1<sup>flx/flx</sup> Cd79a<sup>cre/+</sup>*, referred as cKO conditional Knock Out) were compared to heterozygous littermates (*Satb1<sup>flx/+</sup> Cd79a<sup>cre/+</sup>*, referred as cHet: conditional Heterozygous) and to *wt* littermates carrying an identical conditional allele but devoid of CRE expression (*Satb1<sup>flx/+</sup> Cd79a<sup>+/+</sup>*, referred as WT). Western blot experiments performed on sorted splenic B cells from these three genotypes confirmed that *Satb1*-exon4 deletion induced a complete protein depletion (Fig. 1B).

### ***SATB1 depletion in B lineage cells allows normal B cell development and IgH repertoire***

By questioning ImmGen database (36), we examined *Satb1* expression in developing B-lineage cells and noted that all subtypes of B-lineage cells expressed *Satb1* transcripts (*SI Appendix*, Fig. S1A). Common lymphoid progenitor and pre-pro B cells, also mentioned as Hardy fraction A, both displayed higher SATB1 expression. This was not surprising given that SATB1 has been reported to favor lymphocyte lineage differentiation from hematopoietic stem cells (23). While a consistent drop in *Satb1* transcription was observed at pro- and pre-B stages, a second wave of *Satb1* gene expression occurred in transitional and mature resting B cells present in splenic follicles or the marginal zone as well as in splenic memory B cells. *Satb1* gene expression further decreased in antigen-activated cells such as germinal center centrocytes to reach its minimal level in proliferating cell subsets such as centroblasts and plasmablasts (*SI Appendix*, Fig. S1A). In B cells from WT mice, we experimentally confirmed a significant decrease in *Satb1* transcripts by RT-qPCR assays when comparing naive and GC B cells sorted from Peyer's patches (Fig. 1C). The same findings, corroborating expression profiles described in the ImmGen database, were very recently documented in a mouse model carrying a Tomato-reporter transgene knocked into a *Satb1* allele

(34). This converging evidence for variable expression of *Satb1*, according to B cell subsets, supposes a narrow regulation and suggests an accurate function of this factor in the B cell lineage. To determine SATB1 contribution throughout B cell development and maturation, we analyzed B cell populations from WT, cHet and cKO mice by flow cytometry. Absolute numbers of B-lineage cell subsets from bone marrow (Fig. 1D left and middle) and spleen (Fig. 1D right) were not affected by *Satb1* gene inactivation suggesting, despite its contribution to lymphocyte lineage initiation (22), that SATB1 was not required for B cell development. Our findings that early B cell development was not impaired in our cKO mice excluded any function for SATB1 in B-lineage choice maintenance. This data completed studies from Kanakura and Steidl groups (21, 22) that previously hypothesized that SATB1 function was restricted to stem cell renewal and fate hematopoietic lineages.

Since SATB1 is known to modulate *Rag* gene expression in T cells and incidentally TCR rearrangements (33), its consistent expression in early committed and developing B cells could confer a similar function on Ig gene rearrangements. We investigated this point by first analyzing *Rag* gene expression by RT-qPCR and found that sorted bone marrow pre-B cells from WT and cKO animals expressed similar transcript levels (*SI Appendix*, Fig. S1B). Given that SATB1 acts as a chromatin loop organizer in T cells (5, 37, 38), we suspected a potential effect of its deletion on *IgH V* region accessibility in developing B cells. To assess this point we examined VDJ recombination diversity by Repertoire-Sequencing experiments on RNA samples (RepSeq) extracted from pre-B cell-enriched bone marrow fractions. Our data displayed an equivalent broad distribution of each rearranged and expressed V gene (Fig. 1E and *SI Appendix*, Fig. S1C) in WT and cKO mice. A similar representation of  $V_H$  family usage indicated that SATB1 deletion did not hamper mechanisms leading to a diversified *IgH VDJ* repertoire in developing B cells. VDJ junction analysis in pre-immune repertoire of WT and *Satb1* cKO models revealed comparable lengths of CDR3 regions as well as normal distribution of P and N nucleotides (*SI Appendix*, Fig. S1D) suggesting that *IgH V* region assembly and end-joining occurs normally in the absence of STAB1. Altogether, our results indicate that SATB1 is dispensable to establish *IgH* repertoire and does not affect *Rag* expression or influence *IgH* locus accessibility. Given that SATB1 deletion did not impair VDJ rearrangements and B cell development, further studies could then be performed on peripheral B cell subsets in this model without any bias. In line with this statement, we determined whether SATB1 deletion could impact Ig isotype production and secretion in the mouse. Sera from WT, cHet and cKO animals were collected at two months of age and IgM, IgG1, IgG3 and IgA levels were quantified by ELISA assays. Homozygous and heterozygous KO mice displayed serum Ig levels comparable to WT for each isotype (Fig. 1F), suggesting that SATB1 does not influence global antibody production.

### **SATB1 depletion decreases *IgH* transcription in resting B cells**

Despite normal bone-marrow B cell development and Ig secretion in cKO mice, we however suspected that SATB1 depletion might influence *IgH* locus expression, given its capacity to bind MAR sequences *in vitro* (39) and to modulate MAR-containing reporter genes (29). We first examined surface IgM expression levels, as a component of B cell receptors, on immature and mature B cell subsets from bone marrow and spleen compartments in our model. By measuring IgM mean fluorescence intensity (MFI) by flow cytometry, *Satb1* cKO mice displayed consistent and significant decreased IgM surface expression when compared to WT, on bone marrow immature and recirculating B cells and on splenic transitional, marginal zone and follicular B cell populations (Fig. 2A). The same significant decrease in IgM BCR expression, also observed on naive B cells sorted from Peyer's patches (Fig. 2B), confirmed a specific function for SATB1 as a positive regulator of BCR expression in resting B cells. This role for SATB1 in BCR surface expression could explain its function as a contributing factor to B cell survival as described recently by Ozawa et al (34). Indeed, the knock-in reporter mouse model used in their study leads to inactivation of one *Satb1* allele and it is likely that such a deletion induces intermediate surface IgM expression level similar to the one we observed in our heterozygous mice (*SI Appendix*, Fig. S2). It is reasonable to suppose that reduced BCR expression on resting B cells reduces ability of these cells to respond to anti-IgM mediated *in vitro*-stimulation. One fair interpretation of the survival defect described by Ozawa et al (34) could be simply the consequence of reduced BCR expression in the knock-in model.

The decreased BCR expression observed in SATB1-deficient B lymphocytes raised the question of a potential Ig gene transcription defect in our model. Quantification of *IgH* and *Igk* primary transcription by RT-qPCR in resting splenic B cells showed a significant 2 fold-decrease of transcripts running through the *J<sub>H4</sub>* intron in cKO mice when compared to WT (Fig. 2C). Although not statistically significant, the same trend was also observed for *Igk* transcription in mutants (Fig. 2C). It is established that MAR sequences flanking both sides of the *IgH cE $\mu$*  intronic enhancer are able to bind SATB1 (40). Whereas binding of equivalent regions in the *Igk* locus has never been shown, it has to be noted that only one upstream MAR is associated with the *iE $\kappa$*  enhancer. The presence of either one or two MAR sequences surrounding these intronic enhancers could potentially explain differences observed for *IgH* and *Igk* transcription in our *Satb1* KO model. It is tempting to speculate that, through its DNA binding in proximity to intronic enhancers, SATB1 could physiologically potentiate their transcriptional effect in resting B cells. Although, this hypothesis is unlikely since IgM BCR expression was never compromised in resting B cell populations of mice devoid of either *E $\mu$* , *iE $\kappa$*  full regions or their associated-MARs (41–44). In contrast, the literature reports that BCR expression on resting B cells was decreased upon deletion of components of the 3'regulatory regions of both *IgH* (45) and *Igk* (46) loci, arguing for the importance of close contact

between 3'enhancers and respective promoters of rearranged V exons. In resting B cells, such loops have been reported within the *IgH* locus in many studies (47). For these reasons, a more rational hypothesis, in line with a proposed function for SATB1 as a promoter-enhancer loop regulator (38), could be that this MAR binding protein participates in physical interactions between rearranged  $V_H$  exons and the 3'RR and consequently enhances Ig chain transcription in resting B cells.

In line with the broad chromatin organization function proposed for SATB1 in T cells, we then performed total RNA seq analysis of both WT and SATB1-deficient B cell subsets. Resulting datasets were submitted to principal component analysis to validate reproducibility between samples (*SI Appendix*, Fig. S3A). When comparing datasets obtained from resting splenic B cells, our analysis only disclosed 65 genes displaying significant changes in expression: 53 were overexpressed and only 12 were downregulated ( $\log_2 FC > \pm 1.5$ ; Fig. 2D and *SI Appendix*, Table S1). Indeed, our data supports the hypothesis of an ambivalent function for SATB1, displaying both positive and negative regulatory actions for gene expression in resting B cells. Our study indicates that SATB1 depletion does not induce drastic transcriptional changes in resting B cells. In contrast to the T lineage (7, 27, 48), our data suggest that the intrinsic function of SATB1 in B cells may be relatively focused. In this cell type, SATB1 would have a distinctly different role from that of a major genome organizer.

### ***SATB1 depletion increases IgH transcription and Ig synthesis in activated B cells***

We sought to evaluate effect of SATB1 depletion on B cell activation in response to mitogenic and antigenic stimuli. We first performed *in vitro* stimulation of splenic B cells from both genotypes with LPS and carried out RNA seq in a time-course manner on resting B cells (sorted at day 0), activated B cells (bulk culture at day 2) and plasmablasts (sorted at day 4). Once sample reproducibility validated (*SI Appendix*, Fig. S3A), we compared transcriptional programs induced by transition from “resting to activated” and “activated to plasmablast” cell stages. Venn diagrams mainly indicate that both B cell activation and differentiation programs induced by LPS were not drastically impaired upon SATB1 depletion. Indeed, following *in vitro* culture, a vast majority of transcripts displayed a common regulation profile in WT and SATB1-deficient models: *i.e.* 5460 transcripts during the “resting to activated” transition and 2758 during the “activated to plasmablast” transition (*SI Appendix*, Fig. S3B). Nevertheless, we submitted differentially expressed genes from both transition programs, to a gene ontology analysis for biological processes relevant to B cell activation (see list in *SI Appendix*, Table S2). During the “resting to activated” transition, SATB1 depletion decreased expression of a few genes involved in humoral immune responses mediated by circulating immunoglobulins (Fig 3A red dots) including *Lta*, a gene recently described as a SATB1

target in T cells (38), as well as *Tgfb1* encoding TGF $\beta$ , the cytokine involved in regulation of isotype switching to IgA. Strikingly, the analysis specified that SATB1 depletion increased *Tcf3* expression in activated B cells (Fig. 3A blue dot), while its gene expression, encoding the E2A transcription factor, was unchanged in hematopoietic stem cells devoid of SATB1 (49). During the “activated to plasmablast” transition, SATB1 depletion downregulates *Cd40* and few genes involved in Ig class switching pathway (Fig. 3B red dots); this effect on *Cd40* targeting was expected since already observed upon SATB1 deletion in hematopoietic stem cells and T cells (38, 49).

In parallel, in such *in vitro* stimulation assays, B cells were tested for intrinsic abilities to proliferate, to undergo LPS-driven class switch recombination and to differentiate into plasmablasts. When measured by EDU incorporation, the ability of SATB1-deficient B cells to proliferate was similar to that of WT cells (*SI Appendix*, Fig. S4A). This suggested that, while decreasing IgM BCR expression on naive cells, SATB1-deficiency is dispensable for cell cycle entry of splenic B cells upon activation of TLR pathways. The same normal proliferation capacity in response to TLR or CD40 triggering was also recently reported for B cells devoid of the entire *Satb1* gene region (34). In our model, normal *in vitro* proliferation observed in B cells devoid of SATB1 was consistent to the normal proportion of spontaneous GC B cells in Peyer’s patches of KO animals (Fig. 3C), or the equivalent proportions of splenic GC centroblasts and centrocytes obtained after NP-CGG immunization (*SI Appendix*, Fig. S4B). More strikingly, in clear contrast with the decreased *IgH* expression in resting B cells, such *in vivo*-activated cells express higher BCR levels as evidenced by IgM surface labelling (Fig. 3D). This highlights that the ambivalent function for SATB1 clearly applies to Ig gene expression in developing mature B cells.

Class switching to IgG3 was examined by surface labelling of *in vitro*-activated splenic B cells by LPS at day 4. Data showed equivalent numbers of IgG3-expressing cells in both WT and SATB1 KO models (*SI Appendix*, Fig. S4C) indicating that SATB1 is not required for IgG3 class switching. RNA seq datasets from LPS- activated cells at day 2 also provided relevant portrayals of germline transcription landscape in the *IgH* locus of WT and SATB1 deficient models (*SI Appendix*, Fig. S4D). When examining germline transcription within the *I $\mu$* -donor and *I $\gamma$ 3*- or *I $\gamma$ 2b*-acceptor regions, both WT and SATB1-deficient models displayed identical profiles suggesting that, in agreement with normal CSR to IgG3 observed at day 4, SATB1 depletion does not impede LPS-induced germline transcription occurring prior to CSR.

When measured in the same *in vitro* assays, the intrinsic ability of splenic B cells to differentiate into plasmablasts (CD138<sup>+</sup>) upon LPS activation was also similar in WT and our SATB1 KO model (Fig. 3E). In agreement with *in vitro* data, NP-CGG immunization of mice induced generation of plasma cells in normal proportions in the absence of SATB1 (Fig. 3F). Altogether, these data suggest that SATB1 is dispensable for efficient differentiation into antibody secreting cells.

Although, once differentiated *in vitro* into plasmablasts, SATB1 deficient cells displayed higher levels of intracellular IgM when analyzed by flow cytometry (Fig. 3G). This increase in Ig production by plasmablasts was correlated with a significant increase of both *IgH* and *Igk* primary transcripts (respectively 1.4 and 1.7) in LPS-activated cells from SATB1 deficient animals.(Fig. 3H). Our data indicate that the absence of SATB1 in *in vitro*-activated B cells induces a more pronounced transcription of Ig genes. This is also true for *in vivo*-activated cells since Peyer's patch GC B cells also exhibited the same consistent and significant increase of *IgH* and *Igk* primary transcription upon SATB1 depletion (Fig.3I).

The effect of SATB1 deletion on global transcription of both IgH and IgL chains was also confirmed by comparing RNA seq data from splenic B cells activated by LPS at day 2 (Fig.3J) or from sorted plasmablasts at day 4 (Fig.3K). In the first case, among the 64 genes found upregulated in SATB1-deficient models ( $\log_2 FC > +/ - 1.5$ ), 20 of them corresponded to *IgV* H chain products and 23 others were identified as *IgV* κL chain products (Fig.3J left panel and *SI Appendix*, Table S3). When comparing plasmablasts, *IgV* H chain and *IgV* λL chain products were strongly increased upon SATB1 depletion (Fig. 3K, right panel and *SI Appendix*, Table S4).

Beyond the modest effect of SATB1 deletion on B cell activation programs induced by LPS *in vitro*, our study shows that STAB1 is dispensable for proliferation, CSR and differentiation into plasma cells. However, our data unveil a negative regulatory function of SATB1 for IgH and L chain transcription in activated B cells. This striking effect persists until the terminal stage since SATB1-deficient plasma cells display a higher content of Ig chains. However, SATB1-depletion does not allow plasma cells to produce more Ig since our deficient mice display broadly normal levels of serum antibody isotypes.

### ***SATB1* depletion increases somatic hypermutation**

Since our group recently reported a critical function of *cEμ*-associated MARs for SHM of the *IgH* locus in mouse models (50), and according to historic studies in the *Igk* locus (43), it was questionable whether MAR binding proteins, such as SATB1, could be involved in targeting somatic mutations of Ig genes. We first evaluated the ability of SATB1 deficient B cells to support somatic mutations by analyzing global SHM within intronic regions, not subject to antigen selection, located immediately downstream from V exons of both *IgH* and *Igk* loci and also upstream from *IgH Sμ* regions in spontaneous GC B cells from Peyer's patches. Second, to more accurately evaluate the ability of SATB1-deficient B cells to undergo SHM in response to specific antigen, we immunized mice with NP-CGG (51). By analogy to B cells devoid of *MARs<sub>Eμ</sub>* regions, in which SHM machinery gains access more frequently to the region downstream from *cEμ* (upstream from *Sμ*) (50), we carefully quantified SHM in this same region (Fig 4A, scheme). In GC B cells sorted from Peyer's patches, we found a significant increase of mutations in the absence of SATB1 (1.88 mut/Kb)



compared to WT (1.44 mut/Kb) (Fig. 4A *left bar graph*,  $p=0.0164$  and *SI Appendix*, Table S5A). A similar and significant increase of SHM in this region was also observed in splenic GC B cells sorted after NP-CGG immunization: while WT cells barely reached 0.22 mut/Kb, SATB1-deficient cells underwent two fold more mutations reaching 0.41 mut/Kb (Fig. 4A *right bar graph*,  $p=0.034$  and *SI Appendix*, Table S5B). This common feature shared by  $MAR_{SE\mu}$  KO- and *Satb1* KO-B cells suggests that SATB1 could be involved in limiting access of the SHM machinery to donor  $S\mu$  regions in cells undergoing SHM. When SHM was quantified downstream from the rearranged *IgH* variable regions (Fig 4B, scheme) in GC B cells sorted from Peyer's patches, homozygous *Satb1* deletion also increased, although not significantly, SHM frequencies downstream from  $J_H4$  (Fig. 4B *left bar graph*,  $p=0.054$  and *SI Appendix*, Table S5C). These findings were clearly confirmed upon NP-CGG immunization. Since B cell response to NP-CGG challenge is preferentially dominated by mutated clones expressing the  $V_H186.2$  segment (51), quantification of base substitutions in the  $J_H4$  intron downstream from the  $V_H186.2$ -rearranged exons is considered as a reliable hallmark of antigen-induced SHM. Indeed, in SATB1-deficient GC B cells from immunized mice, NP-CGG-induced mutations were significantly increased within the  $J_H4$  intron: mutant cells displayed 5.06 mut/Kb while WT cells only reached 3.28 mut/Kb (Fig. 4B *right bar graph*,  $p=0.020$  and *SI Appendix*, Table S5D). When global SHM was quantified in the *Igk* locus, within the intronic region downstream from *Jk5* (Fig 4C, scheme), both spontaneous and NP-CGG-induced GC B cells displayed an increased trend of mutation frequencies upon SATB1 depletion (Fig. 4C, *bar graphs* and *SI Appendix*, Table S5E and F). Since B cell responses to DNP hapten preferentially involve Ig composed of  $Ig\lambda 1$  light chains (52), it is likely that changes in mutation frequency within the *Igk* light chain loci, in this case not significant, probably underestimated any potential SHM increase induced by SATB1 depletion. It was also necessary to verify whether SATB1 depletion could modulate, through its transcription factor function, expression of any potential factors, including AID, involved in SHM. Indeed, when extracted from RNA-seq datasets, expression of 14 genes involved in SHM was unchanged upon SATB1 deletion in both LPS-activated B cells as well as LPS-induced plasmablasts (*SI Appendix*, Fig. S5A). Within Ig gene regions submitted to SHM analysis, substitution frequencies calculated at each base displayed a globally increased pattern (although not significant) that did not offer any hypothesis regarding the origin of the changes (*SI Appendix*, Fig. S5B). It is well established that AID targeting for SHM is coupled to transcription initiated at V promoters (53). Given the significant increase of primary transcription occurring at *IgH* and *Igk* in GC B cells devoid of SATB1 (Fig. 3I), one straightforward hypothesis to explain the global SHM increase in our model could be an overall increase in AID targeting of Ig genes. Interestingly, in line with this hypothesis, we recently highlighted increased AID deamination coupled to increased transcription in our mouse model devoid of  $MAR_{SE\mu}$  regions when bred into a DNA repair-deficient background (50). Surprisingly, while genomic deletion of  $MAR_{EIk}$  or  $MAR_{SE\mu}$  regions in the mouse

decreased SHM (43, 50), suppression of SATB1 led to the opposite effect in the regions downstream from the rearranged variable segments of *IgH* and *Igk* loci. This finding also plays in favor of a protective function for SATB1 against mutations. In this case, since the regions involved are located upstream from *E $\mu$*  and *Ei $\kappa$* , one could propose that the potential protective effect of SATB1 could take place in resting B cells in which SHM is not expected to happen. Moreover, the discrepancy between the effect of *MARs<sub>E $\mu$</sub>*  deletion and SATB1-depletion on targeted mutations upstream from intronic enhancers suggest that other interacting factors, beyond SATB1, participate in the complex regulation of *cis*-acting MARs. While we recently proposed that *MARs<sub>E $\mu$</sub>*  favor error-prone repair in its upstream regions in order to optimize SHM (50), our current study suggest that SATB1 does not participate in the unfaithful repair processes associated with SHM.

Taken together, our data suggest that SATB1-deletion increases SHM of Ig genes through a transcription-coupled mechanism that probably favors AID targeting.

### **Concluding remarks**

The fact that SATB1 plays a major role as a “genome organizer” in hematopoietic- and T-lineage cells (7, 28, 30) suggested that it is also important in B-lineage cells, a cell type that also undergoes fine developmental regulation of its expression (34). By using a conditional deletion model in B cells, our study fills in gaps concerning the function of SATB1 in this lymphocyte lineage. In contrast to its function in early-developing T cells, SATB1 is dispensable for the establishment of the Ig gene repertoire and overall early B cell development in the bone marrow. However, our model reveals that SATB1 is involved in the control of Ig gene transcription in mature B cells. Although previously unknown in the B lineage, our findings once again point to a dual function for SATB1 depending on the activation stage. We show that SATB1 promotes Ig gene transcription in resting B cells, while in activated B cells, it acts as a repressor. In contrast to the T cell lineage where SATB1 is considered a major regulatory factor of the enhancer network (38), SATB1 deletion does not induce major disruptions in the B cell transcriptome. Our invalidation model shows that only a reduced number of genes expressed during B differentiation are impacted by SATB1 deletion. In agreement with the repressive function of this factor in activated B cells, our study also shows that Ig genes are the predominant targets of SATB1 in activated B cells. Given the critical effect of SATB1 depletion on Ig gene transcription, it would certainly be interesting to evaluate SATB1 physiological binding to regulatory regions in Ig gene loci in resting and activated B cells.

By clarifying that SATB1 is an essential activator of Ig chain transcription, and consequently BCR expression, in resting B cells, our study extends recent work published by Yokota's team (34) and proposes a rational explanation for the survival function of SATB1 at this stage. These two studies also point out the importance of SATB1 expression levels which could, as demonstrated in HSCs

(22, 49), fine-tune some critical genes or regulatory pathways in lymphocytes. The dual function of SATB1 also observed in the B lineage, switching from activator to repressor of transcription in activated B cells, is reminiscent of the molecular switch observed upon phosphorylation and acetylation of this factor (16).

A surprising observation from our study is that SATB1 deletion increases somatic hypermutation of the *IgH* locus, extending into the *S $\mu$*  region. It is very likely that the *IgK* locus is similarly influenced. In this respect, SATB1 could play a protective role against SHM in resting cells. This finding should be read in conjunction with our recent study of the MARs regions associated with the *E $\mu$*  enhancer of the *IgH* locus (50), known to bind SATB1 (1). The SATB1-induced increase in SHM could logically be correlated with the observed increase in primary transcription of the variable regions of Ig H and L chains. However, it is not excluded that SATB1 modifies the accessibility of these regions to AID or its cofactors. Since it is proposed that SATB1 stabilizes the unpaired DNA regions against unwinding (13), such an action of SATB1 taking place in the *MARs-E $\mu$*  region could contribute to the protection of its surrounding regions against unwanted SHM. In line with the putative protective function of SATB1, it has been shown that this protein acts as an accessory factor of BER through its interaction with Oxo-guanine-glycosylase 1 (OGG1) (54), a DNA glycosylase that usually promotes error-free repair and that is not involved in Ig gene SHM (55).

Further study of these modifications of SATB1 in B cells will be necessary to clarify the origin of its versatility. It has recently been proposed that SATB1 isoforms are subject to phase separation in T cells (56). Consistent with its localization to PML nuclear bodies (57, 58), this mode of regulation deserves to be explored in the context of the B cell nucleus. Moreover, since chromatin loop extrusion mechanisms are now proven to be critical for Ig gene regulation (59), it might be relevant to evaluate a potential implication of SATB1 in this process.

## Materials and Methods

**Mice:** *Satb1<sup>tm1a/wt</sup>* mice come from the Mouse Clinical Institute (MCI) (IR00004167/ P4167). First crossings with 129S4/SvJae-*Gt(ROSA)26Sor<sup>tm2(FLP\*)sor</sup>/J* mice (The Jackson Laboratory) removed reporter and marker cassette placed between *Frt* sites. Second crossings with B6.C(CG)-*Cd79a<sup>tm1(cre)Reth</sup>/EhobJ* mice (The Jackson Laboratory) deleted *Satb1-exon4* due to CRE recombinase expression in B-lineage cells (35). *Aicda<sup>-/-</sup>* homozygous mice (kindly provided by Pr. T. Honjo) were used to prepare control samples devoid of SHM, as reported in (60). All experiments were performed on two-month-old mice. Primers used for genotyping are listed in *SI Appendix*, Table S6. All animal strains were bred and maintained in SPF conditions at 21–23°C with a 12-h

light/dark cycle. Procedures were reviewed and approved by the Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation APAFIS#16639-2018090612249522v2.

**Western Blot:** B cells from spleen were sorted with EASYSEP MOUSE B cell isolation kit (Stem Cell Technologies) and lysed using RIPA buffer (Santa Cruz) completed with protease inhibitor (Orthovanadate, PMSF, anti-protease cocktail). Proteins were quantified with Pierce BCA Protein Assay kit (Thermo Scientific) and denatured 5 minutes at 95°C. TGX Stain-free FastCast 12% acrylamide gels (Bio-Rad Laboratories) were used to separate proteins that were transferred on Transblot Turbo polyvinylidene fluoride membranes (Bio-Rad laboratories) with Transblot Turbo Transfer System. After blocking incubation with Phosphate Buffer Saline (PBS)- 5% skim milk, SATB1 primary antibody was incubated overnight à 4°C. Membrane was incubated with secondary antibody with PBS-3% milk. Proteins were detected and quantified with PIERCE ECL Western Blotting Substrate (Thermo Scientific) and ChemiDoc Touch Imaging System coupled to Image Lab J6.0 software (Bio-Rad Laboratories). Antibodies and concentrations used are in *SI Appendix*, Table S7.

**Enzyme-linked ImmunoSorbent Assay (ELISA):** Sera were collected from the blood of two-month-old mice. Supernatants were obtained after sorting B cells from spleens and *in vitro* LPS-activation during 4 days. Plates were coated overnight with 1µg/ml of primary antibody Goat Anti-Mouse Unlabeled IgM, IgG3, IgG1 or IgA (Southern Biotech) diluted in sodium carbonate buffer. Plates were blocked with PBS-3% bovine serum albumin 1hour at 37°C. Supernatant or diluted sera were incubated 2 hours at 37°C then secondary antibody conjugated with alkaline phosphatase (Southern Biotech) was incubated 1 hour at 37°C. Enzymatic reaction was performed by adding SigmaFast™ p-nitrophényl phosphate tablet. DO were measured at 405 nm on Thermo scientific MULTISKAN FC. Antibodies and concentrations used are in *SI Appendix*, Table S7.

**Flow Cytometry:** To analyze B cell populations, spleen and bone marrow were collected, crushed and filtered through Clearline Streater 40 µm (Fisherbrand). B cells were labeled in FACS Buffer (RPMI 1640' medium (Eurobio), 10% BSA (Stem Cell Technologies) 2mM EDTA) containing antibodies for 20 minutes at 4°C and analyzed on BD LSRFortessa SORP flow cytometer (BD Bioscience). For intracellular staining, cells were first stained for surface markers. After washing, cells were treated with the fixation/permeabilization kit Intraprep (Beckman Coulter, A07803), following the manufacturer's instructions. Mean Fluorescence Intensity values were normalized to WT values; antibodies used are provided in *SI Appendix*, Table S7.

**RNA extraction:** Pre-B cells, splenic resting B cells, *in vitro*-activated cells with LPS for 2 or 4 days, plasma cells and GC B cells were lyzed with TRIZOL Reagent (Ambion Life Technologies)

and RNA extraction was done following recommendations of DirectZol RNA microprep kit (ZymoResearch).

**RT-qPCR:** 1µg RNA from B cell subset samples was treated with DNase I Amplification Grade (Invitrogen) and reverse transcription was done following the High Capacity cDNA Reverse Transcription kit protocol (Applied Biosystem). Quantitative PCR was performed on 20ng cDNA using SensiFAST Probe Hi-ROX kits (Takara) on a Quant Studio III (Applied BioSystem). *IgH* and *Igk* primary transcripts were quantified as previously described (61). *Satb1* transcripts were measured by TaqMan assays with Mm.PT.58.13287891 (IDT), *Rag 1* and *Rag 2* transcripts with Mm01270936\_m1 and Mm00501300\_m1 (ThermoFisher), respectively. Transcript quantification was carried out with normalization to *Hprt* (Mm.PT.58.32092191 - IDT) for resting and *in vitro*-stimulated cells, and to *B220* (Mm.PT.58.32086646 - IDT) for GC B cells.

**Repertoire:** Pre-B cells from bone marrow were enriched using mouse CD25 MicroBead Kit (Miltenyi Biotec). Library preparation was adapted from methods previously described (62). Rapid amplification of cDNA ends (RACE) PCR was done using a specific reverse primer for µ constant region (CAGGTGAAGGAAATGGTGCT), and a cap race primer carrying Unique Molecular Identifiers (UMIs)(63). After purification, cDNA were amplified using a nested reverse primer hybridizing at the beginning of µ CH1 exon and universal forward primers. Finally, Illumina adapter and tag sequences were added by primer extension (62). Resulting libraries were sequenced on Illumina MiSeq sequencing system using MiSeq Reagent kit V3 600 cycles. Paired reads were merged as previously described (62) and UMIs were treated with MIGEC software. Repertoire analysis was done using IMG2/HIGHV-QUEST online tool (<http://img2.org/>) (64).

**In vitro Stimulation:** Splenic B cells isolated with EASYSEP MOUSE B cell isolation kit (Stem Cell Technologies) were cultured at  $1 \times 10^6$  cells/ mL in RPMI 1640 medium (Eurobio) supplemented with 10% fetal bovine serum (Dutscher), 2mM Glutamine (Eurobio), 1% Eagle's Non Essential Amino Acids (Eurobio), 50U/ml of penicillin-streptomycin (Gibco), 1mM sodium pyruvate (Eurobio), 129 µM 2-βmercaptoethanol (Sigma-Aldrich) in the presence of 1µg/ml Lipopolysaccharide (LPS-Invivogen). An aliquot of cells was collected for analysis and sample preparation after 2 or 4 days stimulation.

**In vitro Ethynil-DeoxyUridine (EdU) Incorporation:** Splenic B cells were cultured as described above and EdU was added 48 hours later and incorporated for 24 hours. Cells were then processed by following recommendations of Click-it EdU Alexa Fluor 488 Flow Cytometry Assay Kit (Thermo Fisher Scientific). EdU incorporation was evaluated by flow cytometry on BD LSRFortessa SORP flow cytometer (BD Bioscience)

**NP-CGG immunization:** Nitrophenylacetyl-Chicken Gamma Globulin (NP-CGG) immunization was realized by injecting 100µg NP-CGG precipitated with complete for the first intraperitoneal injection (day 0) and incomplete Freund's adjuvant for the second one at day 12. Pre-immune sera were collected before day 0 and immunized sera were collected at days 8 and 17. Mice were sacrificed at day 17 and splenic GC B cells were sorted to extract DNA for SHM quantification by HTS.

**Cell sorting:** Plasma cells were sorted using B220 and CD138 surface markers on BD FACS ARIA III (BD Bioscience). GC B cells were sorted using B220 and GL7 cell surface markers (*SI Appendix*, Table S7).

**Somatic Hypermutation Analysis:** SHM analysis was performed on B220<sup>+</sup>/GL7<sup>+</sup> GC B cells sorted from Peyer's patches or immunized spleens from *wt*, *Satb1* cKO and *Aicda* KO mice. *J<sub>H</sub>4* and *J<sub>k</sub>5* intronic regions were amplified from 10,000 cells with the following primers: VHJ558 5' CAGCCTGACATCTGAGGACTCTGC 3' ; SHMJH4 5'CAGCAACTACCCTTTTGAGACCGA3'; V<sub>k</sub>cons 5'GGCTGCAGSTTCAGTGGCAGTGGRTCWGGRAC3' and J<sub>k</sub>5SHM 5'AGCGAATTCAACTTAGGAGACAAAAGAGAGAAC3' using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and according to the following program: denaturation (98°C 10s), hybridization (69°C 30s) and amplification (72°C 1mn) during 38 cycles. Libraries were constructed with Ion Xpress Plus gDNA Fragment Library kit (Cat. no. 4471269, Life Technologies) and sequenced on the Ion-Proton System S5. SHM frequencies were determined using Raw Data analyzed with DeMinEr tool (60)

**RNA-Sequencing:** RNA-Seq analysis was performed on splenic resting B cells, *in vitro*-activated B cells for 2 days and *in vitro*-differentiated plasmablasts for 4 days after LPS stimulation. Sample quality controls and library preparation were performed at the GeT-Santé facility (Inserm, Toulouse, France, get.genotoul.fr). i/ Quality control: RNA concentration and purity were determined using a ND-2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, USA). RNA integrity was checked with a Fragment Analyzer (Agilent Technologies, Santa Clara, USA), using the RNA Standard Sensitivity Kit. 260/280 purity ratios were all ≥ 1.8, and integrity indices were of good quality (8.3-10 RIN and > 1.7 28S/18S ratios). ii/ Library preparation (GeT-Santé): RNA-seq paired-end libraries were prepared according to Illumina's protocol with some adjustments, using the TruSeq Stranded Total RNA Gold library prep Kit (Illumina, San Diego, USA). Briefly, between 934-1000 ng total RNA were first ribo-zero depleted using Illumina Ribo-Zero probes. Then, remaining RNA were fragmented during 2 min and retrotranscribed to generate double stranded cDNA. Compatible adaptors were ligated, allowing the barcoding of the samples with unique dual indices. Libraries were amplified by 12 PCR cycles of PCR and an additional final purification step resulted

in 280-700 bp fragments. Libraries quality was assessed using the HS NGS kit on the Fragment Analyzer (Agilent Technologies, Santa Clara, USA). iii/ Library quantification (GeT-PlaGe): quantification and sequencing were performed at the GeT-PlaGe core facility (INRAE, Toulouse, France). Libraries were quantified by qPCR using the KAPA Library Quantification Kit (Roche, Basel, Switzerland) to obtain an accurate quantification. iiiii/ Sequencing (GeT-PlaGe): libraries were equimolarly pooled and RNA sequencing was then performed on one S4 lane of the Illumina NovaSeq 6000 instrument (Illumina, San Diego, USA), using the NovaSeq 6000 S4 v1.5 Reagent Kit (300 cycles), and a paired-end 2 x 150 pb strategy.

**RNA-Sequencing Analysis:** Paired-end reads were mapped on GRCm38 mouse genome that was previously indexed with “Mus\_musculus.GRCm38.dna.primary\_assembly.fa” and “Mus\_musculus.GRCm38.102.chr\_patch\_hapl\_scaff.gtf” files from ENSEMBL release 102. Index and mapping steps were both made with STAR v2.6.0c (65). Then featureCounts v2.0.1 (66) was used to count reads by gene. An R script named template\_script\_DESeq2\_CL.r of SARTools (67) was run first with all count data in order to retrieve a PCA and check if biological variability was the main source of variance in the data. Then, the same script was run for each desired differential analysis with count data from defined reference and interest conditions. Differentially regulated genes with an adjusted p value of 0.05 and a foldchange  $\leq -1.5$  or  $\geq 1.5$  were selected for downstream analysis. Gene SYMBOLS were converted to ENTREZIDs with the bitr function of the R ClusterProfiler package (68). The resulting ENTREZIDs and their associated log2foldchange were then used to calculate enriched biological pathway profiles of different gene clusters (Down in cKO , Up in cKO , Down in WT and Up in WT) using the CompareCluster function of ClusterProfiler (68) with p value and q value thresholds set to 0.01 and 0.05 respectively. The resulting enriched functional profiles were filtered through a GO list consisting of the hierarchical children of the following Biological pathway terms: B cell costimulation, B cell selection, humoral immune response, immunoglobulin production, memory B cell differentiation, regulation of the apoptotic process of mature B cells, SHM (*SI Appendix*, Table S2). Terms that were enriched in all down-regulated gene clusters or all up-regulated clusters were discarded.

**Data availability:** Raw data from RNA-seq, SHM and Rep-Seq have been deposited in the European Nucleotide Archive database under access number PRJEB52320.

## Acknowledgments

We thank BISCEm and the animal core facility team for help with mouse work on both practical and regulatory aspects. We are grateful to Emeline Lhuillier and the Genotoul Plateau GeT-Santé facility (<https://get.genotoul.fr>) for technical assistance with RNA sequencing. MT and OM were

supported by PhD fellowships from the French Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation and the Fonds Européen de Développement Régional (FEDER). This work was supported by La Ligue Contre le Cancer (comités 87, 23 to EP and SLN); the Fondation ARC pour la recherche sur le cancer (PJA 20181207918 to EP and PhD continuation fellowship to MT), Institut CARNOT CALYM, INCa-Cancéropôle GSO Emergence (to EP). We are grateful to Drs. Jeanne Moreau and Amélie Bonaud for critical reading of the manuscript, helpful comments and edits.

## References

1. L. Dickinson, J. Tadashi, K. Yoshinori, K.-S. Terumi, A tissue-specific MARSAR DNA-binding protein with unusual binding site recognition. *Cell* **70**, 631–645 (1992).
2. J. D. Alvarez, *et al.*, The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev.* **14**, 521–535 (2000).
3. S. Cai, H.-J. Han, T. Kohwi-Shigematsu, Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat. Genet.* **34**, 42–51 (2003).
4. D. Yasui, M. Miyano, S. Cai, P. Varga-Weisz, T. Kohwi-Shigematsu, SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* **419**, 641–645 (2002).
5. S. Cai, C. C. Lee, T. Kohwi-Shigematsu, SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* **38**, 1278–1288 (2006).
6. T. Kohwi-Shigematsu, *et al.*, Genome organizing function of SATB1 in tumor progression. *Semin. Cancer Biol.* **23**, 72–79 (2013).
7. T. Zelenka, C. Spilianakis, SATB1-mediated chromatin landscape in T cells. *Nucl. Austin Tex* **11**, 117–131 (2020).
8. Z. Wang, *et al.*, The structural basis for the oligomerization of the N-terminal domain of SATB1. *Nucleic Acids Res.* **40**, 4193–4202 (2012).
9. P. K. Purbey, *et al.*, PDZ domain-mediated dimerization and homeodomain-directed specificity are required for high-affinity DNA binding by SATB1. *Nucleic Acids Res.* **36**, 2107–2122 (2008).
10. K. Yamasaki, T. Akiba, T. Yamasaki, K. Harata, Structural basis for recognition of the matrix attachment region of DNA by transcription factor SATB1. *Nucleic Acids Res.* **35**, 5073–5084 (2007).



11. I. De Belle, S. Cai, T. Kohwi-Shigematsu, The Genomic Sequences Bound to Special AT-rich Sequence-binding Protein 1 (SATB1) In Vivo in Jurkat T Cells Are Tightly Associated with the Nuclear Matrix at the Bases of the Chromatin Loops. *J. Cell Biol.* **141**, 335–348 (1998).
12. J. Seo, M. M. Lozano, J. P. Dudley, Nuclear Matrix Binding Regulates SATB1-mediated Transcriptional Repression. *J. Biol. Chem.* **280**, 24600–24609 (2005).
13. R. P. Ghosh, *et al.*, Satb1 integrates DNA binding site geometry and torsional stress to differentially target nucleosome-dense regions. *Nat. Commun.* **10** (2019).
14. P. P. Kumar, P. K. Purbey, D. S. Ravi, D. Mitra, S. Galande, Displacement of SATB1-Bound Histone Deacetylase 1 Corepressor by the Human Immunodeficiency Virus Type 1 Transactivator Induces Expression of Interleukin-2 and Its Receptor in T Cells. *Mol. Cell. Biol.* **25**, 1620–1633 (2005).
15. S. Galande, P. K. Purbey, D. Notani, P. P. Kumar, The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr. Opin. Genet. Dev.* **17**, 408–414 (2007).
16. P. Pavan Kumar, *et al.*, Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. *Mol. Cell* **22**, 231–243 (2006).
17. P. K. Purbey, *et al.*, Acetylation-dependent interaction of SATB1 and CtBP1 mediates transcriptional repression by SATB1. *Mol. Cell. Biol.* **29**, 1321–1337 (2009).
18. S. P. Khare, *et al.*, NF- $\kappa$ B Signaling and IL-4 Signaling Regulate SATB1 Expression via Alternative Promoter Usage During Th2 Differentiation. *Front. Immunol.* **10**, 667 (2019).
19. I. Patta, *et al.*, Dynamic regulation of chromatin organizer SATB1 via TCR-induced alternative promoter switch during T-cell development. *Nucleic Acids Res.* **48**, 5873–5890 (2020).
20. M. A. Balamotis, *et al.*, Satb1 Ablation Alters Temporal Expression of Immediate Early Genes and Reduces Dendritic Spine Density during Postnatal Brain Development. *Mol. Cell. Biol.* **32**, 333–347 (2012).
21. B. Will, *et al.*, Satb1 regulates the self-renewal of hematopoietic stem cells by promoting quiescence and repressing differentiation commitment. *Nat. Immunol.* **14**, 437–445 (2013).
22. Y. Doi, *et al.*, Variable SATB1 Levels Regulate Hematopoietic Stem Cell Heterogeneity with Distinct Lineage Fate. *Cell Rep.* **23**, 3223–3235 (2018).
23. Y. Satoh, *et al.*, The Satb1 Protein Directs Hematopoietic Stem Cell Differentiation toward Lymphoid Lineages. *Immunity* **38**, 1105–1115 (2013).
24. J. Wen, *et al.*, SATB1 family protein expressed during early erythroid differentiation modifies globin gene expression. *Blood* **105**, 3330–3339 (2005).

25. S. M. Hawkins, T. Kohwi-Shigematsu, D. G. Skalnik, The Matrix Attachment Region-binding Protein SATB1 Interacts with Multiple Elements within the gp91<sup>phox</sup> Promoter and Is Down-regulated during Myeloid Differentiation. *J. Biol. Chem.* **276**, 44472–44480 (2001).
26. Y. Fujii, A. Kumatori, M. Nakamura, SATB1 Makes a Complex with p300 and Represses gp91phox Promoter Activity. *Microbiol. Immunol.* **47**, 803–811 (2003).
27. R. Naik, S. Galande, SATB family chromatin organizers as master regulators of tumor progression. *Oncogene* **38**, 1989–2004 (2019).
28. G. Papadogkonas, D.-A. Papamatheakis, C. Spilianakis, 3D Genome Organization as an Epigenetic Determinant of Transcription Regulation in T Cells. *Front. Immunol.* **13**, 921375 (2022).
29. T. Kohwi-Shigematsu, K. Maass, J. Bode, A thymocyte factor SATB1 suppresses transcription of stably integrated matrix-attachment region-linked reporter genes. *Biochemistry* **36**, 12005–12010 (1997).
30. M. Burute, K. Gottimukkala, S. Galande, Chromatin organizer SATB1 is an important determinant of T-cell differentiation. *Immunol. Cell Biol.* **90**, 852–859 (2012).
31. K. Kakugawa, *et al.*, Essential Roles of SATB1 in Specifying T Lymphocyte Subsets. *Cell Rep.* **19**, 1176–1188 (2017).
32. Y. Kitagawa, *et al.*, Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat. Immunol.* **18**, 173–183 (2017).
33. B. Hao, *et al.*, An anti-silencer– and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J. Exp. Med.* **212**, 809–824 (2015).
34. T. Ozawa, *et al.*, Special AT-Rich Sequence-Binding Protein 1 Supports Survival and Maturation of Naive B Cells Stimulated by B Cell Receptors. *J. Immunol.* **208**, 1937–1946 (2022).
35. E. Hobeika, *et al.*, Testing gene function early in the B cell lineage in mb1-cre mice. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13789–13794 (2006).
36. The Immunological Genome Project Consortium, *et al.*, The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
37. T. Kohwi-Shigematsu, *et al.*, SATB1-mediated functional packaging of chromatin into loops. *Methods San Diego Calif* **58**, 243–254 (2012).
38. T. Zelenka, *et al.*, “The 3D enhancer network of the developing T cell genome is controlled by SATB1” (bioRxiv, 2021).

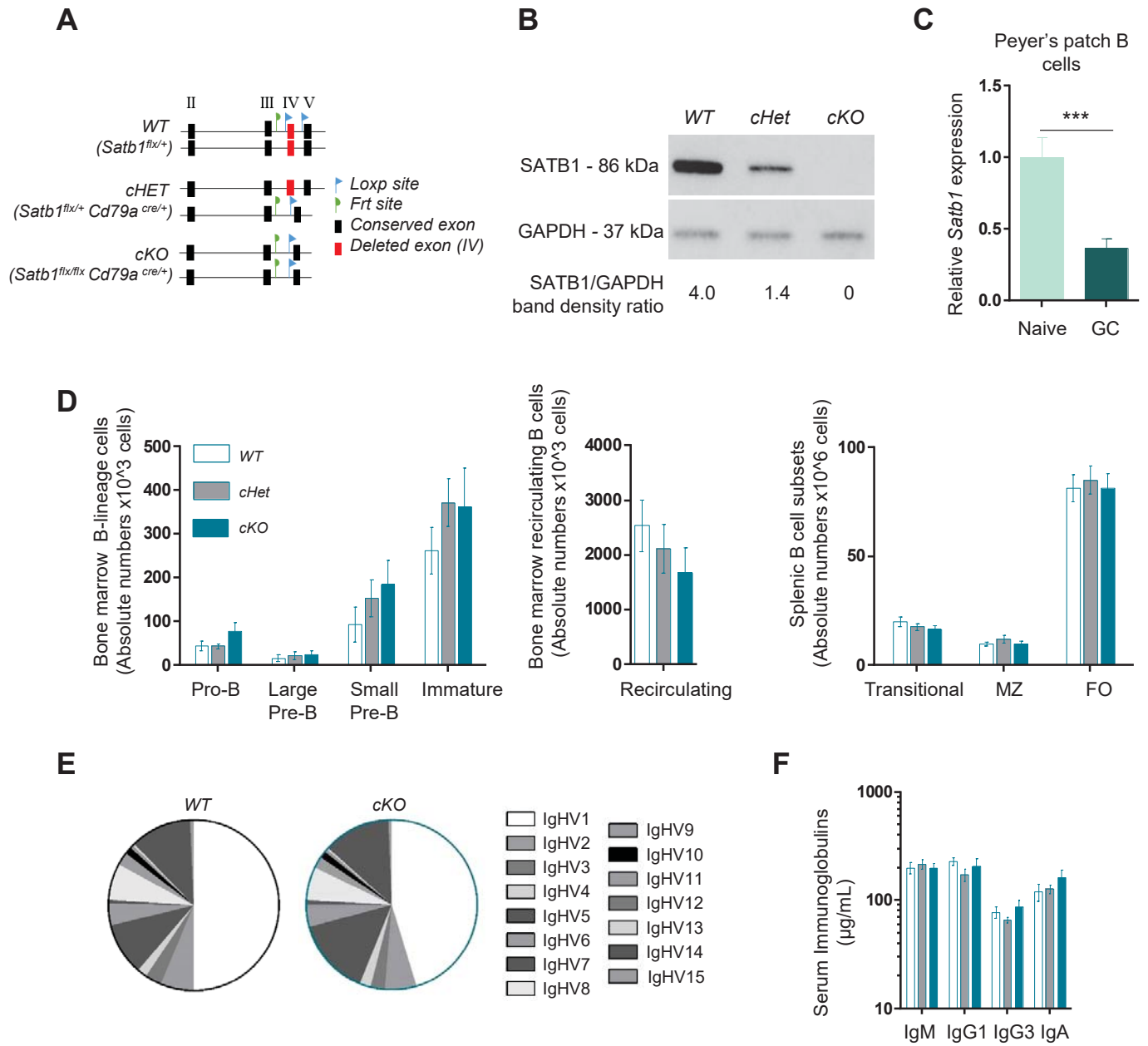
39. L. A. Dickinson, C. D. Dickinson, T. Kohwi-Shigematsu, An Atypical Homeodomain in SATB1 Promotes Specific Recognition of the Key Structural Element in a Matrix Attachment Region. *J. Biol. Chem.* **272**, 11463–11470 (1997).
40. L. A. Dickinson, T. Kohwi-Shigematsu, Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. *Mol. Cell. Biol.* **15**, 456–465 (1995).
41. M. Marquet, *et al.*, The E $\mu$  enhancer region influences H chain expression and B cell fate without impacting IgVH repertoire and immune response in vivo. *J. Immunol. Baltim. Md 1950* **193**, 1171–1183 (2014).
42. Y. Xu, L. Davidson, F. W. Alt, D. Baltimore, Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. *Immunity* **4**, 377–385 (1996).
43. M. Yi, P. Wu, K. W. Trevor, L. Claflin, W. T. Garrard, Evidence that the Igkappa gene MAR regulates the probability of premature V-J joining and somatic hypermutation. *J. Immunol. Baltim. Md 1950* **162**, 6029–6039 (1999).
44. E. Sakai, A. Bottaro, L. Davidson, B. P. Sleckman, F. W. Alt, 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 (1999).
45. A. Garot, *et al.*, 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 (2016).
46. M. Inlay, F. W. Alt, D. Baltimore, Y. Xu, Essential roles of the kappa light chain intronic enhancer and 3' enhancer in kappa rearrangement and demethylation. *Nat. Immunol.* **3**, 463–468 (2002).
47. C. Bruzeau, J. Moreau, S. Le Noir, E. Pinaud, Panorama of stepwise involvement of the IgH 3' regulatory region in murine B cells. *Adv. Immunol.* **149**, 95–114 (2021).
48. T. Yokota, Y. Kanakura, Role of tissue-specific AT-rich DNA sequence-binding proteins in lymphocyte differentiation. *Int. J. Hematol.* **100**, 238–245 (2014).
49. Y. Satoh, *et al.*, The Satb1 protein directs hematopoietic stem cell differentiation toward lymphoid lineages. *Immunity* **38**, 1105–1115 (2013).
50. O. Martin, *et al.*, "The IgH E $\mu$  -MAR regions promote UNG-dependent error-prone repair to optimize somatic hypermutation" (*Immunology*, 2022) <https://doi.org/10.1101/2022.08.15.503996> (August 24, 2022).
51. A. Cumano, K. Rajewsky, Clonal recruitment and somatic mutation in the generation of immunological memory to the hapten NP. *EMBO J.* **5**, 2459–2468 (1986).

52. M. Reth, G. J. Hämmerling, K. Rajewsky, Analysis of the repertoire of anti-NP antibodies in C57BL/6 mice by cell fusion. I. Characterization of antibody families in the primary and hyperimmune response. *Eur. J. Immunol.* **8**, 393–400 (1978).
53. Y. Fukita, H. Jacobs, K. Rajewsky, Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* **9**, 105–114 (1998).
54. S. Kaur, *et al.*, Special AT-rich Sequence-binding Protein 1 (SATB1) Functions as an Accessory Factor in Base Excision Repair. *J. Biol. Chem.* **291**, 22769–22780 (2016).
55. D. B. Winter, *et al.*, Normal somatic hypermutation of Ig genes in the absence of 8-hydroxyguanine-DNA glycosylase. *J. Immunol. Baltim. Md 1950* **170**, 5558–5562 (2003).
56. T. Zelenka, *et al.*, “SATB1 undergoes isoform-specific phase transitions in T cells” (bioRxiv, 2022).
57. P. P. Kumar, *et al.*, Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat. Cell Biol.* **9**, 45–56 (2007).
58. J.-A. T. Tan, *et al.*, SUMO conjugation to the matrix attachment region-binding protein, special AT-rich sequence-binding protein-1 (SATB1), targets SATB1 to promyelocytic nuclear bodies where it undergoes caspase cleavage. *J. Biol. Chem.* **283**, 18124–18134 (2008).
59. Y. Zhang, X. Zhang, H.-Q. Dai, H. Hu, F. W. Alt, The role of chromatin loop extrusion in antibody diversification. *Nat. Rev. Immunol.* **22**, 550–566 (2022).
60. O. A. Martin, *et al.*, 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 (2018).
61. A. Tinguely, *et al.*, Cross talk between immunoglobulin heavy-chain transcription and RNA surveillance during B cell development. *Mol. Cell. Biol.* **32**, 107–117 (2012).
62. V. Javaugue, *et al.*, RNA-based immunoglobulin repertoire sequencing is a new tool for the management of monoclonal gammopathy of renal (kidney) significance. *Kidney Int.* **101**, 331–337 (2022).
63. M. A. Turchaninova, *et al.*, High-quality full-length immunoglobulin profiling with unique molecular barcoding. *Nat. Protoc.* **11**, 1599–1616 (2016).
64. M. Boice, *et al.*, Loss of the HVEM Tumor Suppressor in Lymphoma and Restoration by Modified CAR-T Cells. *Cell* **167**, 405-418.e13 (2016).
65. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
66. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

67. H. Varet, L. Brillet-Guéguen, J.-Y. Coppée, M.-A. Dillies, SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLOS ONE* **11**, e0157022 (2016).
68. T. Wu, *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).



Fig. 1



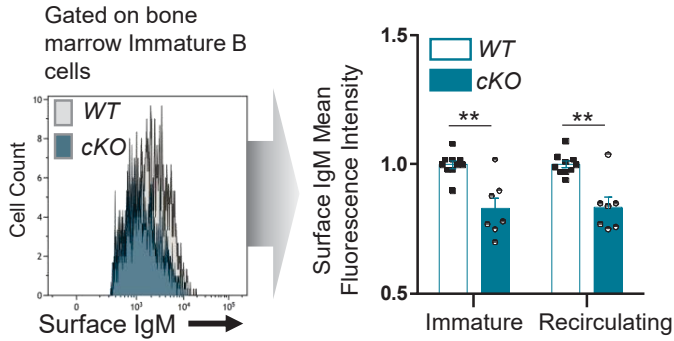
**Figure 1.**

**A)** Conditional deletion of *Satb1*-exon 4 by *Cd79a<sup>cre/+</sup>* recombinase in B cells. Exon 4, cre site, frr site, conditional allele (*Satb1<sup>flx</sup>*) and Wild type allele (*Satb1<sup>+</sup>*) are indicated. **B)** SATB1 protein quantification by Western Blot in splenic B cells from WT (*Satb1<sup>flx/+</sup>*), conditional heterozygous (*Satb1<sup>flx/+</sup> Cd79a<sup>cre/+</sup>*) and conditional homozygous (*Satb1<sup>flx/flx</sup> Cd79a<sup>cre/+</sup>*) mice. **C)** *Satb1* transcript quantification by RT-qPCR (normalized to *Hprt*) in Peyer's patch naive and GC B cells sorted from WT mice (n=8). **D)** Bar graphs displaying absolute numbers of B cell subsets determined by flow cytometry from bone marrow (per femur, n=7-10) and spleen (n=11-15). **E)** Pie chart representation of *IgHV* family gene usage, quantified by RACE-RepSeq, expressed by pre-B cell-enriched bone marrow populations in WT (n=5) and cKO mice (n=6). **F)** Immunoglobulin serum isotype levels quantified by ELISA in 8 to 10 week-old WT and cKO mice (n=8-15). p-value was determined with two tailed Mann Whitney test; significant differences are indicated by: \* P<0.05; \*\* P<0.01; \*\*\* P<0.001; \*\*\*\* P<0.0001 and error bars represent SEM.

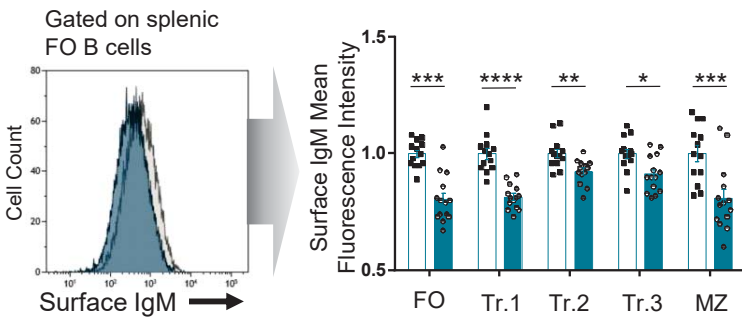


Fig. 2

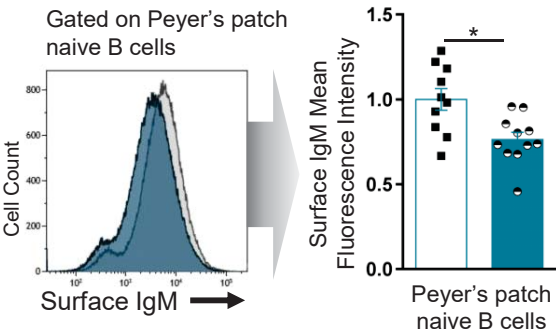
**A**



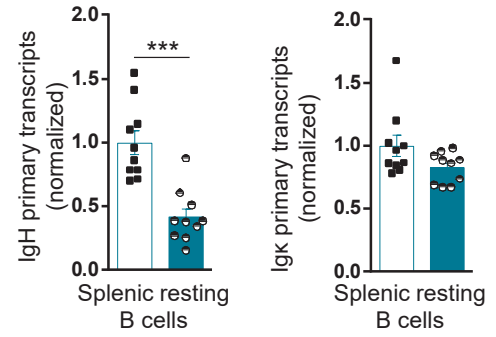
**B**



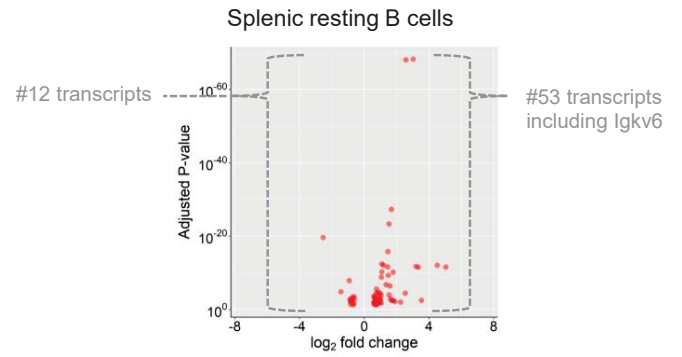
**C**



**D**



**E**

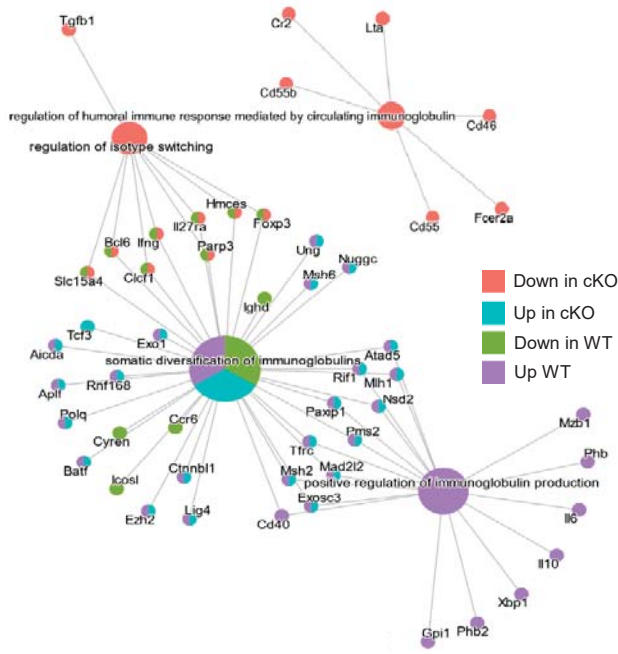


**Figure 2.**

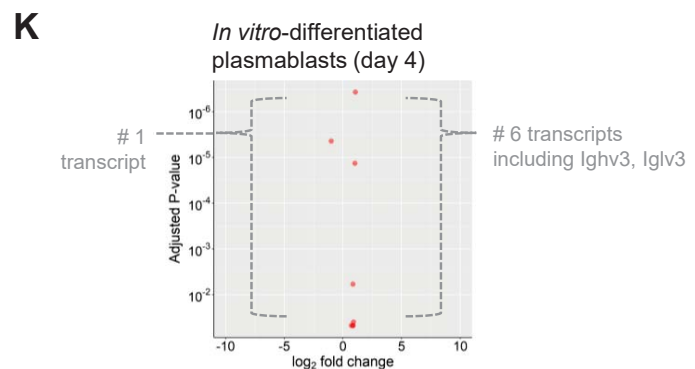
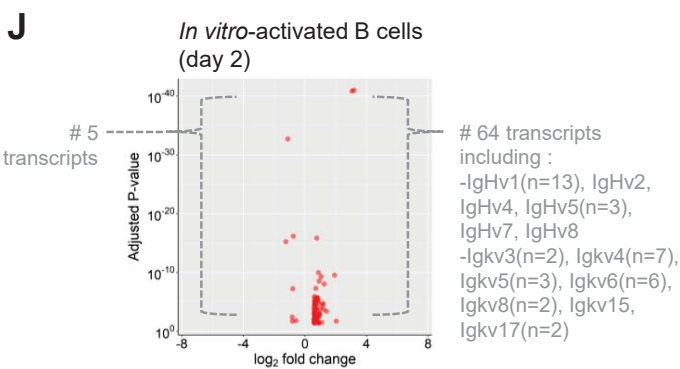
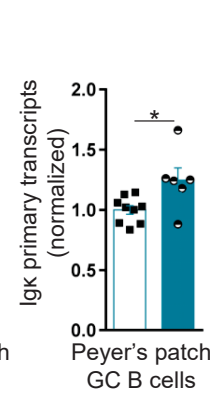
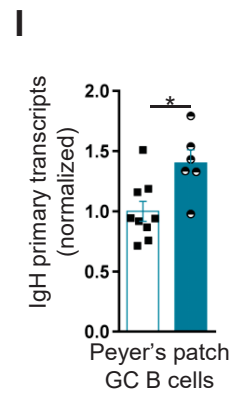
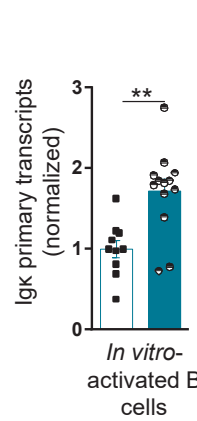
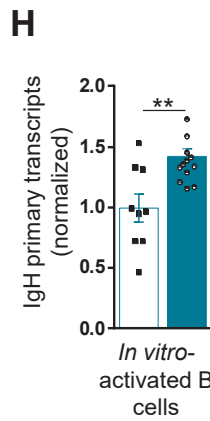
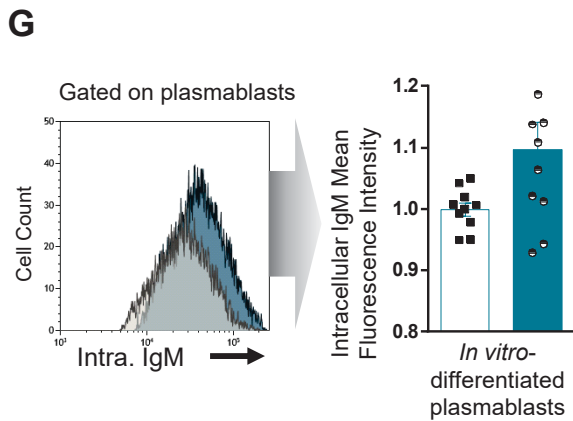
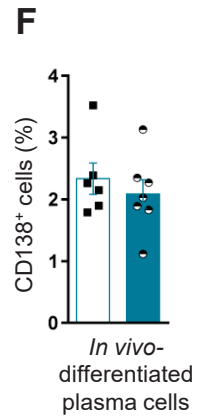
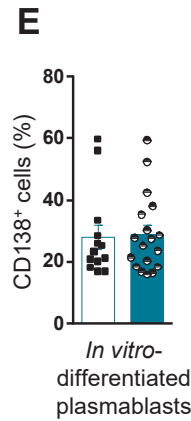
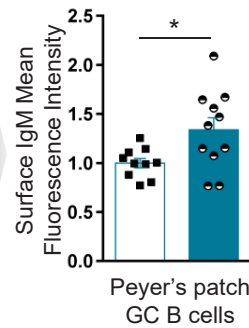
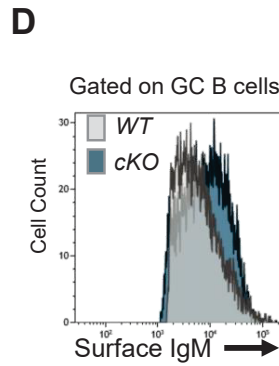
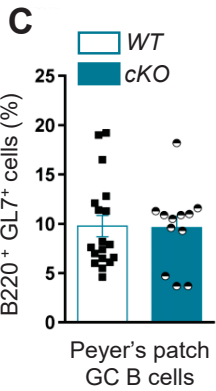
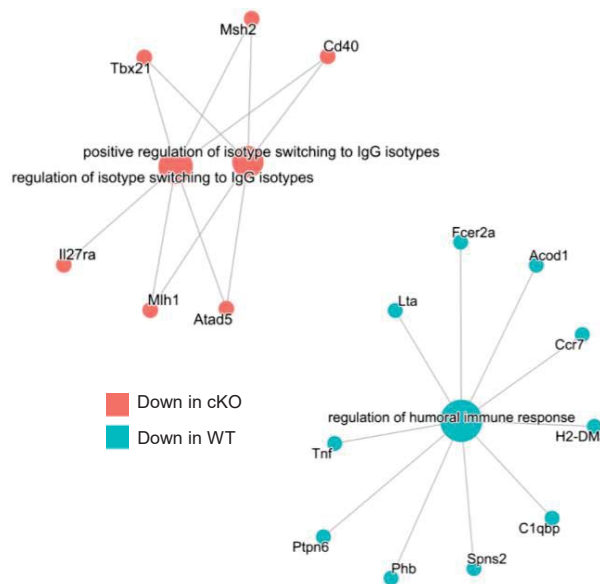
**A)** Comparison of surface IgM fluorescence intensity, evaluated by flow cytometry, on immature and recirculating bone marrow B cell subsets from two-month-old WT and homozygous cKO animals. One representative histogram is shown (left), bar graph (right) displays normalized surface IgM MFI in WT and cKO animals (n=7-10). **B)** Comparison of surface IgM fluorescence intensity, evaluated by flow cytometry, on Follicular (FO), Marginal zone (MZ), Transitional (Tr.1, Tr.2, Tr.3) splenic B cell subsets from two-month-old WT and homozygous cKO animals. One representative histogram is shown (left), bar graph (right) displays normalized surface IgM MFI in WT and cKO animals (n=13-14). **C)** Comparison of surface IgM fluorescence intensity, evaluated by flow cytometry, on naive B cells sorted from Peyer's Patches from two-month-old WT and homozygous cKO animals. One representative histogram is shown (left), bar graph (right) display normalized surface IgM MFI in WT and cKO animals (n=10-11). **D)** *IgH* and *Igk* primary transcripts quantified by RT-qPCR (normalized to *Hprt*) in splenic resting B cells sorted from WT and cKO mice (n= 6-10). **E)** Volcano plot from RNA-seq indicating differentially expressed transcripts when comparing WT and cKO mice (n=2) resting B cells. Transcripts located within Ig gene loci are indicated. p-value was determined by two tailed Mann Whitney test and significant differences are indicated by: \* P<0.05; \*\* P<0.01; \*\*\* P<0.001; \*\*\*\* P<0,.0001; error bars represent SEM.

**Fig. 3**

**A** Resting (day 0) to *in vitro*-activated (day 2) transition



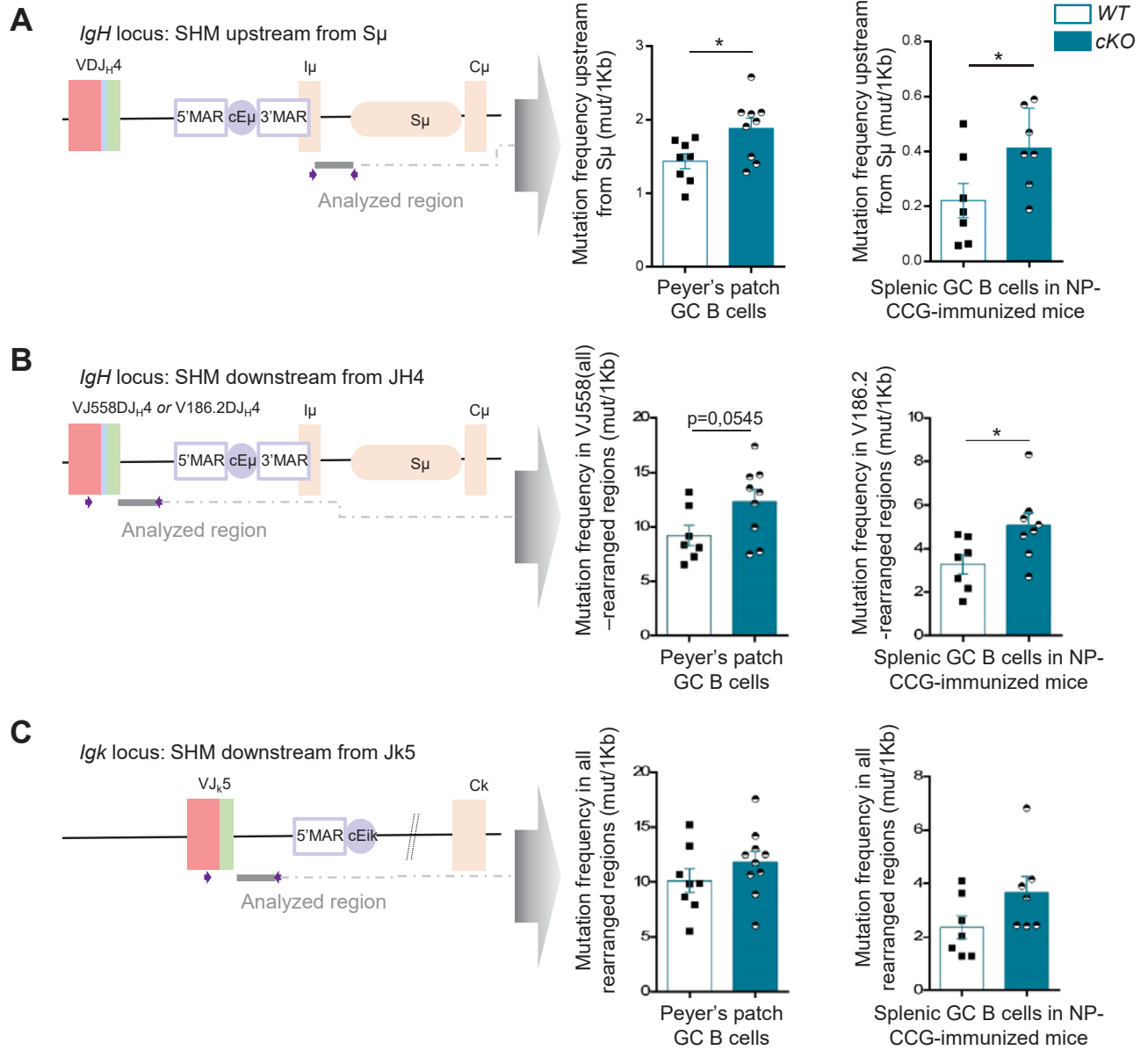
**B** *In vitro*-activated (day 2) to plasmablast (day 4) transition



**Figure 3.**

**A)** Visual representation of Gene Ontology enrichment analysis of genes significantly associated with biological pathways upon B cell transition from resting (day 0) to *in vitro*-activated (day 2) subsets in WT and cKO mice. Datasets from RNA-seq experiments were used for the analysis. **B)** Same representation as above for B cell transition from *in vitro*-activated (day 2) to plasmablast (day 4) subsets. **C)** Germinal center (GC) B cell subsets evaluated by flow cytometry in Peyer's patches of WT and cKO animals (n=12-18). **D)** Comparison of surface IgM fluorescence intensity evaluated by flow cytometry, on Peyer's patch GC B cells from two-month-old WT and homozygous cKO animals. One representative histogram is shown (left), bar graph (right) displays normalized surface IgM MFI in WT and cKO animals (n=10-11) **E)** Proportion of plasmablasts evaluated by flow cytometry, obtained following *in vitro* LPS activation (day 4) of splenic B cells from WT and cKO animals (n=13-18) (left). **F)** Proportion of plasma cells in the spleen evaluated by flow cytometry following NP-CGG immunization of WT and cKO animals (n=6-7). **G)** Comparison of intracellular IgM fluorescence intensity evaluated by flow cytometry, in *in vitro*-differentiated plasmablasts (day4) from WT and homozygous cKO animals. One representative histogram is shown (left), bar graph (right) displays normalized intracellular IgM MFI in WT and cKO animals (n=9-10). **H)** *IgH* and *Igk* primary transcripts quantified by RT-qPCR (normalized to *Hprt*) in *in vitro* LPS-activated B-lineage cells (day 4) of WT and cKO animals (n=9-13). **I)** *IgH* and *Igk* primary transcripts (normalized to *B220*) in GC B cells from Peyer's patches of WT and cKO animals (n=6-9). **J)** Volcano plots from RNA-seq indicating differentially expressed transcripts when comparing WT and cKO mice (n=2) in *in vitro* LPS-activated B cells (day 2). Transcripts located within Ig gene V regions are indicated. **K)** Same analysis as above in *in vitro* LPS-differentiated plasmablasts (day 4). p-value was determined by two tailed Mann Whitney test and significant differences were indicated by: \* P<0.05; \*\* P<0.01; \*\*\* P<0.001; \*\*\*\* P<0.0001,; error bars represent SEM.

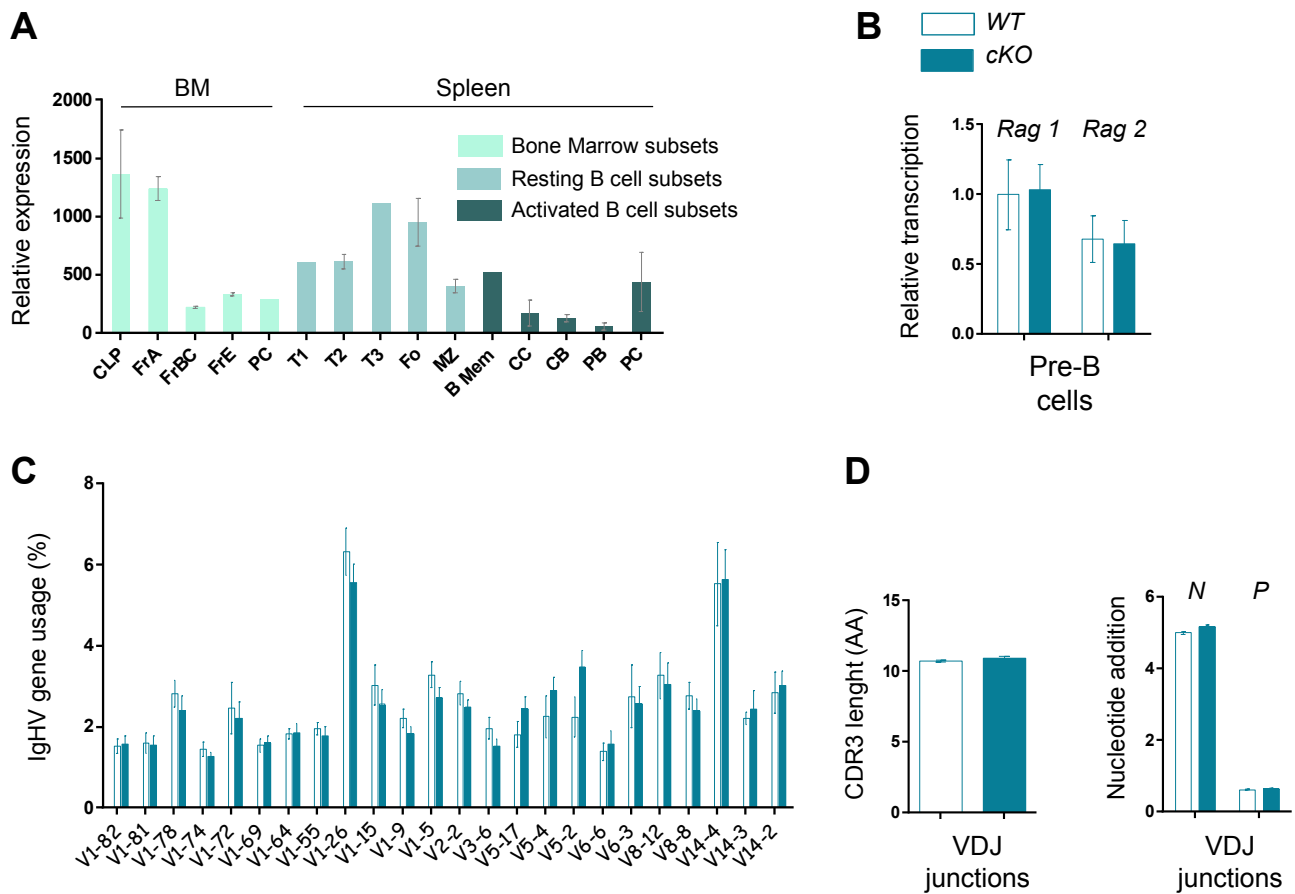
Fig. 4



**Figure 4.**

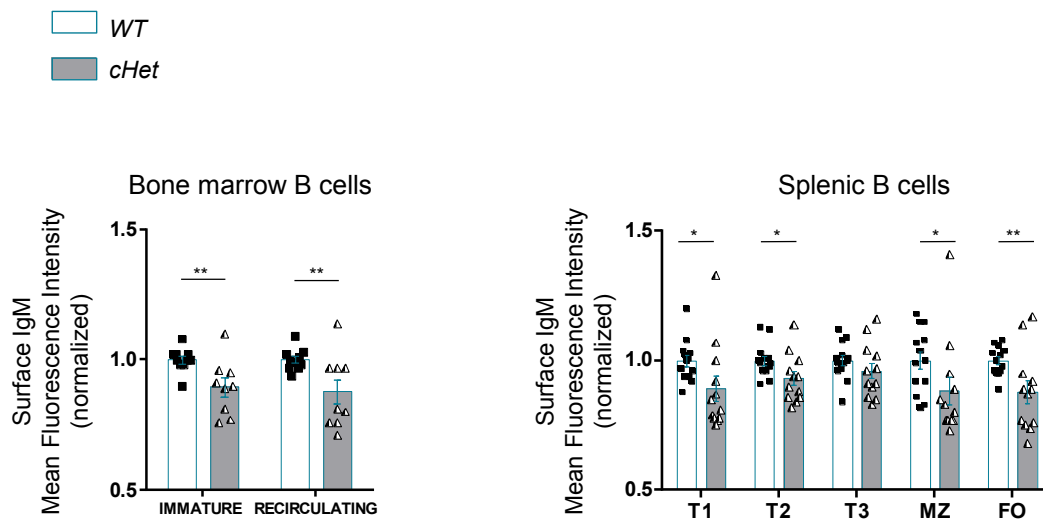
Comparison of somatic hypermutation frequencies in WT and cKO animals, quantified in various regions of Ig genes targeted by AID (left schemas), taking place in spontaneous GC B cells sorted from Peyer's patches (n=7-10, left bar graphs) and in splenic GC B cells obtained following NP-CGG immunization (n=7-8, right bar graphs). Data were obtained by NGS (Ion proton) combined with DeMinEr filtering, an accurate dedicated method to quantify AID-induced substitutions in B-lineage cells (50). **A)** SHM quantification upstream from *IgH S $\mu$*  region. **B)** SHM quantification within the *IgH* intron downstream from *VJ558* (consensus) to *J<sub>H4</sub>*-rearranged exons (in spontaneous GC B cells) or in the same region downstream from *V186.2* to *J<sub>H4</sub>*-rearranged exons (in NP-CGG-induced GC B cells). **C)** SHM quantification within the *Igk* intron downstream from all-rearranged *Jk5* segments. p-value was determined by two tailed Mann Whitney test and significant differences were indicated by: \* P<0.05; \*\* P<0.01; \*\*\* P<0.001; \*\*\*\* P<0.0001; error bars represent SEM.

Fig. S1



**Figure S1.** (A) Representation of SATB1 relative expression in murine *wt* B cell populations with Gene Skyline data browser (ImmGen ULI RNASeq data group, [www.immgen.org](http://www.immgen.org)). Populations include Common Lymphoid progenitor, Hardy fractions A, B, C, E and mature peripheral B cell subsets. (B) Transcription of *Rag1* and *Rag2* genes quantified in bone marrow sorted pre-B cells from WT (n=6) and *Satb1* cKO (n=5) mice. (C) Comparison of *IgHV* gene relative proportion, quantified by RACE-RepSeq, expressed by pre-B cell-enriched bone marrow populations in WT (n=5) and cKO mice (n=6). (D) Comparison of CDR3 length (left panel), N and P-nucleotide addition (right panel) during VDJ recombination for WT (n=5) and cKO mice (n=6). p-value was determined with two tailed Mann Whitney test.

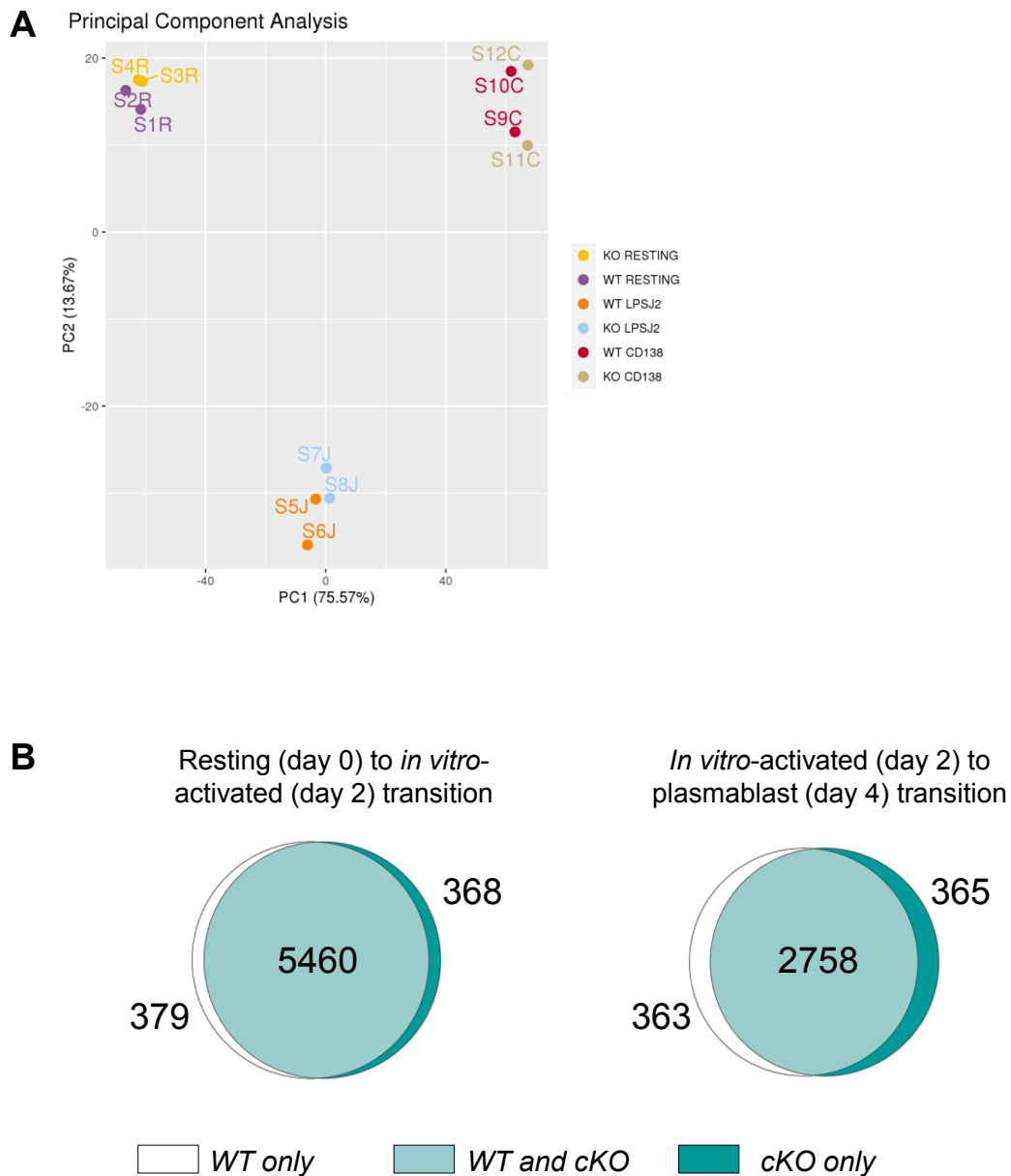
Fig. S2



**Fig. S2.** Mean Fluorescence Intensity measured by FACS analysis (normalized on WT samples) of IgM expressed at the surface of B cell subsets of 2-month-old WT and cKO heterozygous mice. Left panel includes data from bone marrow IgM expressing B cell subsets (n=7-9). Right panel includes data from B cell subset from the spleen (n=12-13). p-value was determined with Mann Whitney test, significant differences were indicated by: \* P<0.05; \*\* P<0.01; \*\*\* P<0.001; \*\*\*\* P<0.0001; error bars represent SEM

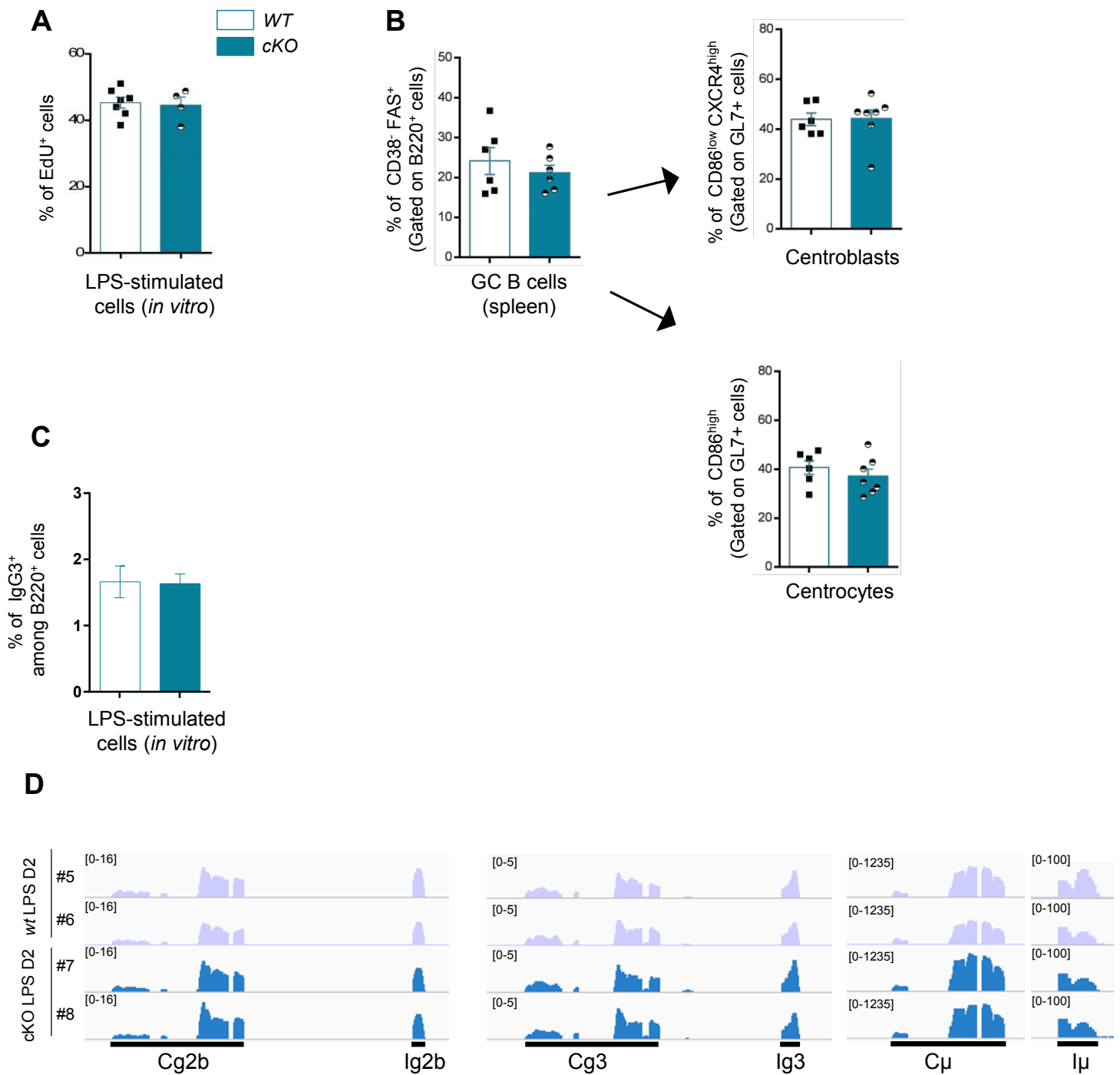


Fig. S3



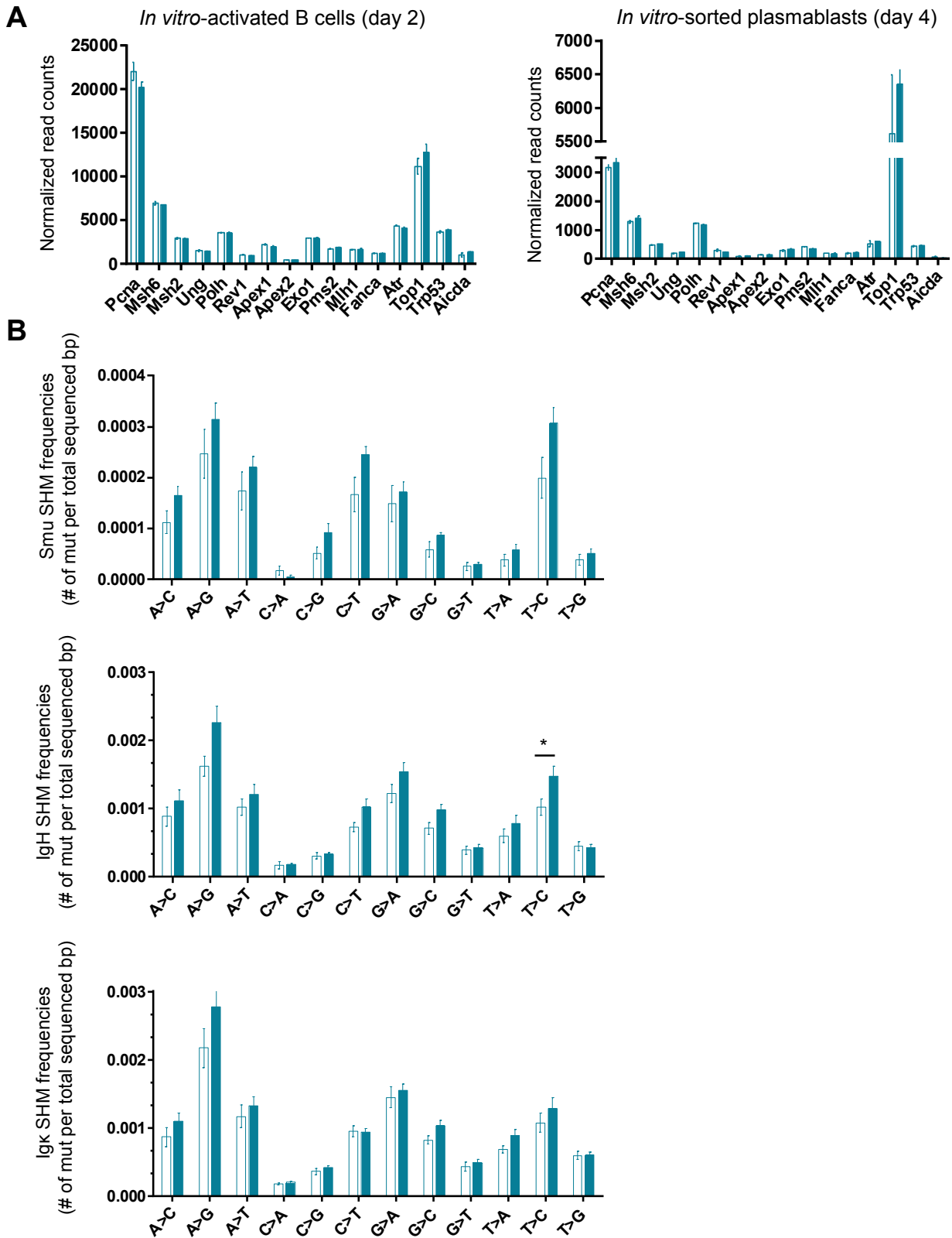
**Fig. S3. (A)** Principal Component Analysis of RNA sequencing of resting (R), LPS stimulated (J) or plasmablast (C) cells from WT and cKO mice (n=2). **(B)** Venn diagrams from RNA-seq displaying numbers and proportions of differentially expressed transcripts of B-lineage cells from WT and cKO mice (n=2) during resting to *in vitro*-activated (day 2) transition (left) and during *in vitro*-activated to plasmablast (day 4) transition (right)

Fig. S4



**Fig. S4.** (A) Percentage of EdU-incorporating splenic B cells from WT and cKO mice during *in vitro* LPS activation (n=4-7). (B) Detailed analysis of germinal center (GC) populations obtained following NP-CGG immunization of WT (n=6) and cKO (n=7) animals: percentage of GC B cells (B220<sup>+</sup>GL7<sup>+</sup>, left) among splenic B cells, percentage of centroblasts (CD86<sup>Low</sup>CXCR4<sup>High</sup>, upper right) & centrocytes (CD86<sup>High</sup>, lower right) within the initial GC B cell subset. (C) IgG3 switched cells obtained after *in vitro* activation of splenic B cells with LPS at day 4. p-value was determined with two tailed Mann Whitney test. (D) Visual representation of relevant *IgH* constant gene transcripts induced in *in vitro* LPS-activated B cells (day 2) from WT and cKO mice (n=2). Bedgraph files from RNAseq analysis were aligned on mouse GRCm38/mm10 assembly using IGV TOOLS viewer.

Fig. S5



**Fig. S5. (A)** Relative expression of genes involved in SHM and SHM-associated DNA repair pathways in *in vitro* activated B cell- (day 2) or enriched-plasma cell- (day 4) subsets of WT and cKO mice evaluated by RNA-seq (n=2). **(B)** Base substitution patterns, reported in frequencies for each base, in GC B cells sorted from Peyer's patches of WT and cKO mice. SHM taking place within the region upstream from *IgH*  $S\mu$  (top), downstream from rearranged- $J_H4$  exons (middle) and downstream from rearranged- $Jk5$  exons (bottom) (n=7-10). Data were obtained by NGS (Ion proton) combined to DeMinEr filtering. p-value was determined with two tailed Mann Whitney test, significant differences were indicated by: \* P<0.05; \*\* P<0.01; \*\*\* P<0.001; \*\*\*\* P<0.0001; error bars represent SEM.

*Supplementary Table S1: List of differentially expressed genes in resting B cells*

Genes downregulated in resting B cells						
MGI SYMBOL	WT	cKO	FoldChange	log2FoldChange	pvalue	padj
Hspa1a	188896	122130	0.647	-0.629	4.58040337476548e-6	0.00191077385433822
Hspd1	49936	32268	0.646	-0.63	5.29346297269999e-7	2.9673168376341e-4
Tnf	1570	1008	0.642	-0.64	9.48629703355245e-5	0.0215398982516283
Snord118	627	392	0.625	-0.679	2.595515118942e-4	0.0484982814620647
Hspe1	33980	20312	0.598	-0.742	8.0046597154752e-7	4.1025024564626e-4
Ccr12	1730	998	0.576	-0.796	5.81615734952946e-6	0.00231844956746354
Oasl1	2640	1502	0.569	-0.814	4.34376907467054e-6	0.00185520308717715
Rnu5g	502	282	0.558	-0.842	1.734644340325e-4	0.0366071178550015
Map3k19	765	416	0.546	-0.873	2.03195896221967e-5	0.00662714179350843
Bag3	1758	955	0.543	-0.881	3.20931750492121e-6	0.00143921843508192
Xlr3b	754	276	0.367	-1.447	1.74137977854962e-8	1.35812480294013e-5
Ifi208	826	141	0.172	-2.542	6.32098751900608e-24	2.2677174823186198e-20

Genes upregulated in resting B cells						
MGI SYMBOL	WT	cKO	FoldChange	log2FoldChange	pvalue	padj
Abca6	34	119	3.443	1.784	8.34323187989419e-6	0.00305430394819474
Abca8b	52	158	3.075	1.621	2.99174456038325e-6	0.00137604907497833
Abcg1	3060	4800	1.569	0.65	2.39080373574238e-5	0.00739417886409428
Adk	420	826	1.963	0.973	4.11208870821991e-8	2.95050588992195e-5
Arhgap5	566	959	1.699	0.765	1.4598619356449e-4	0.0311750040495237
Atp6v0c-ps2	20	182	9.124	3.19	9.38486254275277e-16	1.5304151299263598e-12
Ctse	744	4444	5.974	2.579	9.73537288912741e-73	8.731655944258379e-69
Ctso	1906	3132	1.644	0.717	6.89995361480135e-8	4.58412473860395e-5
Cxcr4	16384	24590	1.501	0.586	4.17674973187954e-7	2.4974178896818e-4
Cyp27a1	312	640	2.054	1.038	1.96906125445597e-7	1.2179662338769e-4
D16Ert472e	1156	2074	1.793	0.843	2.612311866037e-8	1.95248542720715e-5
Dnah8	799	2186	2.732	1.45	1.4451414451258602e-15	2.16024560355564e-12
Dyrk3	286	494	1.726	0.787	1.2085893782092e-4	0.026765032427553
Eya1	342	653	1.908	0.932	7.95376312710286e-7	4.1025024564626e-4
Gnb4	118	354	2.987	1.579	3.95858213616336e-10	3.38138315992849e-7
Hdac9	2135	3452	1.616	0.693	3.56503696134364e-6	0.00155974714664835
Hepacam2	320	894	2.793	1.482	4.3377055635727397e-13	4.57704484702164e-10
Hmga1b	7	153	22.761	4.508	3.91084462440521e-16	7.62067242546415e-13
Igha	843	2421	2.873	1.522	9.67511091545799e-28	4.33880349003714e-24
Ighg2b	593	1632	2.755	1.462	5.56283808835061e-20	1.42551699469762e-16
Ighg2c	317	1088	3.437	1.781	5.3227352877163604e-14	5.967451599441e-11
Igkv6-17	500	810	1.621	0.697	2.5144335575027e-4	0.0475590988698733
Islr2	90	264	2.914	1.543	1.48135847141786e-7	9.49021723581916e-5
Kcnh6	4	150	32.902	5.04	1.8605303978737203e-15	2.56724571361991e-12
Kif18a	1568	2579	1.644	0.717	1.17474921008283e-6	5.6953111703961e-4
Klf4	7503	11306	1.507	0.592	1.3516880439727e-4	0.0295690001619308
Lgals3	174	333	1.922	0.943	5.29357234601036e-5	0.013761753730831
Lmna	1108	1801	1.625	0.701	4.32716357675586e-5	0.0115851731701264
Ltk	402	796	1.988	0.991	5.94914869933211e-6	0.00231990933410042
Myadm	1032	3292	3.191	1.674	8.2306343006773e-32	4.92137060285164e-28
Nampt	7695	16368	2.127	1.089	4.19944119758951e-14	5.02197174682404e-11
Nap1l3	26	99	3.732	1.9	2.07392020698267e-5	0.00664321083443843
Nrip1	1115	2313	2.075	1.053	1.38316010726254e-12	1.3783958891153e-9
P2ry10	12306	18726	1.522	0.606	4.21001126677799e-5	0.0114971040042835
Pira2	42	137	3.234	1.693	9.55066224216315e-6	0.00335921135882201
Prf1	227	446	1.966	0.975	5.33331082939172e-6	0.0021742938558552
Prkca	721	1170	1.622	0.698	3.90387775650536e-5	0.0109418373744052
Prkcg	205	522	2.537	1.343	1.64034911064582e-10	1.47122911733823e-7
Prss41	4	48	11.615	3.538	7.99628982718875e-6	0.00298828014416899
Rfk	2376	3687	1.552	0.634	5.89754438873763e-7	3.2057621589447e-4
Rgcc	864	1410	1.634	0.708	1.4458483240162e-4	0.0311750040495237
Rnps1-ps	16	164	10.114	3.338	2.23727544966743e-15	2.86658907258102e-12
Ryr2	448	3638	8.133	3.024	5.54885820139486e-101	9.953541841662109e-97
Sestd1	19	110	5.761	2.526	5.04666230844608e-8	3.48180878803484e-5
Sgk3	1817	2748	1.513	0.598	2.29052817756475e-5	0.00720833235950115
Spp1	130	268	2.053	1.037	6.14590493314083e-5	0.0154680295637212
Stt3b	4927	8259	1.676	0.745	2.81046981964686e-9	2.29155489203752e-6
Sult2b1	15	73	4.733	2.243	3.12371572585334e-5	0.00889416074450114
Syne1	2812	5927	2.108	1.076	1.65526094327379e-16	3.71150885005564e-13
Tagap1	896	2046	2.284	1.191	4.24834007440303e-16	7.62067242546415e-13
Thns1	304	508	1.676	0.745	2.51873920874e-4	0.0475590988698733
Uchl3	1342	2130	1.587	0.666	4.79346376607084e-5	0.012644875446438
Zfp619	312	564	1.805	0.852	3.07741978779807e-5	0.00889416074450114



GO:0019732	GO:0030451	GO:0002639	GO:0002377	GO:0002639	GO:0048294	GO:0002377	GO:0002377	GO:0016446	GO:0002377
GO:0001867	GO:0006956	GO:0045830	GO:0002381	GO:0048295	GO:0048302	GO:0002639	GO:0002381	GO:0002204	GO:0048289
GO:0006958	GO:0030449	GO:0002638	GO:0002377	GO:0045830	GO:0045191	GO:2000572	GO:0002344	GO:0045190	GO:0048294
GO:0001867	GO:0006956	GO:0045190	GO:0048298	GO:0045190	GO:0002426	GO:0045830	GO:0002381	GO:0016447	GO:0002377
GO:0019731	GO:0006957	GO:0002377	GO:0002639	GO:0048304	GO:0002377	GO:0048304	GO:0002377	GO:0016446	GO:0045190
GO:0001867	GO:0061844	GO:0002637	GO:0002377	GO:0002377	GO:0033152	GO:0002377	GO:0016446	GO:0002381	GO:0002638
GO:0061844	GO:0001867	GO:0048304	GO:0002638	GO:0002381	GO:0002381	GO:0002639	GO:0002377	GO:0002377	GO:0002639
GO:0006958	GO:0061844	GO:0002377	GO:0002377	GO:0002639	GO:0002377	GO:0002377	GO:0002637	GO:0002637	GO:0002377
GO:0019731	GO:0002923	GO:0002637	GO:0002208	GO:0002638	GO:0002639	GO:0048302	GO:0002377	GO:0002381	GO:0045190
GO:0061844	GO:0002925	GO:0048304	GO:0045830	GO:0016446	GO:0045830	GO:0045830	GO:0045190	GO:0002377	GO:0002377
GO:0019731	GO:0030449	GO:0002377	GO:0002377	GO:0002637	GO:0045190	GO:0002208	GO:0002377	GO:0002638	GO:0016446
GO:0061844	GO:0045957	GO:0002639	GO:0045190	GO:0016445	GO:0002639	GO:0045830	GO:0002381	GO:0002639	GO:0033152
GO:0019731	GO:0045959	GO:0048304	GO:0002208	GO:0002377	GO:0002377	GO:0048298	GO:0002638	GO:0002638	GO:0045191
GO:0061844	GO:0045957	GO:0002639	GO:0045830	GO:0045190	GO:0002639	GO:0048304	GO:0002639	GO:0002377	GO:0002377
GO:0019731	GO:0061844	GO:0002638	GO:0002377	GO:0045830	GO:0048304	GO:0002204	GO:0002377	GO:0016446	GO:0002639
GO:0006959	GO:0002786	GO:0033152	GO:0016446	GO:0002381	GO:0048295	GO:0016447	GO:0016445	GO:0048298	GO:0002377
GO:0019731	GO:0006959	GO:0002377	GO:0016447	GO:0016446	GO:0048304	GO:0045190	GO:0016446	GO:0048304	GO:0002344
GO:0061844	GO:0061844	GO:0033152	GO:0045190	GO:0002637	GO:0048295	GO:0016447	GO:0045190	GO:0016446	GO:0002377
GO:0019731	GO:0019731	GO:0045190	GO:0016447	GO:0002381	GO:0002639	GO:0016446	GO:0002639	GO:0016447	GO:0002639
GO:0002760	GO:0061844	GO:0033152	GO:0002377	GO:0002638	GO:0048304	GO:0016447	GO:0048302	GO:0016446	GO:0048298
GO:0002779	GO:0006959	GO:0048290	GO:0045191	GO:0033152	GO:0002639	GO:0002204	GO:0048294	GO:0002377	GO:0071707
GO:0002780	GO:0019731	GO:0016446	GO:0045190	GO:0045190	GO:0048289	GO:0016446	GO:0048289	GO:0048298	GO:0002377
GO:0006958	GO:0006958	GO:0045190	GO:0045830	GO:0033152	GO:0048291	GO:0045190	GO:0002637	GO:0002426	GO:0045830
GO:0061844	GO:0019731	GO:0016447	GO:0016446	GO:0002377	GO:0002639	GO:0016446	GO:0002639	GO:0002381	GO:0002377
GO:0006959	GO:0019732	GO:0016446	GO:0002381	GO:0045190	GO:0048304	GO:0002639	GO:0048304	GO:0002377	GO:0048304
GO:0061844	GO:0061844	GO:0045190	GO:0002377	GO:0045830	GO:0002639	GO:0002377	GO:0002377	GO:0045190	GO:0002377
GO:0006957	GO:0006958	GO:0002377	GO:0016447	GO:0048302	GO:0002377	GO:0048298	GO:0016445	GO:0016446	GO:0002381
GO:0019731	GO:0030449	GO:0002638	GO:0016446	GO:0002377	GO:0002637	GO:0002639	GO:0071707	GO:0002377	GO:0002377
GO:0061844	GO:0048289	GO:0002377	GO:0002377	GO:0048297	GO:0002344	GO:0002377	GO:0016446	GO:0002639	GO:0002639
GO:0019731	GO:0048295	GO:0002381	GO:0071707	GO:0002377	GO:0016446	GO:2000558	GO:0002377	GO:0045829	GO:0002377
GO:0006956	GO:0002377	GO:0048304	GO:0048304	GO:0002208	GO:0002377	GO:0002377	GO:0048298	GO:0002637	GO:0002319
GO:0030449	GO:0033152	GO:0002344	GO:0016446	GO:0045830	GO:2000558	GO:0002381	GO:0048304	GO:0048291	GO:0002906

Supplementary Table S3: List of differentially expressed genes in in vitro-activated B cells (day 2)

Genes downregulated in LPS activated B cells (Day2)						
MGI SYMBOL	WT	cKO	FoldChange	log2FoldChange	pvalue	padj
Tbc1d23	1493	992	0.665	-0.589	1.4379663512243e-4	0.0154285330860781
Igf2bp3	316	179	0.567	-0.818	1.9768790415222e-4	0.018342053761529
St6galnac2	442	246	0.557	-0.844	2.26174031566037e-5	0.00352169516467702
Oasl1	9256	4244	0.458	-1.125	9.28749197296647e-37	1.97637829184726e-33
BC018473	1640	689	0.421	-1.25	4.44834189667717e-19	5.67964293367741e-16



Genes upregulated in LPS activated B cells (Day2)						
MGI SYMBOL	WT	cKO	FoldChange	log2FoldChange	pvalue	padj
ApoE	298	540	1.816	0.861	1.81573739597255e-5	0.00289791688397219
Arhgap31	16	64	4.088	2.031	1.9870840560962e-4	0.018342053761529
Ctse	162	1482	9.113	3.188	1.57292762828546e-82	5.0207849894871706e-79
Gdgd3	248	444	1.796	0.844	4.9041339772894e-4	0.0361942096081108
Gnb4	221	393	1.776	0.829	1.9871185085713e-4	0.018342053761529
Hepacam2	313	606	1.944	0.959	4.49632763275736e-6	9.4113297073845e-4
Igha	83	176	2.12	1.084	3.6973674915831e-4	0.0289619559095296
Ighv1-12	361	714	1.986	0.99	5.06951090721622e-7	1.4710798923485e-4
Ighv1-15	1508	2324	1.541	0.624	7.40335530853516e-8	2.70074401655363e-5
Ighv1-4	493	899	1.825	0.868	3.53546764360655e-8	1.41065158979901e-5
Ighv1-5	324	606	1.874	0.906	4.0459821172115e-6	8.609849945426e-4
Ighv1-59	574	876	1.525	0.608	2.9926435400817e-4	0.0248117355323144
Ighv1-63	144	269	1.869	0.902	5.9253469477392e-4	0.0414736629072173
Ighv1-64	2602	3972	1.527	0.611	1.8496408629221902e-9	1.18081072688953e-6
Ighv1-66	538	838	1.558	0.64	1.26166319526729e-5	0.0022067007776949
Ighv1-72	1966	2973	1.513	0.597	2.40844660273244e-6	5.3949203901206e-4
Ighv1-74	523	948	1.813	0.858	1.22429770556603e-8	6.51326379361127e-6
Ighv1-75	1150	1776	1.546	0.628	4.81417058761007e-9	2.67249261141762e-6
Ighv1-78	734	1161	1.585	0.664	5.53927069622126e-5	0.00693386355385815
Ighv1-81	1140	1819	1.595	0.674	1.72190560641577e-8	8.34493367916676e-6
Ighv2-9-1	620	977	1.576	0.657	1.61387954854543e-5	0.00267610572413351
Ighv4-1	471	708	1.503	0.588	3.3048049394524e-4	0.0263723434168308
Ighv5-12	371	598	1.615	0.692	1.6879600372217e-4	0.0171046617105132
Ighv5-16	1089	2268	2.084	1.059	4.77433382396421e-13	4.68913032802885e-10
Ighv5-9-1	660	1096	1.661	0.732	3.49389624702796e-5	0.00474575183851628
Ighv7-1	268	586	2.192	1.132	8.74210264846859e-8	3.01984854471224e-5
Ighv8-12	1012	1624	1.605	0.682	5.79361119675743e-7	1.5738899523446e-4
Igkv15-103	884	1446	1.637	0.711	1.22433311217471e-5	0.00217115071892316
Igkv17-121	1697	2604	1.534	0.618	1.13783903477559e-6	2.8486134894146e-4
Igkv17-127	2162	3536	1.636	0.71	6.573807157500919e-11	4.93731586982187e-8
Igkv3-12	1044	1871	1.793	0.842	3.83146231752191e-9	2.2364140318726e-6
Igkv3-7	706	1224	1.734	0.794	2.8455666007439504e-9	1.73010449325232e-6
Igkv4-55	876	1326	1.512	0.597	6.54042596310273e-6	0.00128474090302916
Igkv4-57	1073	2028	1.891	0.919	3.3562498027467997e-12	2.8568398320980804e-9
Igkv4-57-1	442	742	1.682	0.75	7.75445977469986e-6	0.00147774540900549
Igkv4-59	1852	2888	1.56	0.641	4.83655880379133e-8	1.81627008255317e-5
Igkv4-61	202	382	1.895	0.922	1.60994968749444e-5	0.00267610572413351
Igkv4-63	412	664	1.612	0.689	9.0568117075039907e-005	0.0103247653465545
Igkv4-68	765	1308	1.711	0.775	7.69625556942995e-7	2.0054243083771e-4
Igkv5-39	1219	2002	1.643	0.717	1.26271126979818e-7	4.13392243404698e-5
Igkv5-45	468	718	1.533	0.616	5.9442984899912e-4	0.0414736629072173
Igkv5-48	1414	2136	1.51	0.594	1.50238092173931e-7	4.79559990219186e-5
Igkv6-13	486	748	1.539	0.622	3.57710546180523e-5	0.00480762974066624
Igkv6-14	122	282	2.299	1.201	5.46043365617278e-7	1.5493070427114e-4
Igkv6-15	2835	4390	1.549	0.632	1.85196308197933e-8	8.34493367916676e-6
Igkv6-17	1384	2542	1.837	0.877	8.67158226982118e-14	1.00653420382797e-10
Igkv6-23	1326	2087	1.575	0.655	8.75112751835471e-8	3.01984854471224e-5
Igkv6-25	514	814	1.582	0.661	3.15389222964394e-5	0.00442515340528504
Igkv8-19	1282	1964	1.532	0.616	6.14025172420464e-6	0.00122498021897883
Igkv8-21	657	1012	1.542	0.625	5.52356193294069e-5	0.00693386355385815
Il9r	221	524	2.371	1.245	1.0811900504023099e-11	8.6278966022104e-9
Lax1	2834	4282	1.511	0.596	4.67891232373424e-7	1.3893105244055e-4
Maged1	292	498	1.708	0.773	1.65786385160679e-5	0.00271379559709174
Myadm	164	371	2.254	1.173	4.74959906327652e-8	1.81627008255317e-5
Oasl2	72	588	8.259	3.046	6.047550256581949e-45	1.54430243352077e-41
Pde8a	537	848	1.58	0.66	1.0115348773303e-5	0.00189930548731666
Pon3	561	854	1.524	0.608	4.2294877988406e-4	0.0323365869554472
Prdm16	234	400	1.71	0.774	8.25736604363825e-5	0.00949820267073632
Ryr2	78	296	3.798	1.925	2.64953991271621e-13	2.81911046713004e-10
Selplg	1664	2546	1.53	0.614	1.85267444878709e-8	8.34493367916676e-6
Sord	515	775	1.502	0.587	2.2488993993367e-4	0.0199402413407862
Spp1	86	228	2.622	1.391	1.55222797931981e-6	3.8113166999914e-4
Tram2	646	982	1.52	0.604	7.25890543153912e-5	0.00858163931017513
Ubc	3298	5570	1.688	0.756	9.853193214380549e-20	1.39783967734679e-16

*Supplementary Table S4: List of differentially expressed genes in in vitro-differentiated plasmablasts (day 4)*

Genes upregulated in plasmablasts (Day4)						
MGI SYMBOL	WT	cKO	FoldChange	log2FoldChange	pvalue	padj
Iglv3	18404	38273	2.079	1.056	1.79921733414964e-12	2.31613247425083e-8
Rps7-ps3	1038	2096	2.02	1.015	4.144740942877341e-9	1.3338812539415e-5
Ighv3-8	6327	11801	1.865	0.899	2.16216554785762e-5	0.0397622244251016
Reep5	3030	5507	1.817	0.861	3.29776904916066e-5	0.0471690899664946
Glb1	1322	2165	1.638	0.712	3.02476519318689e-5	0.0471690899664946
Atf5	412	745	1.805	0.852	3.71673509239989e-5	0.0478455308444638

Supplementary Table S5: Total number of mutations, total number of bp analyzed and mutation frequencies within intron 5' to  $\mu$  (A, B); 3' to  $J_{H4}$  (C, D) and 3' to  $J_{\kappa 5}$  (E, F) in GC B cells of wt and *Satb1* cKO. For each region tested for SHM, data were obtained from spontaneous GC B cells sorted from Peyer's patches (n=6-9 individual mice) or splenic GC B cells sorted upon NP-CGG-immunization (n=7-8 individual mice).

**A**

Intron 5' to Smu Peyer's patches GC							
<i>Satb1</i> <sup>flx/+</sup>				<i>Satb1</i> <sup>flx/flx</sup> <i>Cd79a</i> <sup>cre/+</sup>			
N°	number of mutations	total number of pb analyzed	Frequency (mutation/Kb)	N°	number of mutation	total number of pb analyzed	Frequency (mutation/Kb)
<b>8692</b>	167 611	143 504 465	<b>1,17</b>	<b>8690</b>	151 915	100 956 853	<b>1,50</b>
<b>8649</b>	45 504	48 022 564	<b>0,95</b>	<b>8645</b>	54 050	38 428 176	<b>1,40</b>
<b>8753</b>	143 998	96 361 943	<b>1,50</b>	<b>8756</b>	311 660	148 421 088	<b>2,10</b>
<b>8755</b>	182 738	122 821 454	<b>1,49</b>	<b>9601</b>	144 996	73 409 771	<b>1,98</b>
<b>8644</b>	109 695	63 612 833	<b>1,72</b>	<b>9604</b>	163 545	78 032 498	<b>2,10</b>
<b>7128</b>	250 366	152 000 328	<b>1,65</b>	<b>9877</b>	162 887	85 164 912	<b>1,91</b>
<b>7129</b>	176 860	139 941 043	<b>1,26</b>	<b>9880</b>	113 784	87 909 626	<b>1,29</b>
<b>9881</b>	152 289	86 749 376	<b>1,76</b>	<b>7111</b>	485 120	187 778 619	<b>2,58</b>
				<b>7112</b>	252 555	121 402 038	<b>2,08</b>
<b>Total</b>	1 229 061	853 014 006	<b>1,44</b>	<b>Total</b>	1 840 512	921 503 581	<b>1,88</b>

**B**

Intron 5' to Smu NP CGG Splenic GC							
<i>Satb1</i> <sup>flx/+</sup>				<i>Satb1</i> <sup>flx/flx</sup> <i>Cd79a</i> <sup>cre/+</sup>			
N°	number of mutations	total number of pb analyzed	Frequency (mutation/Kb)	N°	number of mutation	total number of pb analyzed	Frequency (mutation/Kb)
<b>8139</b>	19 700	111 850 026	<b>0,18</b>	<b>6854</b>	26 279	142 403 067	<b>0,19</b>
<b>6858</b>	48 673	211 877 661	<b>0,23</b>	<b>6857</b>	56 795	146 983 925	<b>0,39</b>
<b>6859</b>	8 202	129 181 296	<b>0,06</b>	<b>6864</b>	30 662	52 155 066	<b>0,59</b>
<b>6863</b>	7 520	132 861 196	<b>0,06</b>	<b>6687</b>	106 971	229 058 581	<b>0,47</b>
<b>6692</b>	25 891	179 825 690	<b>0,14</b>	<b>6691</b>	101 092	256 700 586	<b>0,39</b>
<b>6693</b>	109 022	217 449 208	<b>0,50</b>	<b>6791</b>	33 527	118 595 323	<b>0,28</b>
<b>2544</b>	67 300	178 614 933	<b>0,38</b>	<b>2537</b>	112 115	197 775 331	<b>0,57</b>
<b>Total</b>	286 308	1 161 660 010	<b>0,22</b>	<b>Total</b>	467 441	1 143 671 879	<b>0,41</b>

C

Intron 3' to J <sub>H</sub> 4 Peyer's patches GC							
Satb1 <sup>flx/+</sup>				Satb1 <sup>flx/flx</sup> Cd79a <sup>cre/+</sup>			
N°	number of mutations	total number of pb analyzed	Frequency (mutation/Kb)	N°	number of mutation	total number of pb analyzed	Frequency (mutation/Kb)
<b>8692</b>	57 561	7 906 004	<b>7,28</b>	<b>8690</b>	621 327	82 879 451	<b>7,50</b>
<b>8649</b>	654 842	100 297 965	<b>6,53</b>	<b>8645</b>	328 574	32 909 115	<b>9,98</b>
<b>8753</b>	4 064 096	339 723 680	<b>11,96</b>	<b>8756</b>	106 623	13 743 216	<b>7,76</b>
<b>8755</b>	218 499	26 244 087	<b>8,33</b>	<b>9601</b>	1 718 084	140 896 971	<b>12,19</b>
<b>8644</b>	136 754	10 382 456	<b>13,17</b>	<b>9604</b>	1 866 361	127 867 165	<b>14,60</b>
<b>8646</b>	475 950	58 858 803	<b>8,09</b>	<b>9605</b>	2 112 619	142 861 789	<b>14,79</b>
<b>9881</b>	195 215	21 355 525	<b>9,14</b>	<b>9661</b>	2 550 716	146 382 941	<b>17,42</b>
				<b>9877</b>	757 522	57 471 525	<b>13,18</b>
				<b>9880</b>	1 614 250	119 121 342	<b>13,55</b>
<b>Total</b>	5 802 917	564 768 520	<b>9,21</b>	<b>Total</b>	10 061 826	745 012 173	<b>12,33</b>

D

Intron 3' to J <sub>H</sub> 4 NP CCGG Splenic GC							
Satb1 <sup>flx/+</sup>				Satb1 <sup>flx/flx</sup> Cd79a <sup>cre/+</sup>			
N°	number of mutations	total number of pb analyzed	Frequency (mutation/Kb)	N°	number of mutation	total number of pb analyzed	Frequency (mutation/Kb)
<b>8139</b>	152 171	39 872 610	<b>3,81</b>	<b>8133</b>	137 111	28 459 587	<b>4,82</b>
<b>6858</b>	84 367	38 653 923	<b>2,18</b>	<b>6854</b>	252 934	92 925 446	<b>2,72</b>
<b>6859</b>	125 264	78 908 567	<b>1,59</b>	<b>6857</b>	141 510	37 358 940	<b>3,79</b>
<b>6863</b>	153 921	58 613 945	<b>2,62</b>	<b>6864</b>	335 616	40 332 921	<b>8,32</b>
<b>6692</b>	317 757	68 343 790	<b>4,65</b>	<b>6687</b>	777 955	152 063 704	<b>5,12</b>
<b>6693</b>	270 476	59 540 608	<b>4,55</b>	<b>6691</b>	551 228	96 233 795	<b>5,73</b>
<b>2544</b>	151 910	41 648 331	<b>3,64</b>	<b>6791</b>	249 260	54 111 722	<b>4,61</b>
				<b>2537</b>	176 658	32 694 946	<b>5,40</b>
<b>Total</b>	1 255 866	385 581 774	<b>3,29</b>	<b>Total</b>	2 622 272	534 181 061	<b>5,06</b>
<b>Total</b>	1 255 866	385 581 774	<b>3,29</b>	<b>Total</b>	2 622 272	534 181 061	<b>5,06</b>

## E

Intron 3' to Jk5 Peyer's patches GC							
Satb1 <sup>flx/+</sup>				Satb1 <sup>flx/flx</sup> Cd79a <sup>cre/+</sup>			
N°	number of mutations	total number of pb analyzed	Frequency (mutation/Kb)	N°	number of mutation	total number of pb analyzed	Frequency (mutation/Kb)
<b>8692</b>	64 352	8 109 925	<b>7,94</b>	<b>8690</b>	977 250	109 513 034	<b>8,92</b>
<b>8649</b>	187 594	18 973 865	<b>9,89</b>	<b>8645</b>	133 351	11 622 636	<b>11,47</b>
<b>8753</b>	41 647	2 737 783	<b>15,21</b>	<b>8756</b>	1 092 197	76 790 583	<b>14,22</b>
<b>8755</b>	2 497 183	288 443 955	<b>8,66</b>	<b>9601</b>	1 097 124	100 505 774	<b>10,91</b>
<b>8644</b>	400 715	30 094 611	<b>13,32</b>	<b>9604</b>	1 941 371	110 394 569	<b>17,59</b>
<b>9881</b>	958 245	97 655 675	<b>9,81</b>	<b>9605</b>	779 895	62 212 383	<b>12,54</b>
<b>7128</b>	1 571 163	146 409 208	<b>10,73</b>	<b>9877</b>	820 899	65 692 001	<b>12,50</b>
<b>7129</b>	540 065	97 174 751	<b>5,56</b>	<b>9880</b>	2 136 606	163 662 781	<b>13,05</b>
				<b>7111</b>	963 266	90 045 963	<b>10,70</b>
				<b>7112</b>	1 256 410	223 536 225	<b>6,07</b>
<b>Total</b>	6 260 964	689 599 773	<b>10,14</b>	<b>Total</b>	6 842 087	536 730 980	<b>11,80</b>

## F

Intron 3' to Jk5 NP CCG Splenic GC							
Satb1 <sup>flx/+</sup>				Satb1 <sup>flx/flx</sup> Cd79a <sup>cre/+</sup>			
N°	number of mutations	total number of pb analyzed	Frequency (mutation/Kb)	N°	number of mutation	total number of pb analyzed	Frequency (mutation/Kb)
<b>8139</b>	222 655	85 014 498	<b>2,62</b>	<b>6854</b>	171 154	69 778 869	<b>2,45</b>
<b>6858</b>	72 960	56 300 498	<b>1,30</b>	<b>6857</b>	271 677	110 895 960	<b>2,45</b>
<b>6859</b>	2 008	1 253 443	<b>1,60</b>	<b>6864</b>	280 371	67 161 110	<b>4,16</b>
<b>6863</b>	121 103	93 266 492	<b>1,30</b>	<b>6687</b>	371 949	94 961 964	<b>3,92</b>
<b>6692</b>	196 340	96 679 417	<b>2,03</b>	<b>6691</b>	249 263	71 580 665	<b>3,48</b>
<b>6693</b>	323 874	88 558 220	<b>3,66</b>	<b>6791</b>	247 551	102 051 591	<b>2,43</b>
2544	731 196	178 302 885	<b>4,10</b>	<b>2537</b>	622 428	91 007 705	<b>6,84</b>
<b>Total</b>	1 670 136	599 375 453	<b>2,37</b>	<b>Total</b>	2 214 393	607 437 864	<b>3,68</b>

*Supplementary Table S6: summary table of genotyping primers*

<b>Primer name</b>	<b>Sequence</b>	<b>Gene</b>
<b>EF</b>	5'-TTTGCTCATGTGGAATGTCGAGGTA-3'	Satb1
<b>KR</b>	5'-GGGCAAGAACATAAAGTGACCCTC-3'C	Satb1
<b>EF2</b>	5'-AATAATCTGCTCCACTGCTCCACTGAGGACCCAC-3'	Satb1
<b>ER3</b>	5'-CCCTATTGCAGTGGGAATCAGCAT-3'	Satb1
<b>MB1 CRE Wild Type Forward</b>	5'-CTCTTTACCTTCCAAGCACTGA-3'	Mb1
<b>MB1 CRE COMMON</b>	5'-ACTGAGGCAGGAGGATTGG-3'	Mb1
<b>MB1 CRE MUTANT Forward</b>	5-'CATTTCGAGGGAGCTTCA-3'	Mb1

Supplementary Table S7: antibodies used for Western blot, ELISA and flow cytometry

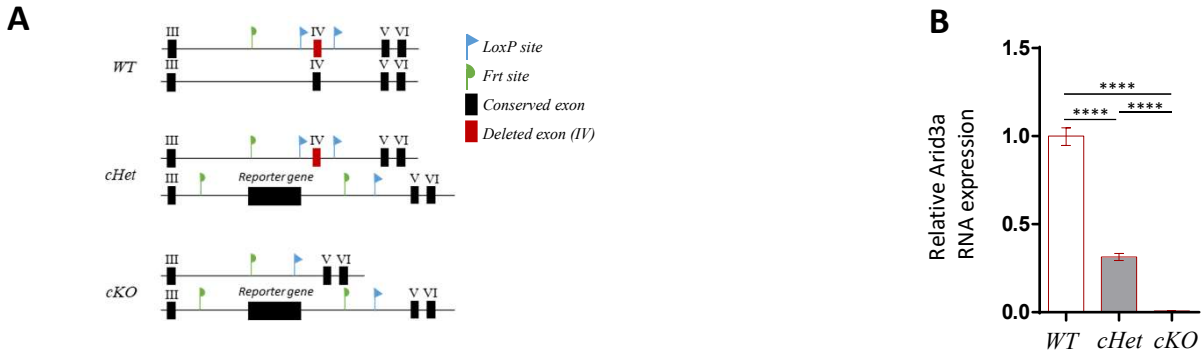
	Antibody	Clone	Dilution	Sources
WESTERN BLOT	SATB1	EPR3951	1/1000	ABCAM AB109122
	GAPDH	AF5718	1/5000	RD SYSTEMS
	Goat anti-rabbit	AB_2632593	1/5000	SOUTHERNBIOTECH
ELISA	Unlabeled Goat Anti Mouse IgM	1020-01	1/500	SOUTHERNBIOTECH
	Unlabeled Goat Anti Mouse IgG1	1070-01	1/500	SOUTHERNBIOTECH
	Unlabeled Goat Anti Mouse IgG3	1100-01	3/1000	SOUTHERNBIOTECH
	Unlabeled Goat Anti Mouse IgA	1040-01	1/250	SOUTHERNBIOTECH
	Goat Anti Mouse IgM-AP	1021-04	1/1000	SOUTHERNBIOTECH
	Goat Anti Mouse IgG1-AP	1070-04	1/1000	SOUTHERNBIOTECH
	Goat Anti Mouse IgG3-AP	1100-04	1/1000	SOUTHERNBIOTECH
	Goat Anti Mouse IgA-AP	1040-04	1/1000	SOUTHERNBIOTECH
Flow cytometry BONE MARROW B CELLS	cKIT BV421	2B8	1/100	BECTON DICKINSON 562609
	B220 BV510	RA3-6B2	1/100	BIOLEGEND 103248
	CD25 APC	PC61	1/100	BECTON DICKINSON 561048
	CD24 FITC	M2/169	1/100	BECTON DICKINSON 553261
	CD43 PE	S7	1/100	BECTON DICKINSON PHARM 553271
	IGM PC7	II /41	1/100	EBIOSCIENCE 25-5790-82
	CD19 APCH7	1D3	1/100	BECTON DICKINSON PHARM 560143
Flow cytometry SPLENIC B CELLS	CD5 FITC	53-7.3	1/100	BECTON DICKINSON 5530211
	CD19 APCH7	1D3	1/100	BECTON DICKINSON PHARM 560143
	CD23 PC7	B3B4	1/100	BIOLEGEND 101614
	CD21 PE	7G6	1/100	BECTON DICKINSON PHARM 552957
	IGM APC	II /41	1/100	EBIOSCIENCE 17-2590-82
	IGD BV421	11-26C-2A	1/100	BIOLEGEND 405725
Flow cytometry PEYER 'S PATCHES CELLS	B220 APC	RA3-6B2	1/100	BIOLEGEND 103212
	GL7 FITC	GL7	1/100	BECTON DICKINSON 553666
Flow cytometry CENTROCYTE / CENTROBLAST	B220 BV510	RA3-6B2	1/100	BIOLEGEND 103248
	CD38 FITC	90	1/100	BECTON DICKINSON 558813
	FAS BV421	Jo2	1/100	BECTON DICKINSON 562633
	CXCR4 PE	2B11/CXCR4	1/50	BECTON DICKINSON PHARM 551966
	CD86 APC	GL1	1/50	BIOLEGEND 105011
Flow Cytometry <i>IN VITRO</i> STIMULATED CELLS	B220 BV421	RA3-6B2	1/100	BECTON DICKINSON 562922
	IGM PC7	II/41	1/100	EBIOSCIENCE 25-5790-82
	IgG3 FITC	R40-82	1/100	BECTON DICKINSON 553403
	CD138 APC	281-2	1/100	BECTON DICKINSON 558626

## Résultats

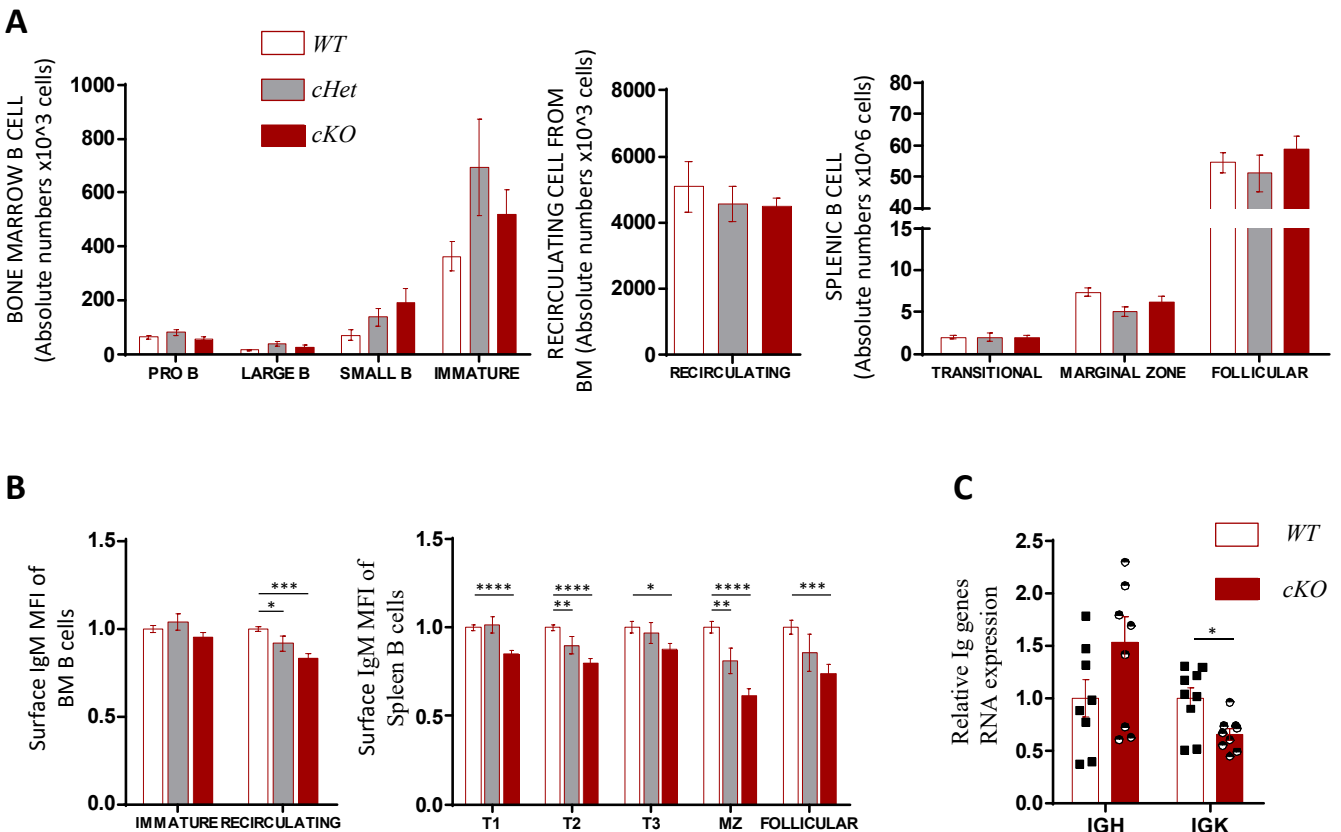
---

### **PARTIE II. Etude du rôle de la protéine ARID3A dans le lignage B**





**FIGURE 1 : A)** Conditional deletion in B cell of *Arid3a*-exon 4 by *Cd79a<sup>cre/+</sup>* recombinase. Exon 4, cre site, frt site, conditional allele (*Arid3a<sup>tm1c</sup>*) and Wild type allele (*Arid3a<sup>+</sup>*) are indicated. **B)** ARID3A transcript level in WT (*Arid3a<sup>tm1c/+</sup>*), conditional Heterozygote (*Arid3a<sup>tm1c/+</sup> Cd79a<sup>cre/+</sup>*) and conditional Homozygote (*Arid3a<sup>tm1c/tm1b</sup> Cd79a<sup>cre/+</sup>*) B cells. RT-qPCR was performed on B cells sorted from spleen. *p*-value was determined with Mann Whitney test, significant differences were indicated by : \* *P* < 0.05; \*\* *P* < 0.01; \*\*\* *P* < 0.001; \*\*\*\* *P* < 0,0001 and error bars represent SEM.



**FIGURE 2 : A)** Absolute numbers of Bone marrow B cells subset from one femur (n=6-15) and spleen (n=12-15) determined by Flow cytometry. **B)** Normalization of IgM MFI in Bone marrow (n=8-15) and spleen B cells subset (n=8-15). **C)** IgH and Igk transcript levels in WT (*Arid3a<sup>tm1c/+</sup>*) and conditional Homozygote (*Arid3a<sup>tm1c/tm1b</sup> Cd79a<sup>cre/+</sup>*) B cells. RT-qPCR was performed on B cells sorted from spleen. *p*-value was determined with Mann Whitney test, significant differences were indicated by : \* *P* < 0.05; \*\* *P* < 0.01; \*\*\* *P* < 0.001; \*\*\*\* *P* < 0,0001 and error bars represent SEM.

## **PARTIE II. Etude du rôle de la protéine ARID3A**

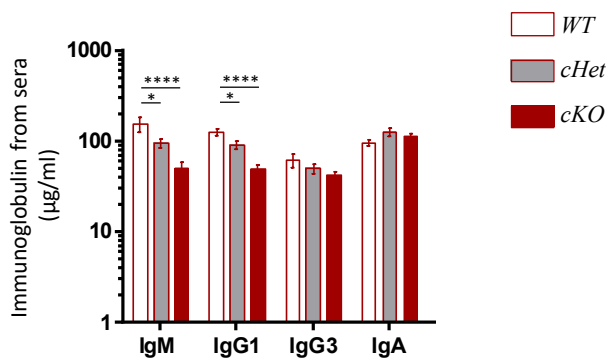
### **Description du modèle animal**

La délétion germinale de *Arid3A* induit un biais dans la différenciation des cellules souches hématopoïétiques et la production de cellules érythrocytaires, myéloïdes et lymphoïdes, conduisant à une létalité très précoce au stade embryonnaire (Webb et al., 2011). Nous avons donc réalisé notre étude sur un modèle de délétion conditionnelle dans la souris. Ce modèle combine un allèle *Arid3a* inactif, nommé *Arid3a<sup>Tm1b</sup>*, qui porte une délétion constitutive de l'exon 4 de façon germinale et un allèle conditionnel, nommé *Arid3a<sup>Tm1c</sup>*. Ce modèle, correspondant à une délétion hétérozygote chez les animaux, permet leur développement. La suppression de l'exon 4 du second allèle sera effectuée, quant à elle, de façon spécifique dans le lignage B par l'action de la recombinase CRE dès les stades précoces (apportée par un allèle *Cd79a-CRE*) (**Figure 1A**) (Hobeika et al., 2006). La fiabilité de la suppression de l'expression de ARID3A est évaluée par quantification des transcrits matures de son gène par RT-qPCR réalisée sur des échantillons d'ARN issus de cellules B triées de la rate des animaux WT et cKO. On observe une diminution d'un facteur 2 de la quantité de transcrits matures *Arid3a* chez les animaux hétérozygote et une absence totale de ces transcrits chez les animaux cKO (**Figure 1B**).

### **Impact de la suppression de ARID3A sur le développement B**

Plusieurs études préalables, établies sur des modèles transgéniques de surexpression et d'expression de dominant négatif, suggèrent une implication de ARID3A dans la génération des cellules B1 (Hayakawa et al., 2019; Nixon et al., 2008). Un modèle de KO conditionnel dans le lignage B, se distinguant de notre modèle par l'utilisation du seul allèle cKO *Arid3a<sup>Tm1c</sup>* (à l'état hétérozygote ou homozygote) associés à l'allèle *Cd79a-CRE* indique que la perte de l'un ou des deux allèles *Arid3a* induit conjointement et progressivement une augmentation des populations B2 et une réduction des populations B1 (Habir et al., 2017).

Nous avons analysé en détail par cytométrie en flux, les populations médullaires et spléniques de type B2 dans notre modèle en comparant les génotypes sauvages (*Arid3a<sup>Tm1c/+</sup> Cd79a<sup>-/+</sup>*), hétérozygotes (*Arid3a<sup>Tm1c/+</sup> Cd79a<sup>-CRE/+</sup>*) et cKO (*Arid3a<sup>Tm1c/b</sup> Cd79a<sup>-CRE/+</sup>*). Contrairement aux modèles décrits dans l'étude de Habir et al, nous n'observons pas de changement significatif dans le nombres de cellules composant les différents sous-compartiment B2 de la moelle et de la rate (**Figure 2A**). En revanche, les populations B issues de la moelle osseuse et de la rate de souris ARID3A cKO présentent toutes une diminution de l'expression du BCR (IgM) à leur surface (**Figure 2B**). Différentes études présentent ARID3A comme un activateur de la transcription de *IgH* mais font état d'un contexte spécifique qui inclut un segment V réarrangé donné (VHS107) et un environnement propice à l'activation par

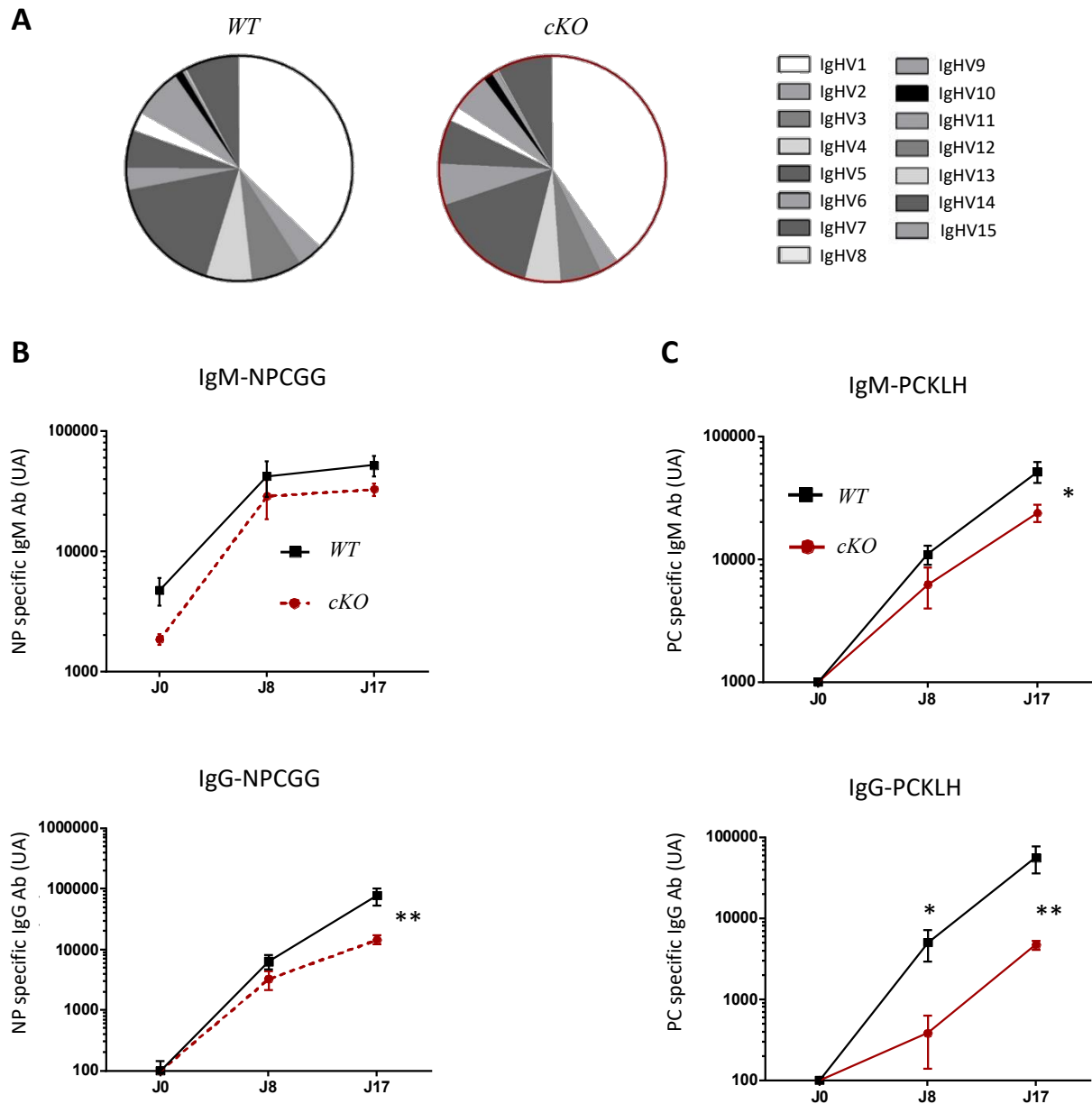


**Figure 3** : Immunoglobulin serum isotype levels quantified by Enzyme-Linked ImmunoSorbent Assay (ELISA) (n=11-13). *p-value* was determined with Mann Whitney test, significant differences were indicated by : \* P < 0.05; \*\* P < 0.01; \*\*\* P < 0.001; \*\*\*\* P < 0,0001 and error bars represent SEM.

l'enhancer  $E_{\mu}$  (Kaplan et al., 2001). Cependant, bien que l'ensemble des populations B « au repos » présente, dans notre modèle, une diminution de l'expression du BCR ; l'analyse des transcrits primaires IgH sur des cellules triées de la rate ne montre aucune diminution d'expression du gène en absence de ARID3A (**Figure 2C**). Ces résultats indiquent que la transcription des chaînes lourdes IgH n'est pas affectée par l'absence d'ARID3A dans les cellules B au repos. Obtenues à partir de populations B *in vivo*, ces données sont en contradiction avec l'étude de Kaplan et al, qui utilisait des gènes rapporteurs et un segment VHS107. Il est en effet probable que, dans la population totale des splénocytes de notre modèle, la diversité des réarrangements composant le répertoire confère une hétérogénéité de sensibilité à ARID3A. Toutefois et de façon inattendue, l'analyse des transcrits primaires au locus Igk, quant à elle, montre une diminution significative de l'expression de la chaîne légère en absence de ARID3A (**Figure 2C**). Il est envisageable qu'une telle altération de la transcription au locus Igk, majoritairement réarrangé dans le répertoire naïf de la souris, conduise à une diminution efficace de l'expression de la chaîne légère  $\kappa$ . Une telle diminution pourrait potentiellement expliquer la diminution de l'expression du BCR à la surface des cellules B au repos dépourvues de ARID3A. Le locus Igk possède également une région MAR associée à l'enhancer Eik, placée directement en amont de ce dernier. Jusqu'ici les études portant sur le rôle d'ARID3A au sein des lymphocytes B ne font pas état de la régulation de l'expression de la chaîne légère. Nos données soulèvent désormais la question d'une liaison possible d'ARID3A au locus Igk. Une telle liaison de ARID3A sur la région MAREik ou sur certains promoteurs V $\kappa$  pourrait apporter un effet activateur « direct » à la transcription de la chaîne légère. La suppression de la région MAREik chez la souris ne fait état d'aucune diminution de la transcription et de l'expression de la chaîne légère, il est toutefois possible que les méthodes d'analyse utilisées à l'époque (Northern blot) n'aient pas permis de saisir la finesse de cette régulation (Yi et al., 1999). L'hypothèse d'un effet direct de ARID3A sur l'expression du locus Igk s'avère intéressante. Des analyses de liaison de ARID3A au génome des cellules B par CHIP-seq sont programmées.

### **Impact de la suppression de ARID3A sur la réponse antigénique**

Notre modèle de délétion conditionnelle de ARID3A se distingue du modèle publié par Habir et al, par son homéostasie B conservée, sans modification substantielle des populations B2 périphériques. Pour autant, l'analyse par ELISA de la production d'immunoglobulines dans le sérum des animaux ARID3A cKO montre une diminution de la sécrétion des isotypes majoritaires IgM et IgG1 (**Figure 3**) ; une conséquence également observée dans le modèle précédemment publié. Dans le but d'évaluer leurs capacités à répondre à un stimulus antigénique, deux cohortes de souris (sauvage et ARID3A cKO) ont été immunisées avec un cocktail composé de NP-CGG (Nitrophénylacétyl-Chicken Gamma Globuline)



**Figure 4** : **A)** IgHV Family genes rearranged in GL7+ sorted B cell from Peyer's Patches in WT and Homozygote. **B)** NPCGG specific immunoglobulin (IgM or IgG) evaluated in sera from WT and ARID3A cKO mice. **C)** PCKLH specific immunoglobulin (IgM or IgG) evaluated in sera from WT and ARID3A cKO mice. *p*-value was determined with Mann Whitney test, significant differences were indicated by : \*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$  and error bars represent SEM.

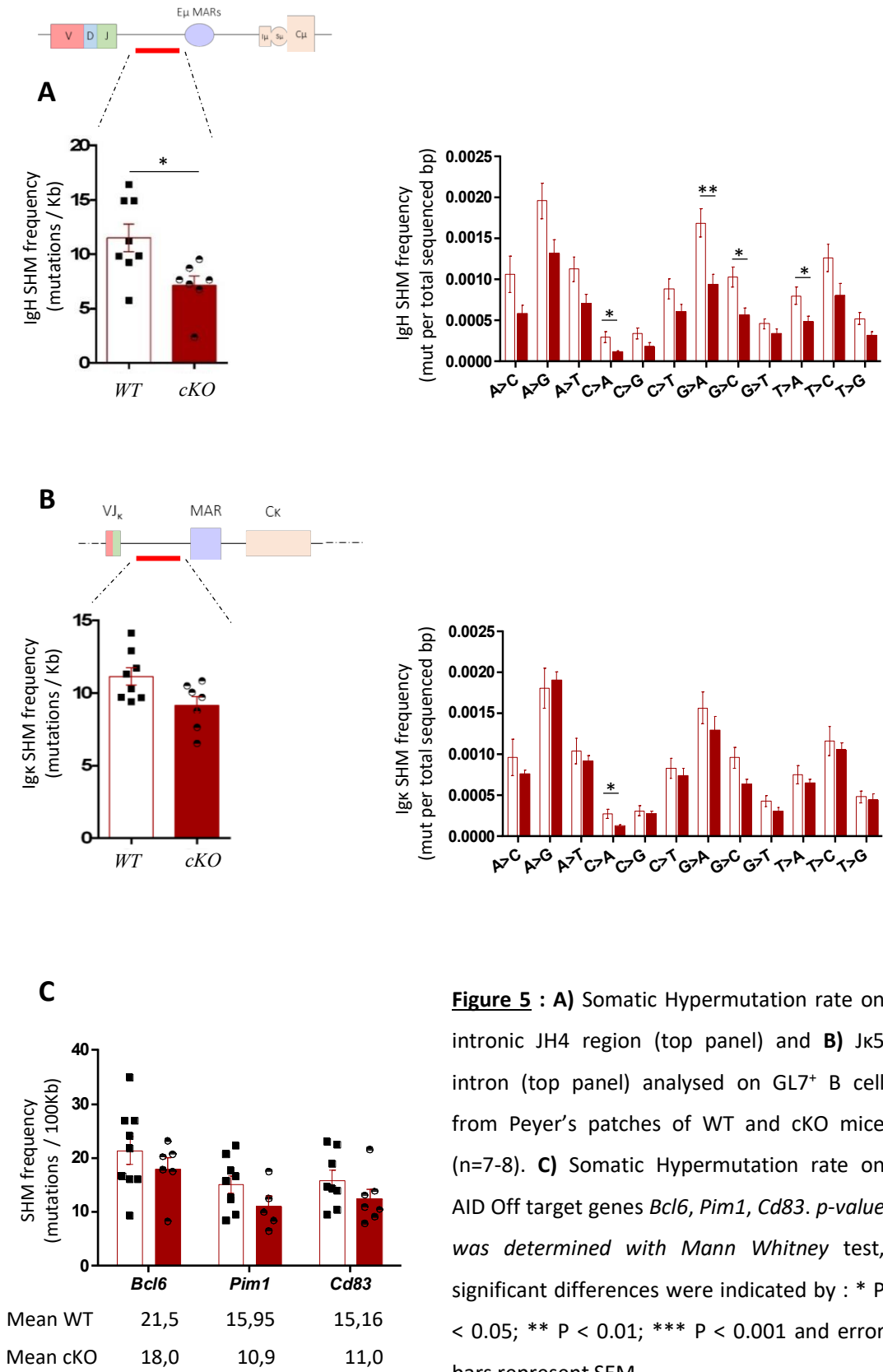
et de PC-KLH (Phosphorylecholesterol-Keyhole Limpet Hemocyanin). Une injection de rappel a été réalisée 8 jours après la première puis le sacrifice est effectué après 17 jours d'immunisation.

Nous avons dans un premier temps comparé les réponses globales des modèles WT et cKO en procédant à une analyse du répertoire des chaînes lourdes d'Ig exprimées par les populations de cellules B activées des centres germinatifs après double immunisation (NP-CGG / PC-KLH). Pour l'ensemble des familles des segments IgHV, la proportion similaire des transcrits d'Ig exprimés par les cellules B activées des deux modèles indique que l'absence de ARID3A n'induit pas de biais dans la sélection du répertoire après immunisation (**Figure 4A**). Ces données indiquent que, bien qu'exprimant plus faiblement leur BCR, nos animaux déficients pour ARID3A sont cependant dotés d'une capacité de réponse comparable à celle des animaux WT.

En complément, nous avons évalué les réponses sériques spécifiques des antigènes NP et PC. En accord avec la concentration d'IgM sériques totale globalement diminuée dans ce modèle, les animaux dépourvus de ARID3A affichent un taux réduit d'IgM dirigés contre ces deux antigènes. Bien que les cellules B dépourvues de ARID3A soient capables de répondre à la présence des deux antigènes, nous constatons que les concentrations en anticorps IgM et IgG spécifiques des haptènes NP et PC se trouvent systématiquement diminués en l'absence de ARID3A (**Figure 4B et 4C**). Un constat similaire a été observé dans le modèle publié par l'équipe de Malin après immunisation par l'antigène PC-KLH (Habir et al., 2017) ; d'autres études basées sur un modèle transgénique appuient également la nécessité de ARID3A pour la production d'Ig dans le sérum (Shankar et al., 2007). Nos données suggèrent que la suppression de ARID3A dans le lignage B n'affecte pas la capacité du pool de lymphocytes B à répondre à un stimulus antigénique mais tend à limiter l'aptitude des cellules B à soutenir une réponse humorale conséquente.

### **La suppression de ARID3A compromet l'hypermutation somatique au locus IgH**

Nos travaux précédents, portant sur l'étude de l'enhancer E $\mu$  et de ses régions flanquantes MARsE $\mu$  (décrites pour fixer ARID3A), ont identifié une fonction spécifique aux régions MAR en tant que cis-activateurs de la SHM. La découverte de cette nouvelle fonction associée aux régions MARsE $\mu$  pose la question d'une éventuelle fonction de ARID3A au cours de ce processus. L'analyse de la fréquence d'hypermutation a donc été évaluée dans notre modèle déficient pour cette MAR BP, dans un premier temps dans les cellules activées des centres germinatifs triées à partir des plaques de Peyer. La fréquence des mutations est analysée dans la région intronique en aval du segment J<sub>H4</sub>, par séquençage haut débit associé à un filtrage réalisé à l'aide de l'outil DeMiNEr (Martin et al., 2018). Les échantillons issus d'animaux cKO présentent une moyenne de SHM de 7.1 mutations/Kb contre 11.5 mut/Kb chez



**Figure 5 : A)** Somatic Hypermutation rate on intronic JH4 region (top panel) and **B)** Jk5 intron (top panel) analysed on GL7<sup>+</sup> B cell from Peyer's patches of WT and cKO mice (n=7-8). **C)** Somatic Hypermutation rate on AID Off target genes *Bcl6*, *Pim1*, *Cd83*. *p*-value was determined with Mann Whitney test, significant differences were indicated by : \* P < 0.05; \*\* P < 0.01; \*\*\* P < 0.001 and error bars represent SEM.

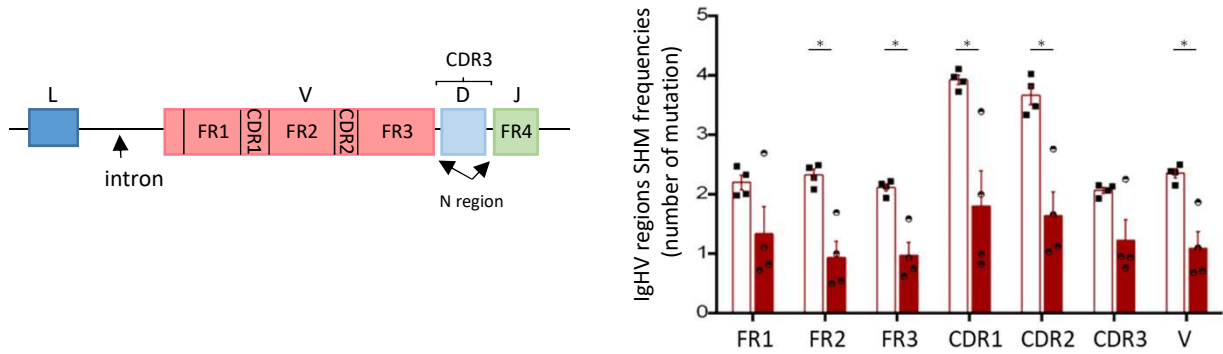
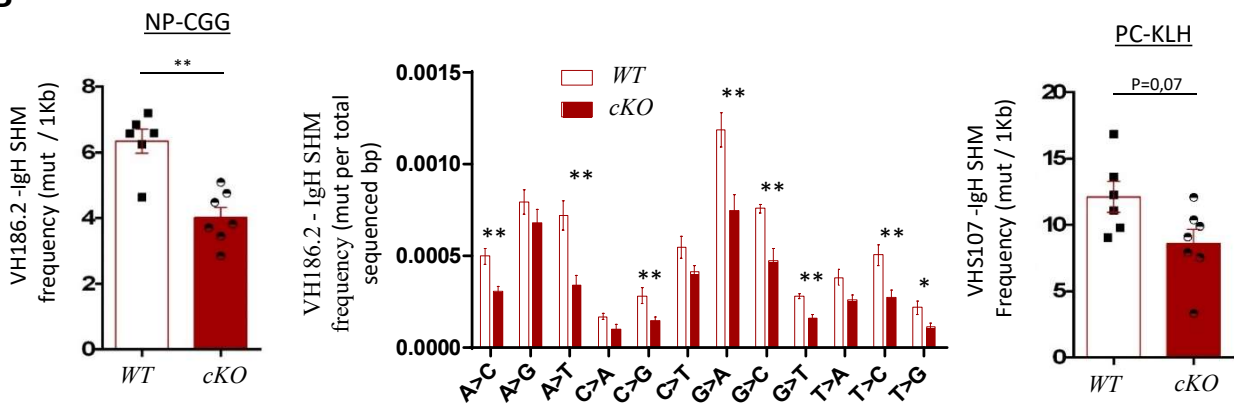
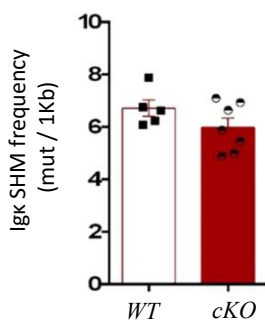
les animaux sauvages (**Figure 5A**). La diminution significative de la SHM observée dans le modèle mutant suggère une implication de ARID3A dans ce processus au locus IgH. Bien que statistiquement non significative, la fréquence de mutation au locus Igk présente également une tendance à la diminution, chez les animaux cKO. Cette observation indique également une potentielle altération du mécanisme de SHM à ce locus (**Figure 5B**). Nos résultats plaident en faveur d'une fonction de ARID3A comme un facteur optimisant l'affinité des Ig en favorisant le mécanisme d'hypermutation aux gènes cibles légitimes d'AID (les loci des chaînes lourdes et légères).

L'analyse de la fréquence des mutations aux gènes cibles illégitimes d'AID tels que *Bcl6*, *Pim1* et *Cd83* est également un indicateur d'une altération globale du mécanisme de SHM. Dans le cas de précédentes collaborations, nous avons montré que notre outil DeMinEr était adapté à la détection de mutations à bas bruit dans ces oncogènes (Martin et al., 2018). Cet outil a permis de prouver que la suppression de contributeurs majeurs de la SHM tels que UNG, un composant du RNA exosome (Lim et al., 2017) et plus récemment un interacteur de UNG nommé FAM72A (**Annexe 2**) (Rogier et al., 2021) influençait de façon significative la SHM dans ces oncogènes. Dans notre modèle, la suppression de ARID3A provoque une légère altération (non significative) de la fréquence d'hypermutation dans les régions ciblées par AID des oncogènes *Pim1* et *Cd83* ; en revanche les mutations dans l'oncogène *Bcl6* ne semblent pas impacter par l'absence de la MAR BP (**Figure 5C**). Les mutations aux gènes cibles illégitimes d'AID étant des dommages collatéraux de l'action de l'enzyme, il semblerait que l'altération du mécanisme (engendrée par la délétion d'ARID3A) au locus IgH impacterait spécifiquement, bien que très faiblement, l'hypermutation de certains gènes cibles illégitimes, ici potentiellement *Pim1* et *Cd83*. Le fait que les mutations « off target » à l'oncogène *Bcl6* surviennent à une fréquence quasi normale en l'absence de ARID3A indique 1) que le processus de déamination par l'enzyme AID demeure actif en l'absence de la MAR BP et 2) que la dynamique de prolifération et de sélection cellulaire dans le centre germinatif ne semble pas être altérée en l'absence de ARID3A.

Bien que la quantification de la SHM dans les cellules du CG des plaques de Peyer permette une évaluation globale du mécanisme, elle demeure dépendante de l'activation continue par les microbiotes des animaux modèles, élevés en zone exempt d'organisme pathogène spécifique (EOPS). Dans ce contexte sanitaire contrôlé, nous avons évalué la SHM de façon plus fine en recherchant les mutations induites en réponse aux deux antigènes NP et PC utilisés lors du protocole de vaccination décrit précédemment.

Dans un premier temps, nous avons mis à profit l'étude du répertoire ARN exprimé par les cellules B de la rate pour en extraire une information d'hypermutation somatique dans les exons VDJ réarrangés. Bien que la diversité du répertoire IgH soit identique dans les deux génotypes, l'hypermutation somatique est significativement réduite dans l'ensemble des régions FR ou CDR (à l'exception du FR1 et du CDR3) du modèle ARID3A KO (**Figure 6A**). Dans un second temps, l'étude de l'hypermutation



**A****B****C**

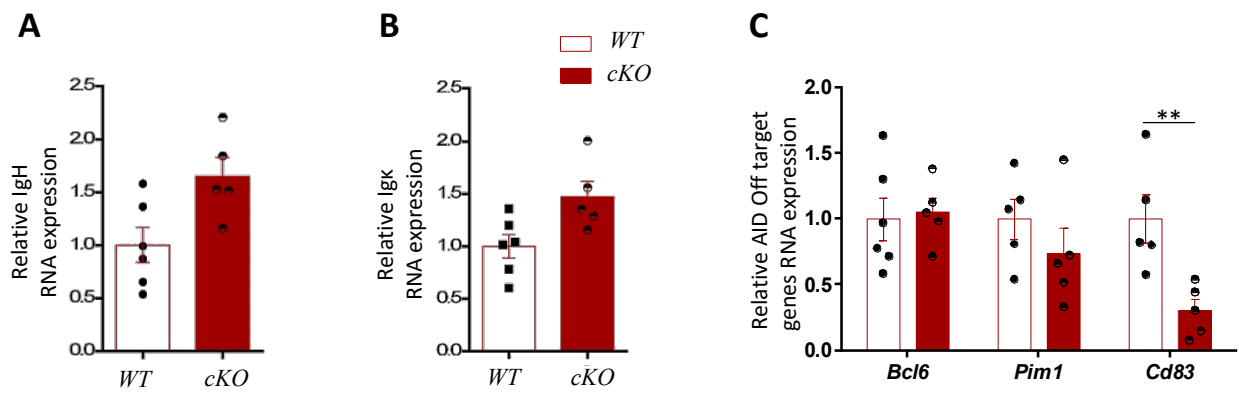
**Figure 6 : A)** Somatic Hypermutation level on FR and CDR fragment analysed on GL7<sup>+</sup> B cell from Spleen of WT and cKO mice (n=4). **B)** Somatic Hypermutation rate on intronic VH186,2-JH4 region induced by NP-CGG immunization (left panel) and VHS107-JH4 intron (right panel) analysed on GL7<sup>+</sup> B cell from Spleen of WT and cKO mice (n=6-7). **C)** Somatic Hypermutation rate on Jk5 intronic region analyzed on GL7<sup>+</sup> B cell from Spleen of WT and cKO mice (n=6-7). *p*-value was determined with Mann Whitney test, significant differences were indicated by : \* *P* < 0.05; \*\* *P* < 0.01; \*\*\* *P* < 0.001 and error bars represent SEM.

somatique a été réalisée cette fois sur l'ADN des cellules activées de centres germinatifs néo synthétisés, triées à partir de la rate des animaux immunisés. Afin d'analyser la fréquence de mutation en réponse aux deux antigènes utilisés, nous avons sélectionné les réarrangements majoritairement impliqués dans ces réponses respectives. Le séquençage est donc réalisé à partir d'amplicons de haute-fidélité générés par PCR en utilisant des amorces « sens » spécifiques des segments  $V_H186.2$  (majoritairement utilisé lors de la réponse NP-CGG) (Hirth et al., 2019) et  $V_H107$  (majoritairement utilisé lors de la réponse PC-KLH). L'expression du segment  $V_H107$  est par ailleurs décrite comme régulée par ARID3A dans des expériences utilisant des gènes rapporteurs (Kaplan et al., 2001). De façon corrélée aux analyses effectuées dans les plaques de Peyer, ces expériences complémentaires montrent que la fréquence de mutations des segments réarrangés  $V_H186.2$  est significativement diminuée chez les souris ARID3A cKO après immunisation avec l'antigène NP-CGG (Moyenne WT : 6.34 mutations /Kb ; Moyenne ARID3A cKO : 4.02 mut/Kb ; p-value : 0.005). Dans le cas de l'immunisation avec l'antigène PC-KLH, la réponse SHM est également diminuée mais de façon non significative en l'absence de ARID3A (Moyenne WT : 12.11 mutation /1Kb ; Moyenne ARID3A cKO : 8.60 mut/Kb ; p-value : 0.073) (**Figure 6B**).

Nous avons également évalué la SHM de la chaîne légère majoritaire dans les cellules B de CG triées après le protocole d'immunisation. Bien que résultant d'un stimulus antigénique, l'évaluation de la SHM, dans la région en aval du segment réarrangé  $J\kappa5$ , ne cible pas les segments VJ préférentiellement impliqués dans la réponse aux antigènes NP et PC mais reflète une réponse globale. Dans ce cas, tout comme dans les cellules des CG des plaques de Peyer, la réduction (bien que modeste et non significative) de la fréquence d'hypermutation observée au locus  $Ig\kappa$  est en concordance avec la diminution significative observée sur les chaînes lourdes de ces mêmes cellules (**Figure 6C**).

La transcription est décrite comme un prérequis indispensable à l'action de AID en tant qu'initiateur de l'hypermutation somatique (Fukita et al., 1998). Nous avons recherché si la délétion de ARID3A dans le lignage B pouvait affecter la transcription préalable aux événements de SHM. Pour ce faire, nous avons évalué par RT-qPCR, selon un protocole précédemment décrit au laboratoire (Rouaud et al., 2013), la transcription primaire aux loci des gènes d'Ig ( $IgH$  et  $Ig\kappa$ ) dans la région située en aval du dernier segment J dans des cellules activées des CG de la rate des animaux immunisés. Bien que la suppression de ARID3A diminue la SHM de ces gènes en réponse aux immunisations par les antigènes NP-CGG et PC-KLH, nous n'observons pas d'altération significative de la transcription primaire aux loci  $IgH$  et  $Ig\kappa$  chez les souris ARID3A cKO. A l'inverse, une tendance à l'augmentation de la transcription de ces régions est même constatée dans les cellules ARID3A cKO (**Figure 7A et 7B**).

Ces résultats indiquent clairement que l'altération du mécanisme de SHM, observée en l'absence de ARID3A, s'opère par un mécanisme indépendant de la transcription de la région ciblée. Dans ces loci des gènes d'Ig, le découplage des processus de SHM et de transcription, observé dans les cellules



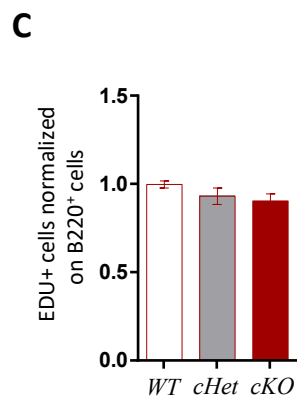
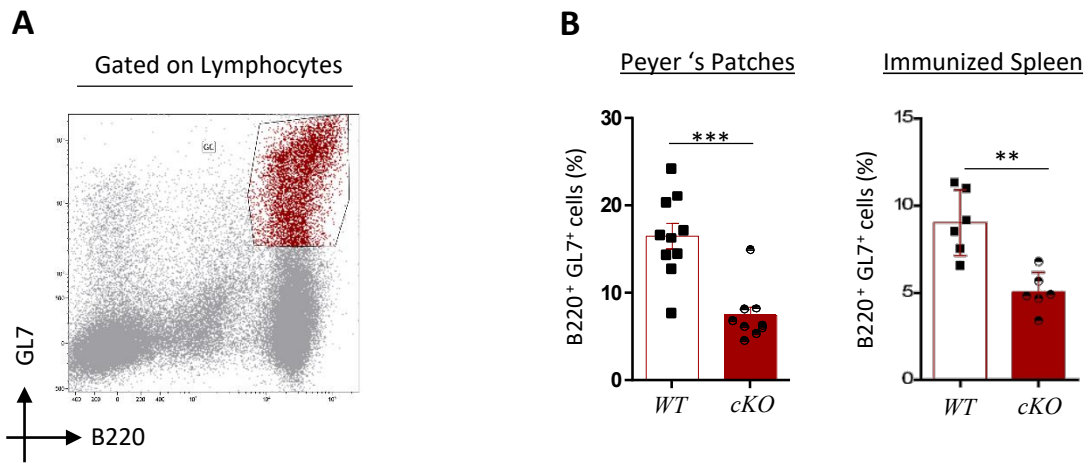
**Figure 7 :** **A)** IgH transcript levels in WT and conditional Homozygote GL7<sup>+</sup> B cell from immunized spleen (n=5-6). **B)** Igk transcript levels in WT and conditional Homozygote GL7<sup>+</sup> B cell from immunized spleen (n=5-6). **C)** AID Off target genes transcript levels in WT and conditional Homozygote GL7<sup>+</sup> B cell from immunized spleen (n=5-6). *p-value was determined with Mann Whitney test, significant differences were indicated by : \* P < 0.05; \*\* P < 0.01; \*\*\* P < 0.001 and error bars represent SEM.*

déficientes pour cette MAR BP, suggère que ce facteur exerce son effet régulateur indépendamment des régions régulatrices 3' des loci *IgH* et *Igk* (Garot et al., 2016; Inlay et al., 2006; Rouaud et al., 2013). A l'inverse, ces résultats sont à rapprocher de l'effet de la délétion des régions MARsE $\mu$  ou MAREik dans des modèles murins chez lesquels une diminution de SHM au loci concerné n'est pas corrélée à une diminution de la transcription (Article annexe n°1 Martin et al., 2022; Yi et al., 1999). Notre étude approfondie du modèle MARsE $\mu$ , menée dans des fonds génétiques déficients pour les voies de réparation BER et MMR (Article annexe n°1 (Martin et al., 2022), précise en effet que ces régions limitent l'action de la réparation infidèle associée à la voie BER, facilitant ainsi la SHM en amont de la région E $\mu$ . Nous pouvons donc émettre l'hypothèse selon laquelle ARID3A participerait à cet effet facilitateur de la réparation infidèle des lésions induites par AID lors de la SHM. La vérification de cette hypothèse constitue un projet ambitieux qui nécessite des croisements dans des fonds génétiques déficients pour les voies de réparation BER et MMR.

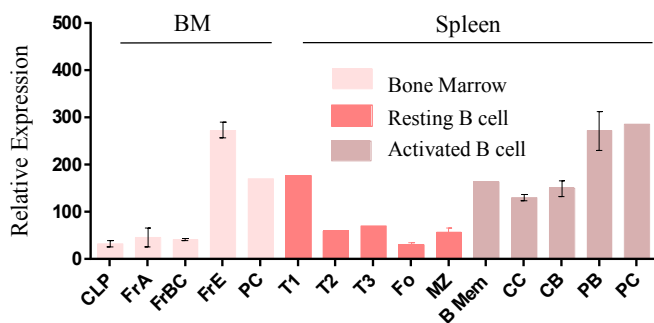
Un rôle potentiellement ambivalent de ARID3A dans le lignage B est également suggéré par l'effet opposé sur la transcription du locus *Igk* aux stades « B au repos » et « B activé ». Nos données semblent montrer que la suppression de la MAR BP influence différemment la transcription de *Igk* selon le stade cellulaire : ARID3A serait un activateur transcriptionnel dans les cellules au repos ; cette fonction serait abolie dans les cellules activées du CG (dans lesquelles il pourrait potentiellement exercer un rôle répresseur).

Nous avons également évalué, dans les cellules B du CG issues du protocole d'immunisation, la transcription des gènes cibles illégitimes de AID. Si l'absence de ARID3A induit une légère diminution de la fréquence de SHM dans certains de ces gènes (*Pim1* et *Cd83* ; dans les cellules B de plaques de Peyer) ; la suppression de ARID3A conduit à des effets variés sur leur transcription respective. L'expression des oncogènes *Bcl6* et *Pim1*, transcrits de façon similaire chez les souris sauvages et les mutantes, ne semble pas contrôlée par ARID3A après activation (**Figure 7B**). A l'inverse, la transcription de *Cd83* diminue significativement dans les cellules B déficientes pour ARID3A (**Figure 7B**). Cependant il demeure délicat de corréliser cette analyse de transcrits réalisée sur des cellules d'animaux immunisés et les données de SHM obtenues à partir de cellules du CG de plaque de Peyer ; il est nécessaire de générer des données complémentaires de SHM et transcription sur les mêmes échantillons.

Nos données détaillées précédemment permettent d'attribuer à ARID3A un rôle de facteur régulateur du mécanisme de SHM. Cependant des analyses complémentaires sont indispensables à la confirmation de ce rôle et la compréhension du mécanisme sous-jacent.



**Figure 8 :** **A)** Gating strategy by flow cytometry to sort GL7<sup>+</sup> B cells from Germinal Center of WT and cKO mice. **B)** Percentage of GL7<sup>+</sup> B cells from Peyer's Patches of WT and cKO mice. **C)** EdU<sup>+</sup> incorporating Splenic B cells during in vitro LPS stimulation normalized on WT values (n=6-11). *p*-value was determined with Mann Whitney test, significant differences were indicated by : \* P < 0.05; \*\* P < 0.01; \*\*\* P < 0.001 and error bars represent SEM.



**Figure 9 :** Representation of ARID3A relative expression in B cell populations (Common Lymphoid progenitor, Fraction A, B, C, E and Mature spleen B cell - Data from IMMGEN)

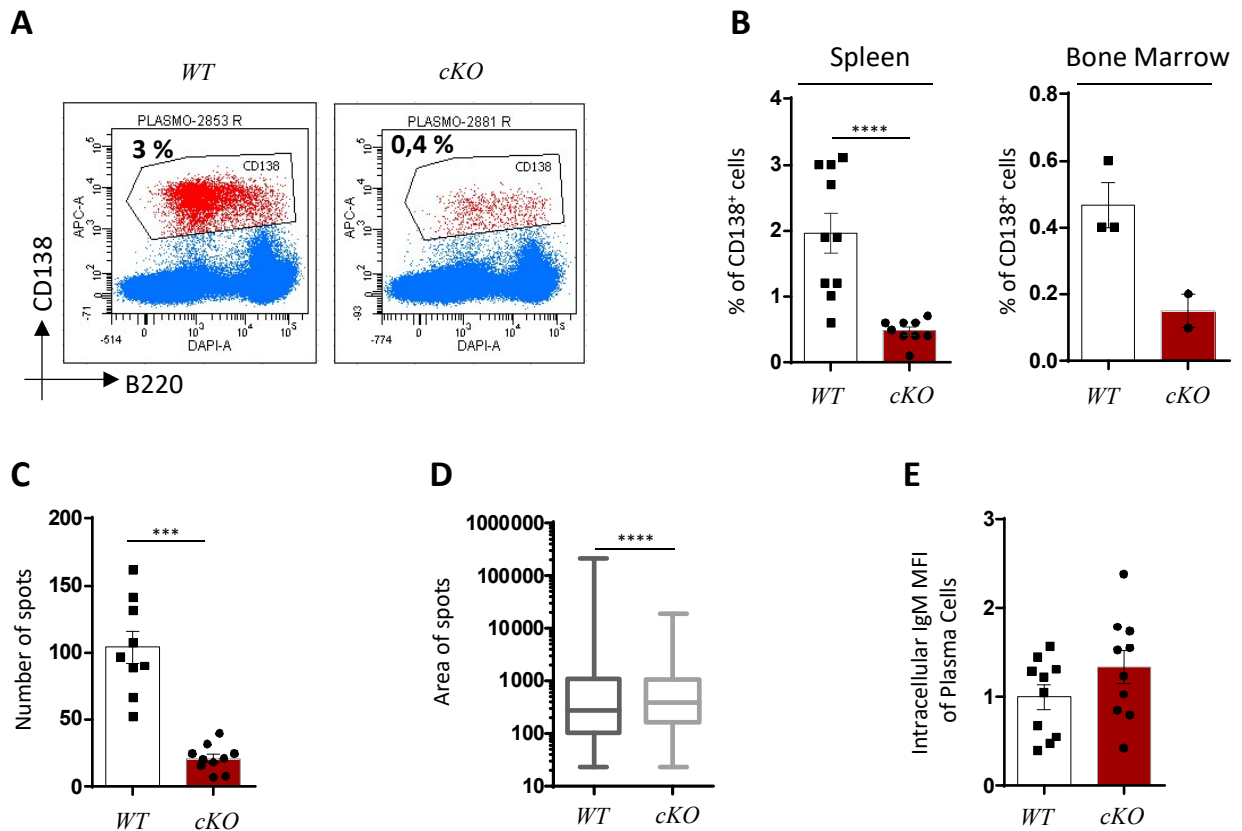
### **La délétion de ARID3A ne compromet pas la prolifération des lymphocytes B**

Bien que notre modèle affiche une représentation normale des population B médullaires et spléniques, il montre tout de même une diminution de la proportion de cellules activées dans le centre germinatif, fait observé en contexte physiologique ou après immunisation (**Figure 8A et 8B**). Certaines données de la littérature pouvaient suggérer une fonction de ARID3A lors de la prolifération cellulaire. Ce facteur fixe préférentiellement les régions MARSEμ lorsque les cellules s'engagent en phase S (Kim et al., 2017). ARID3A coopère également avec le facteur P53 qui contribue à la régulation de gènes impliqués dans le cycle cellulaire (Lestari et al., 2012). Les centroblastes, cibles du mécanisme d'hypermutation, sont des cellules hautement proliférantes. Il est donc envisageable que l'altération du mécanisme de SHM observé en l'absence de ARID3A soit un effet secondaire d'une altération des capacités de prolifération des centroblastes. Nous avons comparé les capacités de prolifération des lymphocytes B après 48H d'activation *in vitro* par le LPS en présence d'EdU (Ethylnyl deoxyuridine). Nos mesures, réalisées en cytométrie de flux, montrent une incorporation identique de l'intercalant par les cellules WT et déficientes pour ARID3A (**Figure 8C**). Ces données semblent indiquer que l'absence de ARID3A ne conduit pas à un défaut de prolifération majeur, dans les conditions d'activation par le LPS. Ces données obtenues *in vitro* nécessitent d'être consolidées par des études de cycle cellulaire réalisées *in vivo* après injection de BrDU (BromoDésoxyUridine).

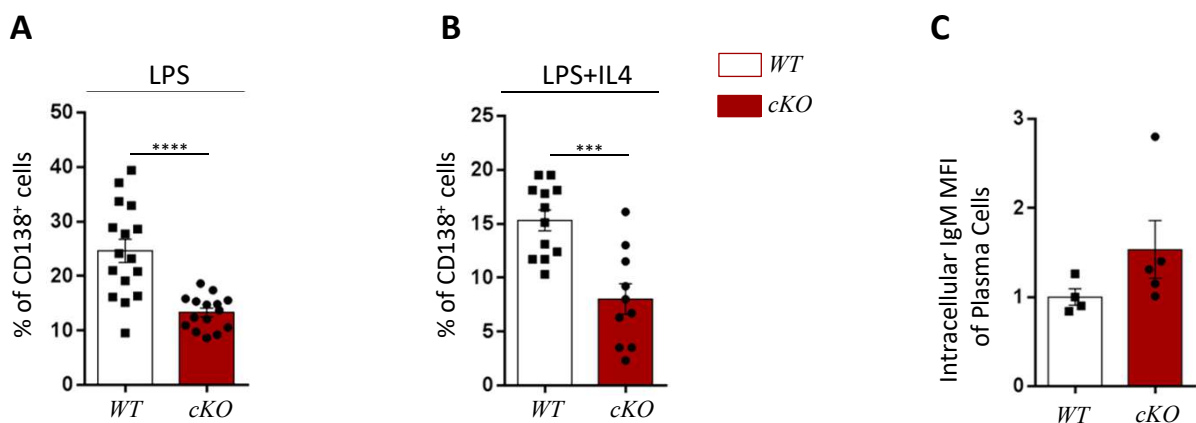
### **La suppression de ARID3A limite la différenciation des plasmocytes**

D'après nos données précédentes, la suppression de ARID3A ne semble pas limiter la transcription des chaînes lourdes d'Ig dans les cellules au repos et activées. Pour autant, notre modèle présente un défaut de production sérique des anticorps de classe IgM et IgG1 (**Figure 3**). Dans l'éventualité d'un effet de la délétion de ARID3A sur la sécrétion de ces isotypes majoritaires, nous avons exploré le compartiment plasmocytaire dans notre modèle animal.

Des données d'expressions tissulaires en RNA-seq issues de la DataBase IMMGEN indiquent que le profil d'expression de ARID3A fluctue au cours du développement B chez a souris : faiblement exprimé dans les précurseurs, son expression est augmentée aux stades transitionnels, diminue progressivement dans les cellules au repos (folliculaire et cellules de la zone marginale) puis augmente à nouveau dans les cellules activées du centre germinatif et dans les stades finaux du développement : plasmablastes et plasmocytes (**Figure 9**) (The Immunological Genome Project Consortium et al., 2008). Ce profil d'expression suggère donc des fonctions plus importantes de la protéine aux stades activés du développement, soutenant ainsi son rôle dans l'hypermutation au sein des B du CG, et posant la question d'une fonction spécifique au stade plasmocytaire.



**Figure 10** : **A**) Gating strategy to plasma cells after SRBC immunization during 7 days of 2 month old WT and Homozygote mice. **B**) Percentage of CD138<sup>+</sup> B cells after SRBC immunization in spleen (left panel) and Bone marrow (right panel). **C**) Object numbers acquired by ELISpot on splenic B cells after SRBC immunization. **D**) Spot's Area of ELISpot on splenic B cells immunized 7 days with SRBC. **E**) Normalization of IgM MFI in SRBC immunized-Splenic Plasma cells from WT and ARID3A cKO mice. *p*-value was determined with Mann Whitney test, significant differences were indicated by : \* *P* < 0.05; \*\* *P* < 0.01; \*\*\* *P* < 0.001 ; \*\*\*\* *P* < 0,0001 and error bars represent SEM.



**Figure 11** : **A**) Percentage of CD138<sup>+</sup> cells generated in vitro after LPS (n=15-16) or **B**) LPS + IL4 (n=10-12) stimulation of splenic B cells from WT and ARID3A cKO mice. **C**) Normalization of IgM MFI in LPS-stimulated Plasma B cell from WT and ARID3A cKO mice (n=4-5). *p*-value was determined with Mann Whitney test, significant differences were indicated by : \* *P* < 0.05; \*\* *P* < 0.01; \*\*\* *P* < 0.001; \*\*\*\* *P* < 0,0001 and error bars represent SEM.

Nous avons analysé la différenciation plasmocytaire des souris *in vivo* après un protocole d'immunisation de 7 jours au SRBC (Sheep Red Blood Cells). L'analyse par cytométrie en flux des cellules B issues de la rate des animaux immunisés révèle un défaut prononcé dans la différenciation plasmocytaire induit par la délétion d'ARID3A. Chez les animaux cKO, seulement 0.5%, en moyenne, des cellules portent le marqueur du stade plasmocytaire CD138 contre 1.96 % chez les souris sauvages (**Figure 10A et 10B**). Ce déficit cellulaire est également observé dans la moelle osseuse des animaux ARID3A cKO qui ont 5 fois moins de cellules plasmocytaires que les souris sauvages (**Figure 10B**). Ces données sont aussi confirmées après la réalisation d'un ELISPOT IgM, sur des cellules de rate issues des animaux immunisés au SRBC. Après incubation des cellules B triées de la rate, provenant des animaux cKO, on observe 4 fois moins de spot (correspondant aux cellules sécrétrices d'Ig) qu'en condition identique pour les souris sauvages (**Figure 10C**). En revanche, en absence de ARID3A, les cellules sécrétrices présentent des spots dont l'aire apparait significativement plus importante suggérant une sécrétion d'immunoglobuline accrue dans ce modèle (**Figure 10D**). Dans ces mêmes échantillons de rate issus d'animaux immunisés, nous avons évalué, par cytométrie en flux, la production d'IgM intracellulaire dans les cellules sécrétrices d'anticorps. Bien que non statistiquement significative, la moyenne de fluorescence (MFI) évaluant la concentration intracellulaire en IgM est légèrement plus importante dans les plasmocytes des animaux cKO (**Figure 10E**). En somme, nos données obtenues *in vivo* indiquent que la délétion de ARID3A altère ou ralentie la différenciation finale des cellules B en plasmocytes mais n'induit pas une diminution de la production d'Ig dans ces cellules.

En complément des expériences de stimulation *in vivo* par le SRBC, induisant une activation polyclonale et la formation de centres germinatifs dans la rate, nous avons testé les capacités des cellules B de nos modèles à initier la différenciation plasmocytaire *in vitro*. Nous avons donc évalué, par cytométrie en flux, la proportion de cellules B exprimant le marqueur CD138 après 4 jours d'activation *in vitro* par le LPS avec ou sans IL4 (interleukine 4) chez les souris sauvages ou déficiente pour la MAR BP. En absence de ARID3A seulement 15% des cellules stimulées par le LPS sont capables de se différencier en plasmablaste (B220<sup>+</sup> CD138<sup>+</sup>) contre 25% pour le modèle sauvage (**Figure 11A**). Un effet identique de réduction significative de la différenciation *in vitro* est également observé après stimulation par le LPS et l'IL4 : dans ces conditions, nous n'observons que 7 % de plasmablastes en l'absence de ARID3A contre 15% en condition sauvage (**Figure 11B**). Cette analyse *in vitro* suggère que la délétion de ARID3A conduit à un déficit intrinsèque des cellules B dans l'engagement de la différenciation plasmocytaire et confirme l'importance de la MAR BP à ce stade. Dans ce modèle d'activation *in vitro*, l'évaluation du contenu intracellulaire en Ig par cytométrie en flux confirme les observations acquises *in vivo* : les plasmablastes dépourvus de ARID3A obtenus après 4 jours de culture présente une augmentation (non significative) de leur contenu en IgM (**Figure 11C**).





L'hypothèse d'une altération de la différenciation plasmocytaire en l'absence de ARID3A est également soutenue par nos données de sérologie en réponse aux immunisations des souris (NP-CGG/ PC-KLH). Dans notre étude, la diminution observée de la production d'anticorps sériques spécifiques des haptènes NP et PC peut être logiquement expliquée par une altération du nombre de plasmocyte chez les animaux KO (**Figure 4B et 4C**). Des résultats identiques, rapportant un défaut de réponse humorale spécifique, ont été décrits dans des études précédentes (Habir et al., 2017; Nixon et al., 2008) sans que les auteurs aient pu y apporter une explication convaincante.

Il s'avère désormais indispensable de confirmer l'implication de ARID3A dans la différenciation plasmocytaire. A ce jour nous envisageons de cibler les transcrits du gène ARID3A humain à l'aide d'oligonucléotides antisens (ASO) et d'étudier l'effet de leur suppression dans un modèle de différenciation plasmocytaire *in vitro* développé à partir de lymphocytes B issus de donneurs sains.

Si une telle implication se confirmait, les analyses différentielles des transcriptomes de différentes sous-populations de cellules de souris WT et dépourvues de ARID3A (cellules B de la rate au repos, cellules B activées 2 jours au LPS et populations B enrichies en plasmablastes après 4 jours d'activation au LPS) pourraient apporter des éléments de réponse concernant les gènes cibles de ARID3A potentiellement impliqués dans la différenciation plasmocytaire.

Dans les plasmocytes issus de nos modèles, il est également indispensable d'étudier l'effet de la délétion de ARID3A sur l'expression de la chaîne légère (Igk dans un premier temps) : sur le plan transcriptionnel (par RT-qPCR) et sur l'expression de la chaîne légère ( $\kappa$ ) intracellulaire dans les plasmocytes (par cytométrie de flux et en western-blot).

Les données apportées par nos modèles murins suggèrent également un intérêt pour l'étude *in vitro* des capacités des cellules B dépourvues de ARID3A à effectuer la CSR : il est en effet envisageable qu'un blocage partiel de la différenciation plasmocytaire conduise à une accumulation de cellules ayant effectuées la CSR. Nous envisageons de mesurer dans nos modèles, de façon cinétique, la production de transcrits germinaux préalables à la CSR et d'évaluer la proportion de cellules ayant effectuée la commutation par cytométrie en flux.

Il demeure intéressant de vérifier, par des moyens de génomique (expériences de ChIP-seq ou de Cut and Run) l'hypothèse qui présente ARID3A comme un facteur répresseur de la transcription des gènes d'Ig dans les cellules B activées.





## ***Discussion***



## Discussion

---

Le modèle murin MARsE $\mu$  KO a révélé un rôle insoupçonné des régions MAR dans les stades tardifs du développement B et notamment lors du processus d'hypermutation somatique. Ce modèle de délétion des deux régions d'attachement à la matrice nucléaire, encadrant l'enhancer intronique E $\mu$  du locus IgH, présente une altération du mécanisme de SHM ciblé à ce locus. La fréquence de mutation se trouve diminuée en amont des régions MARsE $\mu$  et augmentée en aval de ces régions. Ces données dessinent l'hypothèse que les régions MARsE $\mu$  feraient office de barrière via une liaison dynamique à la matrice nucléaire et permettraient ainsi de délimiter et restreindre au mieux la réparation infidèle (error prone) dans la zone la plus appropriée, située en amont de l'enhancer. Les MAR au locus IgH sont définies comme des régions d'attachement facultatives, ceci signifie que leur liaison à la matrice serait dynamique et différente selon le type cellulaire étudié. Faisant office de plateforme de recrutement protéique, leur liaison aux protéines de la matrice pourrait être médiée par plusieurs acteurs dont les MAR binding proteins les plus connues : SATB1 et ARID3A. Ces deux protéines présentent des profils d'expression particuliers au cours du développement B suggérant une régulation fine de leur expression en lien avec leurs fonctions aux différents stades cellulaires.

Mes travaux de thèse ont contribué à déterminer les fonctions de ces protéines, en relation avec leur attachement aux régions MARsE $\mu$  au stade tardif du développement des cellules B.

La protéine SATB1 est essentielle aux cellules T ; elle contribue largement à leur développement et à leurs fonctions. Dans les lymphocytes B, en revanche, les données impliquant SATB1 restent très peu nombreuses. Grâce à l'utilisation d'un modèle murin de délétion conditionnel, nous avons pu d'abord rechercher si cette protéine jouait un rôle majeur dans la maturation des cellules B. D'après notre modèle, la délétion de *Satb1* ne contribue pas au développement des lymphocytes B. De façon inattendue, étant donné sa contribution aux réarrangements du récepteur des cellules T et de son importance pour l'expression des gènes *Rag*, cette protéine ne participe pas à l'établissement du répertoire d'immunoglobulines. En revanche, en accord avec les données de transcription des gènes d'Ig, l'étude des populations cellulaires B montre que SATB1 favorise l'expression du BCR et joue le rôle d'un facteur activateur de la transcription de la chaîne lourde IgH au sein des cellules B matures.

Si la protéine n'est pas indispensable au développement des cellules B, sa délétion n'altère pas non plus leur activation : nous observons en effet une proportion équivalente de cellules du centre germinatif ou de plasmocytes dans les souris sauvages et dépourvues de l'expression de SATB1. Cependant, et de façon opposée à nos observations dans les cellules B matures, la délétion de SATB1

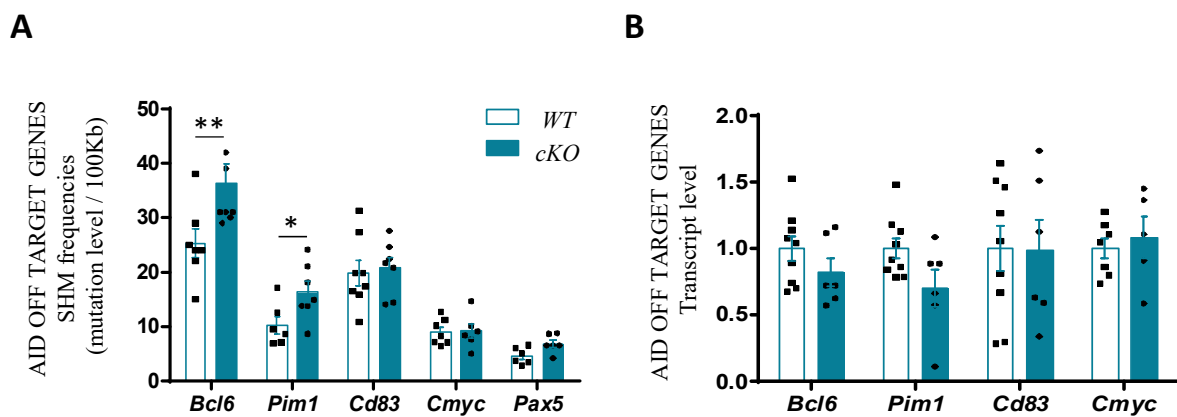


accroît la transcription et la production des Ig, attribuant un rôle répresseur à la protéine aux stades activés. Ces données mettent en évidence la dualité fonctionnelle de SATB1 dans le lignage B selon le stade de développement. Cette dualité pourrait avoir diverses origines. Elle pourrait notamment dépendre des partenaires de SATB1 dont la production et la localisation seraient liées au stade cellulaire étudié ou de la localisation nucléaire. Mais il est aussi probable que l'effet transcriptionnel médié par SATB1 dépende de la concentration de la protéine. Une étude récente démontre que l'expression du gène *Satb1* peut être régulée par la présence de 4 promoteurs distincts dont l'activation (notamment au sein des différentes sous populations lymphocytaires T) est dépendante du type cellulaire ; le niveau d'expression (fort ou faible) étant dépendant du promoteur activé (Patta et al., 2020). Il est alors envisageable que les promoteurs qui soient activés au sein des cellules matures et des cellules activées ne soient pas les mêmes, générant alors une quantité variable de protéine (Patta et al., 2020). En prenant en compte la capacité d'homodimérisation de cette MAR BP, la quantité de protéine générée pourrait avoir une incidence sur sa fonction et notamment sur la régulation de ces gènes cibles (Purbey et al., 2008). Enfin, les modifications post-traductionnelles de la protéine peuvent également influencer son activité (Pavan Kumar et al., 2006).

Nos observations suggèrent que SATB1 pourrait exercer des fonctions d'activateur de la transcription des gènes d'Ig dans les lymphocytes B matures non activés, un stade de développement au cours duquel la cellule B produit des immunoglobulines en quantité limitée, destinée à l'expression membranaire. En revanche, dans les cellules activées et plus particulièrement les plasmocytes qui ont pour fonction de sécréter de façon conséquente des immunoglobulines, nos données suggèrent que SATB1 exerce un rôle de répresseur de la production des Ig. Ces observations mises en parallèle avec l'expression régulée de SATB1 (fortement exprimée aux stades matures et diminuant aux stades activés) posent alors l'hypothèse d'une éventuelle fonction de cette protéine comme « gardien de l'expression des chaînes d'Ig ». L'expression réduite de SATB1 aux stades activés (du centre germinatif) et des plasmocytes permettrait de garder sous contrôle la production des Ig pour : 1) limiter une production accrue de façon prématurée qui serait non indispensable voir préjudiciable (si elle devait survenir dans une cellule trop peu différenciée) ou 2) freiner la production dans un contexte plus pathologique (emballement du système immunitaire – maladie auto-immune – développement tumoral).

Ce rôle « de gardien de l'expression des chaînes d'Ig » semble compatible avec les données d'hypermutation somatique de notre modèle. Dans ce dernier, l'absence de SATB1 induit une augmentation de la fréquence de mutations sur les régions cibles du mécanisme de SHM des gènes d'Ig. Cet effet pourrait être la conséquence directe de l'augmentation de la transcription de ces gènes observée dans les cellules B activées dépourvues de SATB1. En effet, plusieurs études ont montré que





**Figure 20 : Impact de la délétion de SATB1 sur les gènes AID OFF TARGET.**

**A)** Fréquences de mutations des gènes *Bcl6*, *Pim1*, *Cd83*, *Cmyc* et *Pax5* évaluées dans les cellules activées des centres germinatifs issus des plaques de Peyer des souris SATB1 cKO et WT. **B)** Taux de transcription des gènes AID OFF TARGET (*Bcl6*, *Pim1*, *Cd83* et *Cmyc*) dans les cellules des centres germinatifs issus des plaques de Peyer des souris SATB1 déficientes et sauvages (n=6-9).

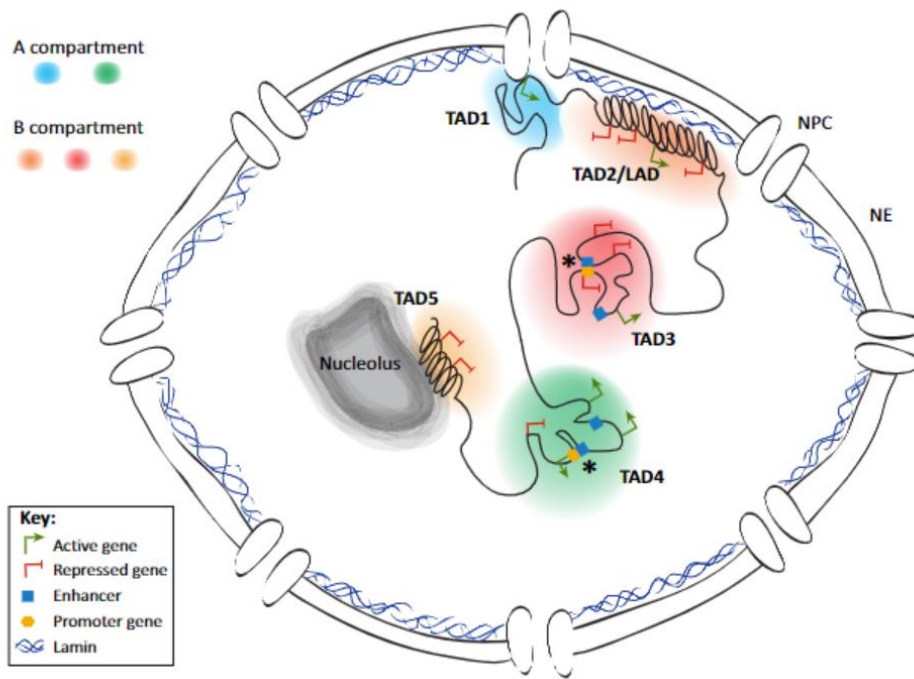
la transcription était un prérequis indispensable aux mutations somatiques (Fukita et al., 1998; Tumas-Brundage and Manser, 1997). L'absence de la protéine SATB1, favorisant la transcription d'IgH et Igk au sein des cellules activées du centre germinatif, contribuerait à augmenter l'accessibilité de ces loci à AID et, par conséquent, à augmenter aussi les mutations somatiques aux cibles « légitimes » que sont les gènes d'Ig. Cependant, nos études portant sur l'hypermutation somatique dans les gènes cibles illégitimes de AID dans notre modèle (tels que *Bcl6*, *Pim1*, *Cd83*, *Pax5* et *C-myc*) suggèrent que l'effet modulateur de SATB1 sur la SHM n'est pas forcément en lien direct avec un effet transcriptionnel. En effet, la fréquence de mutations est significativement augmentée sur *Bcl6* et *Pim1* dans notre modèle KO, en accord ici avec un aspect « protecteur » de la protéine qui limite l'apparition de mutation sur ces oncogènes (**Figure 20**). Pour autant, la transcription de ces gènes n'est pas augmentée dans les cellules B du centre germinatif dépourvues de l'expression de SATB1 (**Figure 20**). Ces données suggèrent alors que si l'augmentation de la SHM aux gènes d'Ig pourrait être un effet secondaire d'une transcription modulée, le mécanisme d'action de SATB1 aux cibles illégitimes d'AID semblerait, quant à lui, impacté d'une toute autre manière. Cependant, nous constatons également que d'autres gènes cibles illégitimes ne subissent pas de biais de mutation en l'absence de SATB1 ; c'est notamment le cas pour *C-myc*, *Pax 5* et *Cd83* (**Figure 20**). Il faut également noter que l'absence la MAR BP ne semble pas non plus moduler leurs transcriptions (**Figure 20**). Les conséquences de la délétion de SATB1 sur les mutations aux gènes cibles illégitimes de AID sont donc différentes et spécifiques de chaque gène ; nos analyses ne permettent pas d'établir un lien avec la transcription de ces régions.

Néanmoins, une hypothèse pour expliquer l'influence de SATB1 sur la SHM des gènes cibles illégitimes pourrait être en lien avec la fonction d'organisateur de la structure nucléaire de SATB1, largement étudiée dans les cellules T.

En lien avec ces hypothèses, il semble également intéressant de rechercher, dans un premier temps à l'aide d'outils bio-informatiques, la présence de sites potentiels pour la fixation de SATB1 et/ou la présence de régions similaires à des MARs dans les cibles illégitimes de AID.

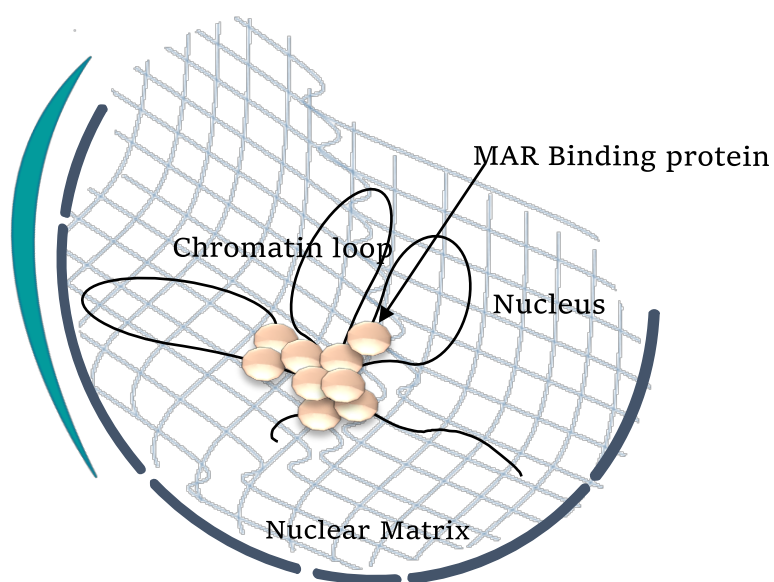
Nous avons par ailleurs entrepris une étude de l'accessibilité du génome des cellules B (matures et plasmocytes) dans notre modèle par ATAC-seq ; cette analyse pourrait apporter des informations sur 1) l'accessibilité des régions soumises à l'hypermutation somatique (cibles légitimes et illégitimes) dans les cellules B dépourvues de la protéine ; 2) d'éventuelles différences dans l'accessibilité du génome des cellules B dépourvues de SATB1 qui pourraient les prédisposer à des variations d'expression ou des mutations (par exemple des gènes cibles de SATB1 connus dans le lignage T).

Les noyaux des cellules eucaryotes sont organisés à différentes échelles. Les compartiments A et B définissent des zones distinctes de la chromatine. Si le compartiment A abrite l'euchromatine



**Figure 21 : Illustration de l'organisation nucléaire sous forme de TAD/LAD.**

Le noyau des cellules eucaryotes est organisé sous forme de domaines topologiquement associés parfois liés aux protéines composant la matrice nucléaire. Cette organisation permet de regrouper des gènes transcriptionnellement régulés de la même façon en établissant des interactions entre promoteurs et enhanceurs. D'après Gonzales-Sandoval & Gasser Trends Genet. Aug 2016 Vol.32



**Figure 22 : Représentation schématique de la formation de boucle de chromatine par l'association des MAR BP avec les régions MAR et la matrice nucléaire.**

transcriptionnellement active, le compartiment B correspond, quant à lui, à l'hétérochromatine caractérisée par un état réprimé. Au sein de chacun de ses compartiments, s'établissent des domaines topologiques associés ou TADs (Topologically Associated Domains) qui définissent une région dans laquelle les gènes sont capables d'interaction en général définies comme des interactions en CIS (bien que théoriquement possibles, les interactions en TRANS sont controversées). Une échelle organisationnelle encore plus fine se définit par une organisation en sous-TADs qui regroupe des gènes dont les interactions sont d'autant plus fortes qu'elles forment des boucles de chromatine (par exemple entre régions régulatrices, enhancers et promoteurs). Les extrémités des TADs, enrichies en protéines de structuration telles que CTCF (CTC-binding factor), inhibent les interactions entre portions de TAD différent (**Figure 21**) (Ea et al., 2015). La délétion de CTCF conduit ainsi à la déstructuration des TADs, mais l'organisation structurelle A-B ne semble pas atteinte, prêtant à CTCF un rôle d'organisateur nucléaire à échelle définie. Le contrôle transcriptionnel des gènes est donc dépendant de leur localisation au sein des sous structures nucléaires.

SATB1 possède aussi la capacité d'organiser la structure nucléaire en boucle de chromatine, notamment via sa liaison aux protéines de la matrice nucléaire (**Figure 22**) (De Belle et al., 1998; Seo et al., 2005). Ce facteur est initialement décrit dans les lymphocytes T comme un organisateur majeur du noyau formant des structures « cage-like ». Plus récemment, il est décrit que cette protéine formerait, dans les cellules T, des structures sphériques et régulières principalement dans les régions d'euchromatine. Ces études comparent, dans les noyaux des cellules T, des données haut débit relatives à CTCF, un organisateur nucléaire ubiquitaire des cellules eucaryotes et SATB1. Les auteurs montrent que la majorité des boucles médiées par SATB1 sont comprises dans les boucles formées par CTCF. La délétion de SATB1 n'entraîne pas, contrairement à celle de CTCF, la déstructuration des TADs. En revanche, les auteurs assignent à SATB1 la capacité de former les boucles à une échelle plus fine, médiant des interactions entre enhancer et promoteur. Si CTCF participe à la structuration plus grossière des TADs dans le noyau, SATB1 contribuerait, quant à lui, à l'organisation interne au sein des TADs (il pourrait s'agir de sous-TAD) (Zelenka et al., 2021a).

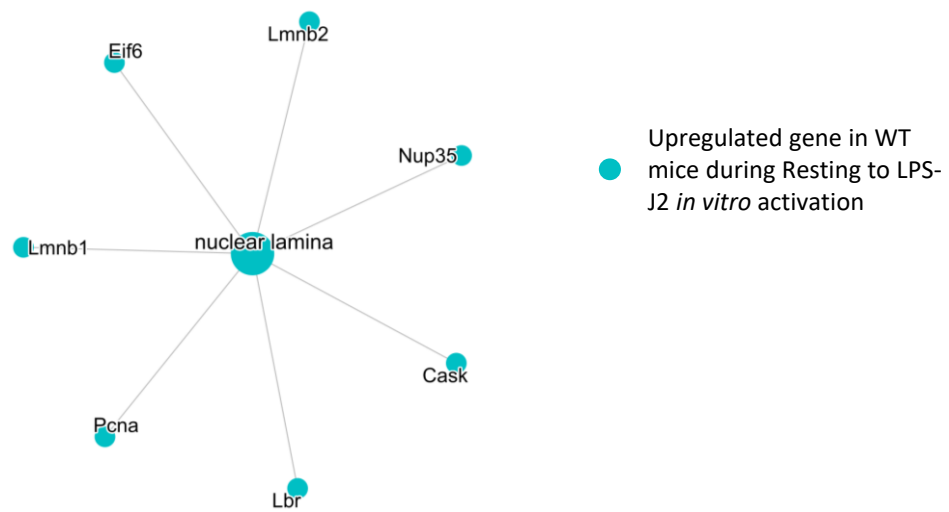
Ces données récentes de la littérature, associées à nos données d'hypermutation somatique des gènes illégitimes de AID, pourraient suggérer une implication de SATB1 dans l'organisation nucléaire du lymphocyte B. Le rôle proposé de « protecteur d'oncogènes des cellules B contre la SHM » pour SATB1, pourrait impliquer sa fonction de maintien de l'organisation du génome, par exemple par le maintien d'une organisation en sous TADs. Aux loci des gènes d'Ig, une telle organisation en sous TAD, soutenue par la présence des régions MARSEμ capables de lier SATB1, permettrait « d'isoler » l'exon variable réarrangé afin d'y favoriser la transcription et la SHM. La conséquence de la formation dynamique d'un tel sous-TAD pourrait limiter le ciblage de mutations qui pourrait être préjudiciable aux régions situées



en aval des enhancers introniques. Il reste néanmoins indispensable de réaliser des expériences de CHIP-seq (Chromatin Immunoprecipitation- sequencing) afin de raffiner l'étude de notre modèle et de valider des éléments de mécanisme contribuant à positionner la machinerie de SHM. Ces éléments pourraient différer selon les chromosomes concernés (gènes d'Ig mais également les cibles illégitimes d'AID) et la fixation de la MAR BP. Ces analyses couplées à des expériences de FISH-3D (Fluorescence *in situ* hybridization) permettant de définir la localisation des gènes d'Ig et des gènes illégitimes, permettraient de valider un éventuel rôle d'organisateur nucléaire de SATB1 au sein des noyaux des cellules B, et contribueraient potentiellement à la compréhension de sa fonction dans le mécanisme de SHM.

Des données non publiées de notre équipe montrent que, dans les cellules de centre germinatifs, la délétion des régions MARsE $\mu$  induit une délocalisation d'IgH vers le centre du noyau, alors que ce dernier adopte une position plus périphérique dans les cellules sauvages. Ces observations soutiennent la fonction d'attachement à la matrice nucléaire du locus IgH médiée par les MAR. Cette relocalisation nucléaire de ce locus et le biais d'hypermutation somatique (décrit dans ce même modèle MARsE $\mu$  KO) évoque l'importance de la localisation des gènes au sein du noyau des cellules B pour le mécanisme d'hypermutation.

Initialement décrites comme des régions inactives de la chromatine, les LADs (Lamin Associated Domain) regroupent un ensemble de gènes interagissant avec des protéines péri-nucléaires, telles que les lamines, qui sont des composants de la matrice du noyau. Les mécanismes de liaisons des gènes à la matrice nucléaire demeurent mal décrits mais certains font intervenir les régions régulatrices des gènes telles que les MAR et leurs protéines associées. Plus récemment, la littérature fait état de plusieurs types de LADs basés sur la dynamique de leur formation et sur le niveau d'expression des gènes qui y sont associés. Les c-LADs (LADs conventionnels) sont associés à un état répressif de la chromatine tandis que les e-LADs sont formés par des régions d'euchromatine transcriptionnellement active. Des expériences d'immuno-FISH, combinant le marquage de gènes ciblés par la SHM et SATB1, pourraient imager la contribution de SATB1 et sa possible fixation sur les gènes d'Ig, en lien avec la matrice nucléaire et plus particulièrement les lamines. Une précédente étude présente la Lamine B1 comme un répresseur du mécanisme d'hypermutation somatique. En revanche, les données du modèle MARsE $\mu$  KO pourraient, à l'inverse, suggérer un rôle potentiellement activateur. Des expériences de « Cut and Run », nouvellement adaptées à la détection de régions associées aux lamines, apporteraient des éléments indispensables à la définition de nos régions d'intérêt (MAR des loci des gènes d'Ig) en tant que e-LAD.



**Figure 23 : Analyse de pathway Gene ontology – cellular component à partir des données de RNA-sèq de cellules B spléniques matures et activées *in vitro* deux jours au LPS des souris SATB1 déficientes comparées aux souris WT.**

Dans un contexte plus général, il est également possible que, indépendamment de la transcription des gènes cibles, la délétion de SATB1 perturbe l'expression et/ou la localisation de facteurs contribuant au mécanisme d'hypermutation. L'analyse transcriptomique que nous avons réalisée en comparant des cellules B spléniques au repos et activées *in vitro* par LPS, ne révèle que peu de gènes différentiellement exprimés en l'absence de SATB1. Ces résultats indiquent que, lors de l'activation des lymphocytes B, SATB1 n'occupe pas la fonction d'un régulateur majeur du transcriptome mais plutôt d'un facteur spécialisé. La comparaison de l'expression des gènes codant les acteurs de la SHM tels que AID, les acteurs des voies BER, MMR et les polymérases infidèles n'indique aucune différence à l'activation des cellules WT et déficientes pour SATB1 ; suggérant que l'origine du défaut de SHM ne résulte pas d'un défaut d'expression de l'un d'entre eux. Une analyse complémentaire des voies d'activation (pathway) met en avant, lors de l'activation des cellules B sauvages, une augmentation de l'expression de plusieurs gènes impliqués dans la fonction de la lamina qui compose la matrice nucléaire. De façon surprenante, la délétion de SATB1 semble éteindre l'expression de ce réseau de gènes parmi lesquels on trouve, entre autre, les gènes codant les lamines B1 et B2 et des protéines reliant la lamina et l'enveloppe nucléaire (**Figure 23**). A l'heure actuelle, la contribution des composants de la matrice nucléaire au développement B et aux remaniements associés n'est pas connue.

En lien avec les données du modèle MAR $\epsilon\mu$  KO et l'altération des voies de réparation « error prone » observée dans ce dernier, notre modèle déficient pour SATB1 pose la question d'une implication éventuelle de cette MAR BP dans les mécanismes de réparation situés en aval de la déamination par AID. Notre hypothèse actuelle propose que les MAR $\epsilon\mu$  formeraient une « barrière » favorisant et délimitant l'action des acteurs infidèles de la voie de réparation BER ; les composants de la barrière restent cependant à identifier. Une étude, dans des modèles cellulaires, met aussi en lien SATB1 et le recrutement de OGG1, une glycosylase contribuant à la voie BER (Kaur et al., 2016). Ces données pourraient suggérer un éventuel impact de la délétion de SATB1 sur l'efficacité des mécanismes de réparation sous-jacent à l'hypermutation somatique. Il serait pertinent de déterminer s'il s'agit de mécanismes « error prone » ou « error free ». L'étude de l'effet de la délétion de la protéine SATB1 dans un contexte déficient pour les voies de réparation (UNG-MSH KO) permettrait de spécifier l'implication de cette MAR BP au cours des mécanismes de réparation associés à la SHM.

Bien que SATB1 semble impliquée dans le mécanisme d'hypermutation somatique, il semblerait, en revanche, que ce facteur ne contribue pas (ou peu) au mécanisme de la CSR: en effet la transcription des régions donneuse *I $\mu$*  et acceptrices *Ig3* et *Ig2b*, induite par la stimulation LPS, n'est pas modifiée en l'absence de SATB1. Cependant nous observons, lors des stimulations *in vitro*, couplant le TGF- $\beta$  (qui favorise le switch à IgA) au LPS, un accroissement de la production d'IgA. Cette potentielle





spécificité nécessite une exploration approfondie par la réalisation d'expériences complémentaires et pose également la question de la présence d'un éventuel site de fixation pour SATB1 au sein du promoteur  $\alpha$ .

D'autres hypothèses, quant au mode d'action de SATB1, ne sont pas à exclure. Récemment l'équipe de Spilianakis montre qu'il existe une isoforme plus courte de la protéine (Zelenka et al., 2021b). Cet isoforme contribuerait à adapter la fonction de la protéine. Cette équipe démontre également que cette MAR BP serait impliquée à la fois dans des processus d'épissage mais également dans le phénomène de séparation de phase. Connu depuis longtemps, ce phénomène est de plus en plus évoqué pour expliquer la formation d'organelles ou de structures auto-assemblées qui contribuent à l'organisation du noyau en concentrant certains composants protéiques à des régions de chromatine. Parmi ces organelles, l'on retrouve notamment les corps Cajal et les corps nucléaires PML (pour revue Wang et al., 2021). L'ensemble de ces structures contribue à l'organisation nucléaire et à la régulation de l'expression des gènes. Une étude a montré que SATB1 interagissait avec la protéine PML1 au sein de corps nucléaires PML afin de réguler l'expression des gènes du locus *CMH1* (Kumar P. et al., 2007).

SATB1 est aussi décrite pour être capable de s'associer à certains lncRNA tels que Xist, Malat et Neat qui sont impliqués dans l'organisation nucléaire (Zelenka et al., 2021b). D'autres études mettent aussi en avant l'implication de ces lncRNA et du RNA exosome lors du processus de SHM. L'étude de l'hypermutation somatique sur certains gènes cibles illégitimes d'AID montre un biais de recrutement de la déaminase en absence du RNA exosome suggérant que l'accumulation des lncRNA pourrait bloquer l'accessibilité des gènes à AID (Pefanis and Basu, 2015; Pefanis et al., 2014).

La complexité des mécanismes de remaniements de la cellule B et les rares études du rôle de la protéine SATB1 au sein de ces cellules ne permettent pas, pour l'instant, d'éliminer ces hypothèses.

De façon similaire à SATB1, la protéine ARID3A ne semble pas être indispensable au développement précoce du lymphocyte B, cependant sa délétion influence l'expression du BCR. L'étude de notre modèle murin indique que cette MAR BP ne contribue pas à l'expression de la chaîne lourde mais il semblerait qu'elle participe à l'expression de la chaîne légère Igk dans les cellules B matures. Cette donnée pourrait expliquer que les cellules B ARID3A déficientes présentent une altération de l'expression du BCR malgré une expression normale de la chaîne lourde.

Néanmoins, ces données nécessitent d'être consolidées par des expériences permettant d'évaluer et comparer la quantité de chaînes légères produites dans le modèle de délétion conditionnelle, dès les stades B immatures. C'est pourquoi nous envisageons de réaliser des marquages intracellulaires des



chaînes légères et des analyses par cytométrie en flux de ces populations B. En complément, l'expression des chaînes légères pourra aussi être déterminée par western blot dans les cellules B isolées de la rate. De plus, bien que le taux de cellules B exprimant une chaîne légère Ig $\lambda$  soit faible chez la souris, il pourrait tout de même être intéressant de vérifier si la délétion de ARID3A influence également l'expression de cette seconde chaîne légère. Si nous confirmons l'hypothèse selon laquelle ARID3A pourrait moduler la transcription de la chaîne légère aux stades B immatures et naïfs, la question d'un effet direct ou indirect de ARID3A présentera alors un intérêt majeur. La littérature ne mentionne pas de liaison de ARID3A à la région MAREik, mais cette interaction reste envisageable. De même que pour SATB1, il semble alors pertinent de rechercher à l'aide d'un outil bio-informatique, la présence de sites potentiels pour la fixation de ARID3A dans les loci des chaînes légères d'Ig. Par la suite, ces données pourraient être complétées par des expériences de ChIP-seq afin de rechercher la fixation de la protéine sur les régions régulatrices, promoteurs et ou enhanceurs, des loci des chaînes légères. Dans le cas d'un éventuel effet indirect lié à la délétion de ARID3A, les expériences de RNA-seq réalisées sur nos modèles permettent une analyse différentielle des transcriptomes des cellules B au repos. Actuellement en cours de réalisation, ces études devraient apporter des éléments de réponse sur d'éventuelles dérégulations globales de gènes cibles de ARID3A tel que le facteur de transcription *Oct4*, exprimé dans le lignage B (Popowski et al., 2014).

Dans les cellules B activées dans lesquelles l'expression d'ARID3A est la plus importante, la délétion de la protéine ne semblerait plus limiter l'expression de la chaîne légère et pourrait même induire un effet opposé.

De nombreuses études répertorient des effets différents de la délétion ou la surexpression d'ARID3A sur le développement des cellules B, néanmoins dans notre modèle, la protéine, dispensable pour le développement précoce, intervient de façon plus importante aux stades plus tardifs. Bien que nos données ne montrent pas de défaut de prolifération (induite par le LPS *in vitro*) des cellules B dépourvues de ARID3A, il semble malgré tout que la suppression de ARID3A impacte la dynamique de formation des cellules du CG ainsi que la différenciation plasmocytaire. Les effets accentués de la délétion de ARID3A, observés à ces stades, sont en corrélation avec l'augmentation de son expression physiologique dans ces populations.

Au sein des cellules B du CG, ARID3A contribuerait à optimiser l'hypermutation somatique. En effet, sa délétion entraîne une altération de la fréquence de SHM de façon significative au locus *IgH*, et de manière plus faible au locus *IgK*. Ces données, mises en parallèle des analyses de la transcription de ces gènes dans les B du CG ARID3A cKO, sont corrélées aux observations faites lors de l'étude du modèle MARSE $\mu$  KO. En effet, la littérature décrit majoritairement ARID3A comme un facteur

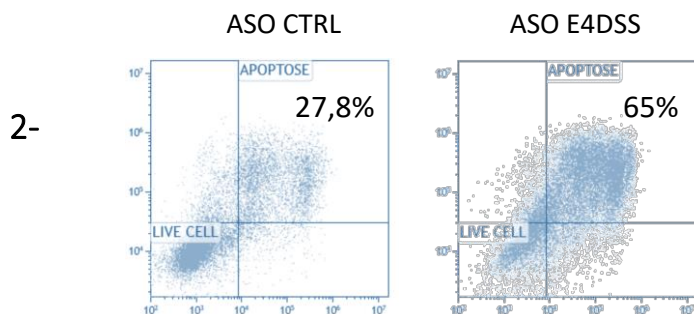
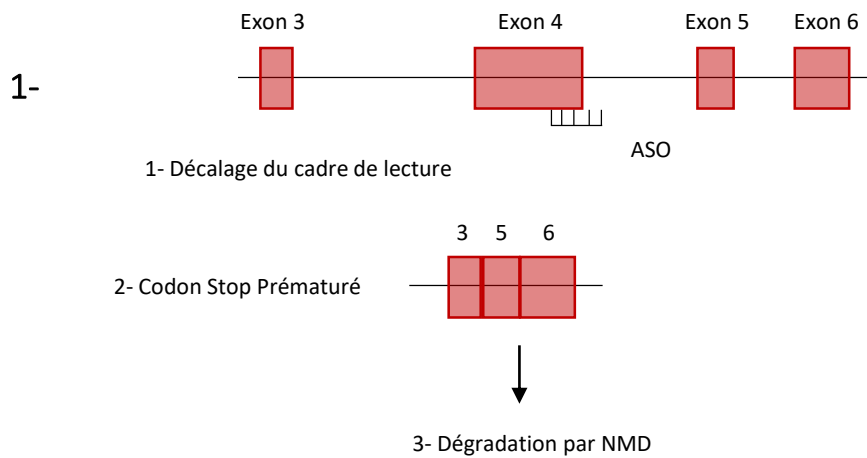


activateur de la transcription d'IgH, notre modèle murin en revanche, indique que ce n'est pas le cas dans les cellules B du CG puisque ces dernières présentent une légère augmentation, non-significative mais reproductible, des transcrits primaires au locus IgH. Les deux modèles (ARID3A cKO et MARsEμ KO) présentent donc, tous deux, une altération de la fréquence d'hypermutation somatique sur *IgH* malgré une transcription qui semble augmentée. Le modèle de délétion des régions MAR, montrant un biais de recrutement des acteurs de la réparation « error prone » (**Annexe1**), et la corrélation des données transcriptionnelles et de SHM entre les deux modèles posent alors la question d'une éventuelle contribution d'ARID3A dans le recrutement des acteurs de la réparation de l'ADN sous-jacente à la SHM. Le croisement de notre modèle conditionnel avec les modèles de délétion des premiers acteurs des voies de réparation (UNG/MSH2) pourrait apporter, tout comme pour le modèle MARsEμ KO, des éléments de réponse pour cette hypothèse.

Dans notre modèle ARID3A cKO, il semble que la diminution de la SHM au locus *IgH* n'a pas pour origine une altération de la transcription, comme observé dans les modèles animaux dépourvus de régions activatrices en 3' du locus (Bruzeau et al., 2021). Néanmoins, il s'avère, à ce stade de notre étude, indispensable d'analyser la composition des sous population B du CG et d'évaluer leur capacité de prolifération dans le but de s'assurer que le défaut de SHM dans notre modèle ne soit pas lié à un défaut de dynamique et de sélection cellulaire.

Cependant, l'étude de l'hypermutation somatique aux gènes cibles illégitimes de AID nous donne des éléments allant à l'encontre de cette hypothèse. Ces mutations, considérées comme des dommages collatéraux de l'action de l'enzyme, constitueraient un bon indicateur de la dynamique cellulaire des populations B dans le centre germinatif. Nous avons montré par le passé, à l'aide de notre outil DeMinEr, que les mutations aux cibles majeures « off targets » que sont *Cd83* et *Bcl6* s'accumulent dans les cellules B de CG des plaques de Peyer avec l'âge de l'animal, vraisemblablement sans biais de sélection (Martin et al., 2018). Dans notre modèle, les études portant sur les fréquences de SHM dans ces cibles illégitimes, indiquent que la suppression de ARID3A n'altère pas la SHM de façon systématique. En effet, les mutations dans les gènes *Bcl6*, *Pim1* et *Cd83* ne sont pas significativement diminuées en l'absence de ARID3A dans des cellules B du CG de plaques de Peyer. Ces données suggèrent qu'une partie du processus de SHM, dont le ciblage par AID constitue l'origine, se trouve fonctionnel dans les cellules du CG dépourvues de ARID3A. Une confirmation de cette hypothèse pourrait être apportée par l'évaluation des mutations cumulées dans ces oncogènes chez des animaux plus âgés.

En dehors d'un processus de sélection et de dynamique du centre germinatif, plusieurs hypothèses peuvent être formulées pour expliquer un rôle de ARID3A dans le mécanisme d'hypermutation.



**Figure 24 : Test de l'efficacité de l'oligonucléotide antisens ciblant ARID3A.**

1- Représentation schématique du fonctionnement de l'oligonucléotide antisens (ASO) ciblant l'exon 4 qui génère un saut d'exon créant l'apparition d'un codon stop prématuré induisant ainsi la dégradation de l'ARN par le mécanisme spécialisé NMD (nonsense mediated decay)

2- Analyse par cytométrie en flux de la survie cellulaire des cellules Ocity 3 (DLBCL) après 48 heures d'incubation en présence d'un oligonucléotide antisens ciblant ARID3A (saut de l'exon 4).

ARID3A permet aussi d'accroître l'accessibilité de ses gènes cibles et on le retrouve impliqué également au sein des corps nucléaires PML (Fukuyo et al., 2004; Zong et al., 2000). La description « historique » de cette localisation renforce l'hypothèse selon laquelle ARID3A interagit à la fois avec la chromatine et les composants de la lamina péri nucléaire. La question de la fonction d'un tel attachement sur l'accessibilité, l'expression et la réparation de gènes au sein des lymphocytes B reste, à ce jour inexplorée. Nous avons entrepris des expériences d'ATAC-seq afin d'évaluer l'impact de la suppression de ARID3A sur l'accessibilité globale du génome des cellules B ; ces données seront mises en parallèle de nos données de RNA-seq afin de tenter de corrélérer, dans un premier temps, accessibilité et transcription.

Enfin, nos données indiquent clairement une contribution de ARID3A dans la différenciation plasmocytaire. Nous souhaitons confirmer cette hypothèse à l'aide d'un modèle complémentaire de différenciation plasmocytaire *in vitro*, initié à partir de cellules primaires humaines. Ces expériences sont actuellement en cours de mise au point. Dans ce modèle cellulaire, utilisé depuis peu dans notre laboratoire et mis au point par des collaborateurs, l'expression de ARID3A sera supprimée par une stratégie utilisant un oligonucléotide anti-sens (ASO) permettant d'induire un saut d'exon qui conduit à la production d'un messager non-fonctionnel qui sera éliminé par la machinerie de surveillance des ARN (**Figure 24**) (Marchalot et al., 2020). Des premiers tests réalisés sur une lignée cellulaire de DLBCL (diffus large B cell lymphoma) montrent que la délétion de ARID3A provoque la mort d'une partie des cellules transduites (**Figure 24**). Si ces résultats se confirment par des expériences complémentaires dans des modèles cellulaires de même type, ces derniers ouvrent un champ d'investigation intéressant en lien avec les pathologies cancéreuses du lignage B.

En effet, l'impact de la délétion de ARID3A compromettant à la fois la différenciation plasmocytaire et la survie de lignées tumorales pourrait appuyer l'hypothèse de son implication dans le développement de processus tumoraux tels que les myélomes multiples ou les DLBCL. Des études réalisées sur un modèle murin développant de façon spontanée des leucémies lymphocytaires chroniques (LLC) montre que la délétion de la protéine pourrait freiner l'émergence de la maladie (Hayakawa et al., 2019). L'expression de ARID3A est également décrite comme étant une signature des lymphomes DLBCL de type ABC (Care et al., 2013). Enfin, une surexpression du gène *Arid3a* a également été remarquée dans un modèle murin, établi dans notre laboratoire, exprimant une version mutée constitutivement active de la protéine MYD88 (MYD88 L252P) impliquée dans différentes pathologies telles que la maladie de Waldenström et des DLBCL. Ce modèle présente une augmentation importante du pourcentage de cellule lympho-plasmocytaires considérées comme tumorales (**Annexe 3**) (Ouk et al., 2021). Le croisement des souris MYD88 L252P avec notre modèle de délétion de la protéine ARID3A pourrait confirmer l'implication de ARID3A dans la survenue de tumeurs et présenter





des données intéressantes (retard de développement de la pathologie) quant à la fonction de la MAR BP sur l'émergence de pathologie tumorale et la différenciation plasmocytaire.

Les deux protéines étudiées au cours de notre étude présentent toutes deux des rôles opposés sur la régulation du processus d'hypermutation somatique, suggérant une possible compétition pour leur fixation sur les régions MAR ; un tel mécanisme a, par ailleurs, été décrit entre CUX1 et ARID3A (Kim et al., 2017).

SATB1 et ARID3A sont aussi capables d'interagir avec leur homologues respectifs (SATB2 et ARID3B). Il serait donc intéressant d'étudier l'effet de la délétion de SATB2, notamment en raison de sa régulation par le facteur d'épissage SFSR1-3 qui contribue au mécanisme de l'hypermutation et qui favorise son adressage au locus des chaînes légères (Singh et al., 2020). Enfin, il est décrit que chacune des protéines étudiées au cours de ma thèse présente certaines fonctions redondantes avec leurs homologues. La réalisation de modèles présentant la délétion conjointe des deux facteurs (SATB1/SATB2 et ARID3A/ARID3B) pourrait, bien que techniquement complexe, contribuer à la compréhension de leur implication respective au cours du processus d'hypermutation.



## Références bibliographiques

---

- Acevedo, N., Reinius, L.E., Vitezic, M., Fortino, V., Söderhäll, C., Honkanen, H., Veijola, R., Simell, O., Toppari, J., Ilonen, J., et al. (2015). Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes. *Clin. Epigenetics* 7. <https://doi.org/10.1186/s13148-015-0064-6>.
- Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L.A.M., et al. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* 121, 295–306. <https://doi.org/10.1016/j.cell.2005.02.013>.
- Agrelo, R., Kishimoto, H., Novatchkova, M., Peraza, V., Paolino, M., Souabni, A., and Wutz, A. (2013). SATB1 collaborates with loss of p16 in cellular transformation. *Oncogene* 32, 5492–5500. <https://doi.org/10.1038/onc.2013.158>.
- Ahlfors, H., Limaye, A., Elo, L.L., Tuomela, S., Burute, M., Gottimukkala, K.V.P., Notani, D., Rasool, O., Galande, S., and Lahesmaa, R. (2010). SATB1 dictates expression of multiple genes including IL-5 involved in human T helper cell differentiation. *Blood* 116, 1443–1453. <https://doi.org/10.1182/blood-2009-11-252205>.
- Akhavantabasi, S., Sapmaz, A., Tuna, S., and Erson-Bensan, A.E. (2012). miR-125b Targets ARID3B in Breast Cancer Cells. *Cell Struct. Funct.* 37, 27–38. <https://doi.org/10.1247/csf.11025>.
- Alcamo, E.A., Chirivella, L., Dautzenberg, M., Dobрева, G., Fariñas, I., Grosschedl, R., and McConnell, S.K. (2008). Satb2 Regulates Callosal Projection Neuron Identity in the Developing Cerebral Cortex. *Neuron* 57, 364–377. <https://doi.org/10.1016/j.neuron.2007.12.012>.
- 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. <https://doi.org/10.1128/mcb.11.4.2096-2107.1991>.
- 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. <https://doi.org/10.1038/ni1100>.
- Allen, C.D.C., Okada, T., and Cyster, J.G. (2007). Germinal-Center Organization and Cellular Dynamics. *Immunity* 27, 190–202. <https://doi.org/10.1016/j.immuni.2007.07.009>.
- 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. <https://doi.org/10.1016/j.dnarep.2007.01.009>.
- 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. <https://doi.org/10.1002/j.1460-2075.1984.tb01955.x>.
- Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., and Kohwi-Shigematsu, T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev.* 14, 521–535. <https://doi.org/10.1101/gad.14.5.521>.

- 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* 28, 1560–1567. <https://doi.org/10.1002/stem.491>.
- An, N., Khan, S., Imgruet, M.K., Gurbuxani, S.K., Konecki, S.N., Burgess, M.R., and McNerney, M.E. (2018). Gene dosage effect of CUX1 in a murine model disrupts HSC homeostasis and controls the severity and mortality of MDS. *Blood* 131, 2682–2697. <https://doi.org/10.1182/blood-2017-10-810028>.
- Angelin-Duclos, C., Cattoretti, 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.* 165, 5462–5471. <https://doi.org/10.4049/jimmunol.165.10.5462>.
- Aoki-Ota, M., Torkamani, A., Ota, T., Schork, N., and Nemazee, D. (2012). Skewed Primary Igk Repertoire and V–J Joining in C57BL/6 Mice: Implications for Recombination Accessibility and Receptor Editing. *J. Immunol.* 188, 2305–2315. <https://doi.org/10.4049/jimmunol.1103484>.
- Arakawa, H., Shimizu, T., and Takeda, S. (1996). Re-evaluation of the probabilities for productive rearrangements on the  $\kappa$  and  $\lambda$  loci. *Int. Immunol.* 8, 91–99. <https://doi.org/10.1093/intimm/8.1.91>.
- Bacharier, L.B., and Geha, R.S. (2000). Molecular mechanisms of IgE regulation. *J. Allergy Clin. Immunol.* 105, S547–S558. [https://doi.org/10.1016/S0091-6749\(00\)90059-9](https://doi.org/10.1016/S0091-6749(00)90059-9).
- Balamotis, M.A., Tamberg, N., Woo, Y.J., Li, J., Davy, B., Kohwi-Shigematsu, T., and Kohwi, Y. (2012). Satb1 Ablation Alters Temporal Expression of Immediate Early Genes and Reduces Dendritic Spine Density during Postnatal Brain Development. *Mol. Cell. Biol.* 32, 333–347. <https://doi.org/10.1128/MCB.05917-11>.
- Balázs, M., Martin, F., Zhou, T., and Kearney, J.F. (2002). Blood Dendritic Cells Interact with Splenic Marginal Zone B Cells to Initiate T-Independent Immune Responses. *Immunity* 17, 341–352. [https://doi.org/10.1016/S1074-7613\(02\)00389-8](https://doi.org/10.1016/S1074-7613(02)00389-8).
- Banan, M., Rojas, I.C., Lee, W.-H., King, H.L., Harriss, J.V., Kobayashi, R., Webb, C.F., and Gottlieb, P.D. (1997). Interaction of the Nuclear Matrix-associated Region (MAR)-Binding Proteins, SATB1 and CDP/Cux, with a MAR Element (L2a) in an Upstream Regulatory Region of the Mouse CD8a Gene. *J. Biol. Chem.* 272, 18440–18452. <https://doi.org/10.1074/jbc.272.29.18440>.
- Bankovich, A.J., Raunser, S., Juo, Z.S., Walz, T., Davis, M.M., and Garcia, K.C. (2007). Structural Insight into Pre-B Cell Receptor Function. *Science* 316, 291–294. <https://doi.org/10.1126/science.1139412>.
- Barajas-Mora, E.M., Kleiman, E., Xu, J., Carrico, N.C., Lu, H., Oltz, E.M., Murre, C., and Feeney, A.J. (2019). A B-Cell-Specific Enhancer Orchestrates Nuclear Architecture to Generate a Diverse Antigen Receptor Repertoire. *Mol. Cell* 73, 48–60.e5. <https://doi.org/10.1016/j.molcel.2018.10.013>.
- Baranello, L., Kouzine, F., and Levens, D. (2013). DNA Topoisomerases: Beyond the standard role. *Transcription* 4, 232–237. <https://doi.org/10.4161/trns.26598>.
- Barber, C.L., Montecino-Rodriguez, E., and Dorshkind, K. (2011). Reduced production of B-1–specified common lymphoid progenitors results in diminished potential of adult marrow to generate B-1 cells. *Proc. Natl. Acad. Sci.* 108, 13700–13704. <https://doi.org/10.1073/pnas.1107172108>.

- Bassing, C.H., Alt, F.W., Hughes, M.M., D'Auteuil, M., Wehrly, T.D., Woodman, B.B., Gärtner, F., White, J.M., Davidson, L., and Sleckman, B.P. (2000). Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule. *Nature* *405*, 583–586. <https://doi.org/10.1038/35014635>.
- Basso, K., Saito, M., Sumazin, P., Margolin, A.A., Wang, K., Lim, W.-K., Kitagawa, Y., Schneider, C., Alvarez, M.J., Califano, A., et al. (2010). Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. *Blood* *115*, 975–984. <https://doi.org/10.1182/blood-2009-06-227017>.
- 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. <https://doi.org/10.1016/j.cell.2011.01.001>.
- Baumgarth, N. (2011). The double life of a B-1 cell: self-reactivity selects for protective effector functions. *Nat. Rev. Immunol.* *11*, 34–46. <https://doi.org/10.1038/nri2901>.
- Bekeredjian-Ding, I., and Jego, G. (2009). Toll-like receptors - sentries in the B-cell response. *Immunology* *128*, 311–323. <https://doi.org/10.1111/j.1365-2567.2009.03173.x>.
- Bergmann, B., Grimsholm, O., Thorarinsdottir, K., Ren, W., Jirholt, P., Gjertsson, I., and Mårtensson, I.-L. (2013). Memory B Cells in Mouse Models. *Scand. J. Immunol.* *78*, 149–156. <https://doi.org/10.1111/sji.12073>.
- Beyer, M., Thabet, Y., Müller, R.-U., Sadlon, T., Classen, S., Lahl, K., Basu, S., Zhou, X., Bailey-Bucktrout, S.L., Krebs, W., et al. (2011). Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. *Nat. Immunol.* *12*, 898–907. <https://doi.org/10.1038/ni.2084>.
- Bobbs, A., Gellerman, K., Hallas, W.M., Joseph, S., Yang, C., Kurkewich, J., and Cowden Dahl, K.D. (2015). ARID3B Directly Regulates Ovarian Cancer Promoting Genes. *PLOS ONE* *10*, e0131961. <https://doi.org/10.1371/journal.pone.0131961>.
- 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. <https://doi.org/10.1084/jem.20092449>.
- 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. In *Advances in Immunology*, (Elsevier), pp. 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. (2004). Antisense intergenic transcription in V(D)J recombination. *Nat. Immunol.* *5*, 630–637. <https://doi.org/10.1038/ni1068>.
- Boreström, C., Forsman, A., Rüetschi, U., and Rymo, L. (2012). E2F1, ARID3A/Bright and Oct-2 factors bind to the Epstein–Barr virus C promoter, EBNA1 and oriP, participating in long-distance promoter–enhancer interactions. *J. Gen. Virol.* *93*, 1065–1075. <https://doi.org/10.1099/vir.0.038752-0>.
- 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.* *100*, 4102–4107. <https://doi.org/10.1073/pnas.0730835100>.

- Bredemeyer, A.L., Sharma, G.G., Huang, C.-Y., Helmink, B.A., Walker, L.M., Khor, K.C., Nuskey, B., Sullivan, K.E., Pandita, T.K., Bassing, C.H., et al. (2006). ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* *442*, 466–470. <https://doi.org/10.1038/nature04866>.
- 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: A novel transcription factor interacting with MARs. *Eur. J. Neurosci.* *21*, 658–668. <https://doi.org/10.1111/j.1460-9568.2005.03897.x>.
- Bruzeau, C., Moreau, J., Le Noir, S., and Pinaud, E. (2021). Panorama of stepwise involvement of the IgH 3' regulatory region in murine B cells. In *Advances in Immunology*, (Elsevier), pp. 95–114.
- 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.
- Burrows, P.D., Stephan, R.P., Wang, Y.-H., Lassoued, K., Zhang, Z., and Cooper, M.D. (2002). The transient expression of pre-B cell receptors governs B cell development. *Semin. Immunol.* *14*, 343–349. [https://doi.org/10.1016/S1044-5323\(02\)00067-2](https://doi.org/10.1016/S1044-5323(02)00067-2).
- 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. <https://doi.org/10.1038/ng1146>.
- 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. <https://doi.org/10.1038/ng1913>.
- 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). The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nat. Immunol.* *13*, 1092–1100. <https://doi.org/10.1038/ni.2418>.
- Care, M.A., Barrans, S., Worrillow, L., Jack, A., Westhead, D.R., and Tooze, R.M. (2013). A Microarray Platform-Independent Classification Tool for Cell of Origin Class Allows Comparative Analysis of Gene Expression in Diffuse Large B-cell Lymphoma. *PLoS ONE* *8*, e55895. <https://doi.org/10.1371/journal.pone.0055895>.
- Carrington, E.M., Vikstrom, I.B., Light, A., Sutherland, R.M., Londrigan, S.L., Mason, K.D., Huang, D.C.S., Lew, A.M., and Tarlinton, D.M. (2010). BH3 mimetics antagonizing restricted prosurvival Bcl-2 proteins represent another class of selective immune modulatory drugs. *Proc. Natl. Acad. Sci.* *107*, 10967–10971. <https://doi.org/10.1073/pnas.1005256107>.
- Carson, S., and Wu, G.E. (1989). A linkage map of the mouse immunoglobulin lambda light chain locus. *Immunogenetics* *29*, 173–179. <https://doi.org/10.1007/BF00373642>.
- Casanova, J.C., Uribe, V., Badia-Careaga, C., Giovinazzo, G., Torres, M., and Sanz-Ezquerro, J.J. (2011). Apical ectodermal ridge morphogenesis in limb development is controlled by *Arid3b* -mediated regulation of cell movements. *Development* *138*, 1195–1205. <https://doi.org/10.1242/dev.057570>.
- Casola, S., Otipoby, K.L., Alimzhanov, M., Humme, S., Uyttersprot, N., Kutok, J.L., Carroll, M.C., and Rajewsky, K. (2004). B cell receptor signal strength determines B cell fate. *Nat. Immunol.* *5*, 317–327. <https://doi.org/10.1038/ni1036>.

- 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. .
- Cerutti, A., Cols, M., and Puga, I. (2013). Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat. Rev. Immunol.* 13, 118–132. <https://doi.org/10.1038/nri3383>.
- 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. <https://doi.org/10.1016/j.molcel.2007.07.010>.
- Chao, M.P., Seita, J., and Weissman, I.L. (2008). Establishment of a Normal Hematopoietic and Leukemia Stem Cell Hierarchy. *Cold Spring Harb. Symp. Quant. Biol.* 73, 439–449. <https://doi.org/10.1101/sqb.2008.73.031>.
- Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* 4, 541–552. <https://doi.org/10.1038/nri1395>.
- Chauveau, C., and Cogné, M. (1996). Palindromic structure of the IgH 3'locus control region. *Nat. Genet.* 14, 15–16. <https://doi.org/10.1038/ng0996-15>.
- Chen, K., and Cerutti, A. (2011). The function and regulation of immunoglobulin D. *Curr. Opin. Immunol.* 23, 345–352. <https://doi.org/10.1016/j.coi.2011.01.006>.
- Chen, C., Prak, E.L., and Weigert, M. (1997). Editing Disease-Associated Autoantibodies. *Immunity* 6, 97–105. [https://doi.org/10.1016/S1074-7613\(00\)80673-1](https://doi.org/10.1016/S1074-7613(00)80673-1).
- 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. <https://doi.org/10.1002/j.1460-2075.1993.tb06152.x>.
- Cheng, C., Lu, X., Wang, G., Zheng, L., Shu, X., Zhu, S., Liu, K., Wu, K., and Tong, Q. (2010). Expression of SATB1 and heparanase in gastric cancer and its relationship to clinicopathologic features: SATB1 AND HEPARANASE IN GASTRIC CANCER. *APMIS* 118, 855–863. <https://doi.org/10.1111/j.1600-0463.2010.02673.x>.
- Cheng, C., Wan, F., Liu, L., Zeng, F., Xing, S., 'an, Wu, X., Chen, X., and Zhu, Z. (2014). Overexpression of SATB1 Is Associated with Biologic Behavior in Human Renal Cell Carcinoma. *PLoS ONE* 9, e97406. <https://doi.org/10.1371/journal.pone.0097406>.
- Chowdhury, D. (2001). Stepwise activation of the immunoglobulin micro heavy chain gene locus. *EMBO J.* 20, 6394–6403. <https://doi.org/10.1093/emboj/20.22.6394>.
- Chu, S.-H., Ma, Y.-B., Feng, D.-F., Zhang, H., Zhu, Z.-A., Li, Z.-Q., and Jiang, P.-C. (2012). Upregulation of SATB1 is associated with the development and progression of glioma. *J. Transl. Med.* 10, 149. <https://doi.org/10.1186/1479-5876-10-149>.
- Chung, J., Lau, J., Cheng, L.S., Grant, R.I., Robinson, F., Ketela, T., Reis, P.P., Roche, O., Kamel-Reid, S., Moffat, J., et al. (2010). SATB2 augments  $\Delta Np63\alpha$  in head and neck squamous cell carcinoma. *EMBO Rep.* 11, 777–783. <https://doi.org/10.1038/embor.2010.125>.
- 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. <https://doi.org/10.1038/ni1542>.



- Cobaleda, C., Schebesta, A., Delogu, A., and Busslinger, M. (2007). Pax5: the guardian of B cell identity and function. *Nat. Immunol.* *8*, 463–470. <https://doi.org/10.1038/ni1454>.
- Coqueret, O. (1998). The mammalian Cut homeodomain protein functions as a cell-cycle-dependent transcriptional repressor which downmodulates p21WAF1/CIP1/SDI1 in S phase. *EMBO J.* *17*, 4680–4694. <https://doi.org/10.1093/emboj/17.16.4680>.
- Coqueret, O., Bérubé, G., and Nepveu, A. (1996). DNA Binding by Cut Homeodomain Proteins Is Down-modulated by Protein Kinase C. *J. Biol. Chem.* *271*, 24862–24868. <https://doi.org/10.1074/jbc.271.40.24862>.
- Coqueret, O., Martin, N., Bérubé, G., Rabbat, M., Litchfield, D.W., and Nepveu, A. (1998). DNA Binding by Cut Homeodomain Proteins Is Down-modulated by Casein Kinase II. *J. Biol. Chem.* *273*, 2561–2566. <https://doi.org/10.1074/jbc.273.5.2561>.
- Cubelos, B., Sebastián-Serrano, A., Beccari, L., Calcagnotto, M.E., Cisneros, E., Kim, S., Dopazo, A., Alvarez-Dolado, M., Redondo, J.M., Bovolenta, P., et al. (2010). Cux1 and Cux2 Regulate Dendritic Branching, Spine Morphology, and Synapses of the Upper Layer Neurons of the Cortex. *Neuron* *66*, 523–535. <https://doi.org/10.1016/j.neuron.2010.04.038>.
- Cui, J., Qi, S., Liao, R., Su, D., Wang, Y., and Xue, S. (2020). MiR-574–5p promotes the differentiation of human cardiac fibroblasts via regulating ARID3A. *Biochem. Biophys. Res. Commun.* *521*, 427–433. <https://doi.org/10.1016/j.bbrc.2019.09.107>.
- Dausinas, P., Pulakanti, K., Rao, S., Cole, J.M., Dahl, R., and Cowden Dahl, K.D. (2020). ARID3A and ARID3B induce stem promoting pathways in ovarian cancer cells. *Gene* *738*, 144458. <https://doi.org/10.1016/j.gene.2020.144458>.
- Davies, D.R., and Metzger, H. (1983). Structural Basis of Antibody Function. *Annu. Rev. Immunol.* *1*, 87–115. <https://doi.org/10.1146/annurev.iy.01.040183.000511>.
- De Belle, I., Cai, S., and Kohwi-Shigematsu, T. (1998). The Genomic Sequences Bound to Special AT-rich Sequence-binding Protein 1 (SATB1) In Vivo in Jurkat T Cells Are Tightly Associated with the Nuclear Matrix at the Bases of the Chromatin Loops. *J. Cell Biol.* *141*, 335–348. .
- De Silva, N.S., Simonetti, G., Heise, N., and Klein, U. (2012). The diverse roles of IRF4 in late germinal center B-cell differentiation: IRF4 in late GC B-cell differentiation. *Immunol. Rev.* *247*, 73–92. <https://doi.org/10.1111/j.1600-065X.2012.01113.x>.
- 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. <https://doi.org/10.1038/311752a0>.
- Di Noia, J., and Neuberger, M.S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* *419*, 43–48. <https://doi.org/10.1038/nature00981>.
- Di Noia, J.M., Rada, C., and Neuberger, M.S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J.* *25*, 585–595. <https://doi.org/10.1038/sj.emboj.7600939>.
- 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. <https://doi.org/10.1084/jem.20030481>.

Dickinson, L., Tadashi, J., Yoshinori, K., and Terumi, K.-S. (1992). A tissue-specific MARSAR DNA-binding protein with unusual binding site recognition. *Cell* *70*, 631–645. [https://doi.org/10.1016/0092-8674\(92\)90432-C](https://doi.org/10.1016/0092-8674(92)90432-C).

Dickinson, L.A., Dickinson, C.D., and Kohwi-Shigematsu, T. (1997). An Atypical Homeodomain in SATB1 Promotes Specific Recognition of the Key Structural Element in a Matrix Attachment Region. *J. Biol. Chem.* *272*, 11463–11470. <https://doi.org/10.1074/jbc.272.17.11463>.

Dobrev, G., Dambacher, J., and Grosschedl, R. (2003). SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin  $\mu$  gene expression. *Genes Dev.* *17*, 3048–3061. <https://doi.org/10.1101/gad.1153003>.

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. <https://doi.org/10.1016/j.cell.2006.05.012>.

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. <https://doi.org/10.1038/ni.1814>.

Dominguez-Sola, D., Victora, G.D., Ying, C.Y., Phan, R.T., Saito, M., Nussenzweig, M.C., and Dalla-Favera, R. (2012). The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry. *Nat. Immunol.* *13*, 1083–1091. <https://doi.org/10.1038/ni.2428>.

Dowrey, T., Schwager, E.E., Duong, J., Mercuri, F., Zarate, Y.A., and Fish, J.L. (2019). *Satb2* regulates proliferation and nuclear integrity of pre-osteoblasts. *Bone* *127*, 488–498. <https://doi.org/10.1016/j.bone.2019.07.017>.

Durdik, J., Moore, M.W., and Selsing, E. (1984). Novel  $\kappa$  light-chain gene rearrangements in mouse  $\lambda$  light chain-producing B lymphocytes. *Nature* *307*, 749–752. <https://doi.org/10.1038/307749a0>.

Ea, V., Baudement, M.-O., Lesne, A., and Forné, T. (2015). Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. *Genes* *6*, 734–750. <https://doi.org/10.3390/genes6030734>.

Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* *19*, 981–992. [https://doi.org/10.1016/0092-8674\(80\)90089-6](https://doi.org/10.1016/0092-8674(80)90089-6).

Eberhard, J., Gaber, A., Wangefjord, S., Nodin, B., Uhlén, M., Ericson Lindquist, K., and Jirström, K. (2012). A cohort study of the prognostic and treatment predictive value of SATB2 expression in colorectal cancer. *Br. J. Cancer* *106*, 931–938. <https://doi.org/10.1038/bjc.2012.34>.

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. <https://doi.org/10.1016/j.immuni.2011.02.005>.

Edelman, G.M. (1973). Antibody Structure and Molecular Immunology. *Science* *180*, 830–840. <https://doi.org/10.1126/science.180.4088.830>.

Ellis, T., Gambardella, L., Horcher, M., Tschanz, S., Capol, J., Bertram, P., Jochum, W., Barrandon, Y., and Busslinger, M. (2001). The transcriptional repressor CDP (*Cutl1*) is essential for epithelial cell

differentiation of the lung and the hair follicle. *Genes Dev.* *15*, 2307–2319. <https://doi.org/10.1101/gad.200101>.

Emorine, L., Kuehl, M., Weir, L., Leder, P., and Max, E.E. (1983). A conserved sequence in the immunoglobulin  $\kappa$ - $\kappa$  intron: possible enhancer element. *Nature* *304*, 447–449. <https://doi.org/10.1038/304447a0>.

Enders, A., Short, A., Miosge, L.A., Bergmann, H., Sontani, Y., Bertram, E.M., Whittle, B., Balakishnan, B., Yoshida, K., Sjollem, G., et al. (2014). Zinc-finger protein ZFP318 is essential for expression of IgD, the alternatively spliced *Igh* product made by mature B lymphocytes. *Proc. Natl. Acad. Sci.* *111*, 4513–4518. <https://doi.org/10.1073/pnas.1402739111>.

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.* *106*, 5773–5778. <https://doi.org/10.1073/pnas.0811632106>.

Fang, W., Nath, K.A., Mackey, M.F., Noelle, R.J., Mueller, D.L., and Behrens, T.W. (1997). CD40 inhibits B cell apoptosis by upregulating bcl-xL expression and blocking oxidant accumulation. *Am. J. Physiol.-Cell Physiol.* *272*, C950–C956. <https://doi.org/10.1152/ajpcell.1997.272.3.C950>.

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. <https://doi.org/10.1002/j.1460-2075.1990.tb08087.x>.

FitzPatrick, D.R. (2003). Identification of SATB2 as the cleft palate gene on 2q32-q33. *Hum. Mol. Genet.* *12*, 2491–2501. <https://doi.org/10.1093/hmg/ddg248>.

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. <https://doi.org/10.1016/j.dnarep.2006.10.008>.

Fredholm, S., Willerslev-Olsen, A., Met, Ö., Kubat, L., Gluud, M., Mathiasen, S.L., Friese, C., Blümel, E., Petersen, D.L., Hu, T., et al. (2018). SATB1 in Malignant T Cells. *J. Invest. Dermatol.* *138*, 1805–1815. <https://doi.org/10.1016/j.jid.2018.03.1526>.

Freeman, G.J., Long, A.J., Iwai, Y., Bourque, K., Chernova, T., Nishimura, H., Fitz, L.J., Malenkovich, N., Okazaki, T., Byrne, M.C., et al. (2000). Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J. Exp. Med.* *192*, 1027–1034. <https://doi.org/10.1084/jem.192.7.1027>.

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J., and Schatz, D.G. (2000). The RAG Proteins and V(D)J Recombination: Complexes, Ends, and Transposition. *Annu. Rev. Immunol.* *18*, 495–527. <https://doi.org/10.1146/annurev.immunol.18.1.495>.

Fujii, Y., Kumatori, A., and Nakamura, M. (2003). SATB1 Makes a Complex with p300 and Represses gp91<sup>phox</sup> Promoter Activity. *Microbiol. Immunol.* *47*, 803–811. <https://doi.org/10.1111/j.1348-0421.2003.tb03438.x>.

Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic Hypermutation in the Heavy Chain Locus Correlates with Transcription. *Immunity* *9*, 105–114. [https://doi.org/10.1016/S1074-7613\(00\)80592-0](https://doi.org/10.1016/S1074-7613(00)80592-0).

Fukuyo, Y., Mogi, K., Tsunematsu, Y., and Nakajima, T. (2004). E2FBP1/hDril1 modulates cell growth through downregulation of promyelocytic leukemia bodies. *Cell Death Differ.* *11*, 747–759. <https://doi.org/10.1038/sj.cdd.4401412>.

Galante, S., Dickinson, L.A., Mian, I.S., Sikorska, M., and Kohwi-Shigematsu, T. (2001). SATB1 Cleavage by Caspase 6 Disrupts PDZ Domain-Mediated Dimerization, Causing Detachment from Chromatin Early in T-Cell Apoptosis. *Mol. Cell. Biol.* *21*, 5591–5604. <https://doi.org/10.1128/MCB.21.16.5591-5604.2001>.

Galante, S., Purbey, P.K., Notani, D., and Kumar, P.P. (2007). The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr. Opin. Genet. Dev.* *17*, 408–414. <https://doi.org/10.1016/j.gde.2007.08.003>.

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.* *113*, 1618–1623. <https://doi.org/10.1073/pnas.1514090113>.

Gauthier, L., Rossi, B., Roux, F., Termine, E., and Schiff, C. (2002). Galectin-1 is a stromal cell ligand of the pre-B cell receptor (BCR) implicated in synapse formation between pre-B and stromal cells and in pre-BCR triggering. *Proc. Natl. Acad. Sci.* *99*, 13014–13019. <https://doi.org/10.1073/pnas.202323999>.

Ge, X., Gao, J., Sun, Q., Wang, C., Deng, W., Mao, G., Li, H., Guo, S., Cheng, J., Wu, Y., et al. (2020). MiR-34a inhibits the proliferation, migration, and invasion of oral squamous cell carcinoma by directly targeting SATB2. *J. Cell. Physiol.* *235*, 4856–4864. <https://doi.org/10.1002/jcp.29363>.

Geier, J.K., and Schlissel, M.S. (2006). Pre-BCR signals and the control of Ig gene rearrangements. *Semin. Immunol.* *18*, 31–39. <https://doi.org/10.1016/j.smim.2005.11.001>.

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. <https://doi.org/10.1128/mcb.14.9.6153-6163.1994>.

van Gent, D.C. (1997). Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J.* *16*, 2665–2670. <https://doi.org/10.1093/emboj/16.10.2665>.

Gilfillan, S., Dierich, A., Lemeur, M., Benoist, C., and Mathis, D. (1993). Mice Lacking TdT: Mature Animals with an Immature Lymphocyte Repertoire. *Science* *261*, 1175–1178. <https://doi.org/10.1126/science.8356452>.

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. <https://doi.org/10.1084/jem.20161576>.

Giudicelli, V. (2004). IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* *33*, D256–D261. <https://doi.org/10.1093/nar/gki010>.

Gluch, A., Vidakovic, M., and Bode, J. (2008). Scaffold/Matrix Attachment Regions (S/MARs): Relevance for Disease and Therapy. In *Protein-Protein Interactions as New Drug Targets*, E. Klussmann, and J. Scott, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 67–103.

- Gong, F., Sun, L., and Sun, Y. (2010). A novel SATB1 binding site in the BCL2 promoter region possesses transcriptional regulatory function. *J. Biomed. Res.* 24, 452–459. [https://doi.org/10.1016/S1674-8301\(10\)60060-7](https://doi.org/10.1016/S1674-8301(10)60060-7).
- Gong, Y., Qian, Y., Yang, F., Wang, H., and Yu, Y. (2014). Lentiviral-mediated expression of SATB2 promotes osteogenic differentiation of bone marrow stromal cells in vitro and in vivo. *Eur. J. Oral Sci.* 122, 190–197. <https://doi.org/10.1111/eos.12122>.
- Good-Jacobson, K.L., and Shlomchik, M.J. (2010). Plasticity and Heterogeneity in the Generation of Memory B Cells and Long-Lived Plasma Cells: The Influence of Germinal Center Interactions and Dynamics. *J. Immunol.* 185, 3117–3125. <https://doi.org/10.4049/jimmunol.1001155>.
- Goodnow, C.C., Crosbie, J., Jorgensen, H., Brink, R.A., and Basten, A. (1989). Induction of self-tolerance in mature peripheral B lymphocytes. *Nature* 342, 385–391. <https://doi.org/10.1038/342385a0>.
- Gordon, M.S., Kanegai, C.M., Doerr, J.R., and Wall, R. (2003). Somatic hypermutation of the B cell receptor genes *B29* (*Ig*  $\beta$ , CD79b) and *mb1* (*Ig*  $\alpha$ , CD79a). *Proc. Natl. Acad. Sci.* 100, 4126–4131. <https://doi.org/10.1073/pnas.0735266100>.
- Gottimukkala, K.P., Jangid, R., Patta, I., Sultana, D.A., Sharma, A., Misra-Sen, J., and Galande, S. (2016). Regulation of SATB1 during thymocyte development by TCR signaling. *Mol. Immunol.* 77, 34–43. <https://doi.org/10.1016/j.molimm.2016.07.005>.
- Grawunder, U., Leu, T.M.J., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F., and Winkler, T.H. (1995). Down-regulation of RAG1 and RAG2 gene expression in PreB cells after functional immunoglobulin heavy chain rearrangement. *Immunity* 3, 601–608. [https://doi.org/10.1016/1074-7613\(95\)90131-0](https://doi.org/10.1016/1074-7613(95)90131-0).
- Grzanka, A., Grzanka, D., Gagat, M., Tadrowski, T., Sokołowska-Wojdyło, M., Marszałek, A., and Placek, W. (2012). Correlation of SATB1 expression with clinical course of cutaneous T-cell lymphomas. *Pol. J. Pathol. Off. J. Pol. Soc. Pathol.* 63, 101–105. .
- Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., Oltz, E.M., and Sen, R. (2011). Two Forms of Loops Generate the Chromatin Conformation of the Immunoglobulin Heavy-Chain Gene Locus. *Cell* 147, 332–343. <https://doi.org/10.1016/j.cell.2011.08.049>.
- Guo, H.-M., Zhou, Z.-Y., Huang, Y.-Q., Li, X., and Wang, X.-J. (2019). Investigation of the roles of dysbindin-1 and SATB2 in the progression of Parkinson’s disease. *Eur. Rev. Med. Pharmacol. Sci.* 23, 7510–7516. [https://doi.org/10.26355/eurrev\\_201909\\_18865](https://doi.org/10.26355/eurrev_201909_18865).
- 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. <https://doi.org/10.1111/j.1460-9568.2008.06061.x>.
- 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. <https://doi.org/10.3389/fimmu.2017.01387>.
- Hagman, J., Rudin, C.M., Haasch, D., Chaplin, D., and Storb, U. (1990). A novel enhancer in the immunoglobulin lambda locus is duplicated and functionally independent of NF kappa B. *Genes Dev.* 4, 978–992. <https://doi.org/10.1101/gad.4.6.978>.

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.

Han, H.-J., Russo, J., Kohwi, Y., and Kohwi-Shigematsu, T. (2008). SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* *452*, 187–193. <https://doi.org/10.1038/nature06781>.

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. <https://doi.org/10.1084/jem.20142207>.

Hao, J., Zhang, L., Cong, G., Ren, L., and Hao, L. (2016). MicroRNA-34b/c inhibits aldosterone-induced vascular smooth muscle cell calcification via a SATB2/Runx2 pathway. *Cell Tissue Res.* *366*, 733–746. <https://doi.org/10.1007/s00441-016-2469-8>.

Hardy, R.R., and Hayakawa, K. (1991). A developmental switch in B lymphopoiesis. *Proc. Natl. Acad. Sci.* *88*, 11550–11554. <https://doi.org/10.1073/pnas.88.24.11550>.

Hardy, R.R., Kincade, P.W., and Dorshkind, K. (2007). The Protean Nature of Cells in the B Lymphocyte Lineage. *Immunity* *26*, 703–714. <https://doi.org/10.1016/j.immuni.2007.05.013>.

Hassan, M.Q., Gordon, J.A.R., Beloti, M.M., Croce, C.M., Wijnen, A.J. v., Stein, J.L., Stein, G.S., and Lian, J.B. (2010). A network connecting Runx2, SATB2, and the miR-23a 27a 24-2 cluster regulates the osteoblast differentiation program. *Proc. Natl. Acad. Sci.* *107*, 19879–19884. <https://doi.org/10.1073/pnas.1007698107>.

Hawkins, S.M., Kohwi-Shigematsu, T., and Skalnik, D.G. (2001). The Matrix Attachment Region-binding Protein SATB1 Interacts with Multiple Elements within the gp91<sup>phox</sup> Promoter and Is Down-regulated during Myeloid Differentiation. *J. Biol. Chem.* *276*, 44472–44480. <https://doi.org/10.1074/jbc.M104193200>.

Hayakawa, K., Li, Y.-S., Shinton, S.A., Bandi, S.R., Formica, A.M., Brill-Dashoff, J., and Hardy, R.R. (2019). Crucial Role of Increased Arid3a at the Pre-B and Immature B Cell Stages for B1a Cell Generation. *Front. Immunol.* *10*. <https://doi.org/10.3389/fimmu.2019.00457>.

Hendrickson, B.A., Conner, D.A., Ladd, D.J., Kendall, D., Casanova, J.E., Corthesy, B., Max, E.E., Neutra, M.R., Seidman, C.E., and Seidman, J.G. (1995). Altered hepatic transport of immunoglobulin A in mice lacking the J chain. *J. Exp. Med.* *182*, 1905–1911. <https://doi.org/10.1084/jem.182.6.1905>.

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. <https://doi.org/10.1101/gad.9.24.3067>.

Hirth, G., Svensson, C.-M., Böttcher, K., Ullrich, S., Figge, M.T., and Jungnickel, B. (2019). Regulation of the Germinal Center Reaction and Somatic Hypermutation Dynamics by Homologous Recombination. *J. Immunol.* *203*, 1493–1501. <https://doi.org/10.4049/jimmunol.1900483>.

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.* *103*, 13789–13794. <https://doi.org/10.1073/pnas.0605944103>.

- Hu, C.-C.A., Dougan, S.K., McGehee, A.M., Love, J.C., and Ploegh, H.L. (2009). XBP-1 regulates signal transduction, transcription factors and bone marrow colonization in B cells. *EMBO J.* *28*, 1624–1636. <https://doi.org/10.1038/emboj.2009.117>.
- 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. <https://doi.org/10.1038/ni.2543>.
- Huang, C., Gonzalez, D.G., Cote, C.M., Jiang, Y., Hatzi, K., Teater, M., Dai, K., Hla, T., Haberman, A.M., and Melnick, A. (2014). The BCL6 RD2 Domain Governs Commitment of Activated B Cells to Form Germinal Centers. *Cell Rep.* *8*, 1497–1508. <https://doi.org/10.1016/j.celrep.2014.07.059>.
- Hulea, L., and Nepveu, A. (2012). CUX1 transcription factors: From biochemical activities and cell-based assays to mouse models and human diseases. *Gene* *497*, 18–26. <https://doi.org/10.1016/j.gene.2012.01.039>.
- Hwang\*, J.K., Alt\*, F.W., and Yeap, L.-S. (2015). Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination. *Microbiol. Spectr.* *3*. <https://doi.org/10.1128/microbiolspec.MDNA3-0037-2014>.
- Imgruet, M.K., Lutze, J., An, N., Hu, B., Khan, S., Kurkewich, J., Martinez, T.C., Wolfgeher, D., Gurbuxani, S.K., Kron, S.J., et al. (2021). Loss of a 7q gene, *CUX1*, disrupts epigenetically driven DNA repair and drives therapy-related myeloid neoplasms. *Blood* *138*, 790–805. <https://doi.org/10.1182/blood.2020009195>.
- 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. <https://doi.org/10.4049/jimmunol.177.2.1146>.
- Jankowska-Konsur, A., Kobierzycki, C., Reich, A., Grzegorzolka, J., Bieniek, A., and Dziegiel, P. (2016). Expression of SATB1, MTI/II and Ki-67 in Mycosis Fungoides. *Anticancer Res.* *36*, 189–197. .
- 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. <https://doi.org/10.1084/jem.20052227>.
- Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* *7*, 335–346. <https://doi.org/10.1038/nrm1907>.
- Johnson, K., Angelin-Duclos, C., Park, S., and Calame, K.L. (2003). Changes in Histone Acetylation Are Associated with Differences in Accessibility of V<sub>H</sub> Gene Segments to V-DJ Recombination during B-Cell Ontogeny and Development. *Mol. Cell. Biol.* *23*, 2438–2450. <https://doi.org/10.1128/MCB.23.7.2438-2450.2003>.
- 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. <https://doi.org/10.4049/jimmunol.176.7.4221>.
- Joseph, S., Deneke, V.E., and Cowden Dahl, K.D. (2012). ARID3B Induces Tumor Necrosis Factor Alpha Mediated Apoptosis While a Novel ARID3B Splice Form Does Not Induce Cell Death. *PLoS ONE* *7*, e42159. <https://doi.org/10.1371/journal.pone.0042159>.

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. <https://doi.org/10.1016/j.celrep.2017.04.038>.

Kanehiro, Y., Todo, K., Negishi, M., Fukuoka, J., Gan, W., Hikasa, T., Kaga, Y., Takemoto, M., Magari, M., Li, X., et al. (2012). Activation-induced cytidine deaminase (AID)-dependent somatic hypermutation requires a splice isoform of the serine/arginine-rich (SR) protein SRSF1. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 1216–1221. <https://doi.org/10.1073/pnas.1120368109>.

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. *J. Biol. Chem.* *276*, 21325–21330. <https://doi.org/10.1074/jbc.M100836200>.

Kassambara, A., Jourdan, M., Bruyer, A., Robert, N., Pantesco, V., Elemento, O., Klein, B., and Moreaux, J. (2017). Global miRNA expression analysis identifies novel key regulators of plasma cell differentiation and malignant plasma cell. *Nucleic Acids Res.* *45*, 5639–5652. <https://doi.org/10.1093/nar/gkx327>.

Kaul-Ghanekar, R. (2004). SMAR1 and Cux/CDP modulate chromatin and act as negative regulators of the TCR enhancer (E). *Nucleic Acids Res.* *32*, 4862–4875. <https://doi.org/10.1093/nar/gkh807>.

Kaur, S., Coulombe, Y., Ramdzan, Z.M., Leduy, L., Masson, J.-Y., and Nepveu, A. (2016). Special AT-rich Sequence-binding Protein 1 (SATB1) Functions as an Accessory Factor in Base Excision Repair. *J. Biol. Chem.* *291*, 22769–22780. <https://doi.org/10.1074/jbc.M116.735696>.

Kawaguchi, Y., Nariki, H., Kawamoto, N., Kanehiro, Y., Miyazaki, S., Suzuki, M., Magari, M., Tokumitsu, H., and Kanayama, N. (2017). SRSF1-3 contributes to diversification of the immunoglobulin variable region gene by promoting accumulation of AID in the nucleus. *Biochem. Biophys. Res. Commun.* *485*, 261–266. <https://doi.org/10.1016/j.bbrc.2017.02.097>.

Kawai, T., Sato, S., Ishii, K.J., Coban, C., Hemmi, H., Yamamoto, M., Terai, K., Matsuda, M., Inoue, J., Uematsu, S., et al. (2004). Interferon- $\alpha$  induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nat. Immunol.* *5*, 1061–1068. <https://doi.org/10.1038/ni1118>.

Khare, S.P., Shetty, A., Biradar, R., Patta, I., Chen, Z.J., Sathe, A.V., Reddy, P.C., Lahesmaa, R., and Galande, S. (2019). NF- $\kappa$ B Signaling and IL-4 Signaling Regulate SATB1 Expression via Alternative Promoter Usage During Th2 Differentiation. *Front. Immunol.* *10*. <https://doi.org/10.3389/fimmu.2019.00667>.

Kim, D., and Tucker, P.W. (2006). A Regulated Nucleocytoplasmic Shuttle Contributes to Bright's Function as a Transcriptional Activator of Immunoglobulin Genes. *Mol. Cell. Biol.* *26*, 2187–2201. <https://doi.org/10.1128/MCB.26.6.2187-2201.2006>.

Kim, D., Probst, L., Das, C., and Tucker, P.W. (2007). REKLES Is an ARID3-restricted Multifunctional Domain. *J. Biol. Chem.* *282*, 15768–15777. <https://doi.org/10.1074/jbc.M700397200>.

Kim, D., Schmidt, C., Brown, M., and Tucker, H. (2017). Competitive Promoter-Associated Matrix Attachment Region Binding of the Arid3a and Cux1 Transcription Factors. *Diseases* *5*, 34. <https://doi.org/10.3390/diseases5040034>.

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.* *22*, 1117–1123. <https://doi.org/10.1038/ejhg.2013.306>.

Kitagawa, Y., Ohkura, N., Kidani, Y., Vandenbon, A., Hirota, K., Kawakami, R., Yasuda, K., Motooka, D., Nakamura, S., Kondo, M., et al. (2017). Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat. Immunol.* *18*, 173–183. <https://doi.org/10.1038/ni.3646>.

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. <https://doi.org/10.1016/j.immuni.2011.03.025>.

Kobayashi, K., Jakt, L.M., and Nishikawa, S.-I. (2013). Epigenetic regulation of the neuroblastoma genes, *Arid3b* and *Mycn*. *Oncogene* *32*, 2640–2648. <https://doi.org/10.1038/onc.2012.285>.

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 (Mosc.)* *36*, 12005–12010. <https://doi.org/10.1021/bi971444j>.

Kondo, M., Tanaka, Y., Kuwabara, T., Naito, T., Kohwi-Shigematsu, T., and Watanabe, A. (2016). SATB1 Plays a Critical Role in Establishment of Immune Tolerance. *J. Immunol.* *196*, 563–572. <https://doi.org/10.4049/jimmunol.1501429>.

Kortschak, R.D., Reimann, H., Zimmer, M., Eyre, H.J., Saint, R., and Jenne, D.E. (1998). The Humandead ringer/brightHomolog, DRIL1:cDNA Cloning, Gene Structure, and Mapping to D19S886, a Marker on 19p13.3 That Is Strictly Linked to the Peutz–Jeghers Syndrome. *Genomics* *51*, 288–292. <https://doi.org/10.1006/geno.1998.5259>.

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. <https://doi.org/10.1002/eji.1830240407>.

Kowalczyk, A.E., Godlewski, J., Krazinski, B.E., Kiewisz, J., Sliwinska-Jewsiewicka, A., Kwiatkowski, P., Pula, B., Dziegiel, P., Janiszewski, J., Wierzbicki, P.M., et al. (2015). Divergent expression patterns of SATB1 mRNA and SATB1 protein in colorectal cancer and normal tissues. *Tumor Biol.* *36*, 4441–4452. <https://doi.org/10.1007/s13277-015-3084-0>.

Krokan, H.E., and Bjoras, M. (2013). Base Excision Repair. *Cold Spring Harb. Perspect. Biol.* *5*, a012583–a012583. <https://doi.org/10.1101/cshperspect.a012583>.

Kumar, P.P., Purbey, P.K., Ravi, D.S., Mitra, D., and Galande, S. (2005). Displacement of SATB1-Bound Histone Deacetylase 1 Corepressor by the Human Immunodeficiency Virus Type 1 Transactivator Induces Expression of Interleukin-2 and Its Receptor in T Cells. *Mol. Cell. Biol.* *25*, 1620–1633. <https://doi.org/10.1128/MCB.25.5.1620-1633.2005>.

Kumar P., P., Bischof, O., Purbey, P.K., Notani, D., Urlaub, H., Dejean, A., and Galande, S. (2007). Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat. Cell Biol.* *9*, 45–56. <https://doi.org/10.1038/ncb1516>.

Kumar Singh, A., Tamrakar, A., Jaiswal, A., Kanayama, N., Agarwal, A., Tripathi, P., and Kodgire, P. (2019). Splicing regulator SRSF1-3 that controls somatic hypermutation of IgV genes interacts with topoisomerase 1 and AID. *Mol. Immunol.* *116*, 63–72. <https://doi.org/10.1016/j.molimm.2019.10.002>.

- 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, e0161468. <https://doi.org/10.1371/journal.pone.0161468>.
- Kwon, K., Hutter, C., Sun, Q., Bilic, I., Cobaleda, C., Malin, S., and Busslinger, M. (2008). Instructive Role of the Transcription Factor E2A in Early B Lymphopoiesis and Germinal Center B Cell Development. *Immunity* 28, 751–762. <https://doi.org/10.1016/j.immuni.2008.04.014>.
- 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. <https://doi.org/10.1016/j.jmb.2016.12.021>.
- Lai, A.Y., and Kondo, M. (2008). T and B lymphocyte differentiation from hematopoietic stem cell. *Semin. Immunol.* 20, 207–212. <https://doi.org/10.1016/j.smim.2008.05.002>.
- 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. <https://doi.org/10.1371/journal.pone.0080414>.
- 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. <https://doi.org/10.1084/jem.20070902>.
- Lara-Ochoa, F., Vargas-Madrado, E., and Almagro, J.C. (1995). Distributions of the use frequencies of amino acids in the hypervariable regions of immunoglobulins. *J. Mol. Evol.* 41. <https://doi.org/10.1007/BF00174045>.
- 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. <https://doi.org/10.1128/MCB.01046-07>.
- Lederman, H.M., and Winkelstein, J.A. (1985). X-Linked Agammaglobulinemia: An Analysis of 96 Patients. *Medicine (Baltimore)* 64, 145–156. <https://doi.org/10.1097/00005792-198505000-00001>.
- 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. <https://doi.org/10.1038/sj.onc.1210872>.
- Leoyklang, P., Suphapeetiporn, K., Siriwan, P., Desudchit, T., Chaowanapanja, P., Gahl, W.A., and Shotelersuk, V. (2007). Heterozygous nonsense mutation SATB2 associated with cleft palate, osteoporosis, and cognitive defects. *Hum. Mutat.* 28, 732–738. <https://doi.org/10.1002/humu.20515>.
- Lestari, W., Ichwan, S.J.A., Otsu, M., Yamada, S., Iseki, S., Shimizu, S., and Ikeda, M.-A. (2012). Cooperation between ARID3A and p53 in the transcriptional activation of p21WAF1 in response to DNA damage. *Biochem. Biophys. Res. Commun.* 417, 710–716. <https://doi.org/10.1016/j.bbrc.2011.12.003>.
- 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 E $\mu$  in Somatic Hypermutation and Class-Switch Recombination. *J. Immunol.* 185, 6049–6057. <https://doi.org/10.4049/jimmunol.0902992>.
- Li, Q.-Q., Chen, Z.-Q., Xu, J.-D., Cao, X.-X., Chen, Q., Liu, X.-P., and Xu, Z.-D. (2010b). Overexpression and involvement of special AT-rich sequence binding protein 1 in multidrug resistance in human breast carcinoma cells. *Cancer Sci.* 101, 80–86. <https://doi.org/10.1111/j.1349-7006.2009.01372.x>.

- Li, Q.-Q., Chen, Z.-Q., Cao, X.-X., Xu, J.-D., Xu, J.-W., Chen, Y.-Y., Wang, W.-J., Chen, Q., Tang, F., Liu, X.-P., et al. (2011). Involvement of NF- $\kappa$ B/miR-448 regulatory feedback loop in chemotherapy-induced epithelial–mesenchymal transition of breast cancer cells. *Cell Death Differ.* *18*, 16–25. <https://doi.org/10.1038/cdd.2010.103>.
- Li, S., Moy, L., Pittman, N., Shue, G., Aufiero, B., Neufeld, E.J., LeLeiko, N.S., and Walsh, M.J. (1999). Transcriptional Repression of the Cystic Fibrosis Transmembrane Conductance Regulator Gene, Mediated by CCAAT Displacement Protein/cut Homolog, Is Associated with Histone Deacetylation. *J. Biol. Chem.* *274*, 7803–7815. <https://doi.org/10.1074/jbc.274.12.7803>.
- Li, Z., Otevrel, T., Gao, Y., Cheng, H.-L., Seed, B., Stamato, T.D., Taccioli, G.E., and Alt, F.W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* *83*, 1079–1089. [https://doi.org/10.1016/0092-8674\(95\)90135-3](https://doi.org/10.1016/0092-8674(95)90135-3).
- Liao, T.-T., Hsu, W.-H., Ho, C.-H., Hwang, W.-L., Lan, H.-Y., Lo, T., Chang, C.-C., Tai, S.-K., and Yang, M.-H. (2016). let-7 Modulates Chromatin Configuration and Target Gene Repression through Regulation of the ARID3B Complex. *Cell Rep.* *14*, 520–533. <https://doi.org/10.1016/j.celrep.2015.12.064>.
- Liao, T.-T., Lin, C.-C., Jiang, J.-K., Yang, S.-H., Teng, H.-W., and Yang, M.-H. (2020). Harnessing stemness and PD-L1 expression by AT-rich interaction domain-containing protein 3B in colorectal cancer. *Theranostics* *10*, 6095–6112. <https://doi.org/10.7150/thno.44147>.
- 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. <https://doi.org/10.1146/annurev.biochem.052308.093131>.
- Lieberman, R., Giannini, S.L., Birshtein, B.K., and Eckhardt, L.A. (1991). An enhancer at the 3' end of the mouse immunoglobulin heavy chain locus. *Nucleic Acids Res.* *19*, 933–937. <https://doi.org/10.1093/nar/19.4.933>.
- 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. <https://doi.org/10.1016/j.cell.2017.03.043>.
- 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*. <https://doi.org/10.1186/1476-4598-6-23>.
- 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. <https://doi.org/10.1038/nature06547>.
- Liu, N., Sun, Q., Wan, L., Wang, X., Feng, Y., Luo, J., and Wu, H. (2020). CUX1, A Controversial Player in Tumor Development. *Front. Oncol.* *10*. <https://doi.org/10.3389/fonc.2020.00738>.
- 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. <https://doi.org/10.1016/j.immuni.2007.09.005>.
- Liu, Z.-M., George-Raizen, J.B., Li, S., Meyers, K.C., Chang, M.Y., and Garrard, W.T. (2002). Chromatin Structural Analyses of the Mouse Igk Gene Locus Reveal New Hypersensitive Sites Specifying a Transcriptional Silencer and Enhancer. *J. Biol. Chem.* *277*, 32640–32649. <https://doi.org/10.1074/jbc.M204065200>.

- 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. <https://doi.org/10.1016/j.jaut.2011.12.005>.
- Livingston, S., Carlton, C., Sharma, M., Kearns, D., Baybutt, R., and Vanden Heuvel, G.B. (2019). Cux1 regulation of the cyclin kinase inhibitor p27kip1 in polycystic kidney disease is attenuated by HDAC inhibitors. *Gene* *721*, 100007. <https://doi.org/10.1016/j.gene.2019.100007>.
- Löffert, D., Ehlich, A., Müller, W., and Rajewsky, K. (1996). Surrogate Light Chain Expression Is Required to Establish Immunoglobulin Heavy Chain Allelic Exclusion during Early B Cell Development. *Immunity* *4*, 133–144. [https://doi.org/10.1016/S1074-7613\(00\)80678-0](https://doi.org/10.1016/S1074-7613(00)80678-0).
- Low, M.S.Y., Brodie, E.J., Fedele, P.L., Liao, Y., Grigoriadis, G., Strasser, A., Kallies, A., Willis, S.N., Tellier, J., Shi, W., et al. (2019). IRF4 Activity Is Required in Established Plasma Cells to Regulate Gene Transcription and Mitochondrial Homeostasis. *Cell Rep.* *29*, 2634–2645.e5. <https://doi.org/10.1016/j.celrep.2019.10.097>.
- Ludérus, M.E.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. [https://doi.org/10.1016/0092-8674\(92\)90245-8](https://doi.org/10.1016/0092-8674(92)90245-8).
- Lund, J., Sato, A., Akira, S., Medzhitov, R., and Iwasaki, A. (2003). Toll-like Receptor 9-mediated Recognition of Herpes Simplex Virus-2 by Plasmacytoid Dendritic Cells. *J. Exp. Med.* *198*, 513–520. <https://doi.org/10.1084/jem.20030162>.
- Ma, C., Olevian, D., Miller, C., Herbst, C., Jayachandran, P., Kozak, M.M., Chang, D.T., and Pai, R.K. (2019). SATB2 and CDX2 are prognostic biomarkers in DNA mismatch repair protein deficient colon cancer. *Mod. Pathol.* *32*, 1217–1231. <https://doi.org/10.1038/s41379-019-0265-1>.
- Ma, K., Araki, K., Ichwan, S.J.A., Suganuma, T., Tamamori-Adachi, M., and Ikeda, M.-A. (2003). E2FBP1/DRIL1, an AT-rich interaction domain-family transcription factor, is regulated by p53. *Mol. Cancer Res. MCR* *1*, 438–444. .
- Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M.R. (2002). Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination. *Cell* *108*, 781–794. [https://doi.org/10.1016/S0092-8674\(02\)00671-2](https://doi.org/10.1016/S0092-8674(02)00671-2).
- 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. <https://doi.org/10.1101/gad.8.18.2212>.
- Maier, H., and Hagman, J. (2002). Roles of EBF and Pax-5 in B lineage commitment and development. *Semin. Immunol.* *14*, 415–422. <https://doi.org/10.1016/S1044532302000763>.
- Mailly, F., Bérubé, G., Harada, R., Mao, P.L., Phillips, S., and Nepveu, A. (1996). The human cut homeodomain protein can repress gene expression by two distinct mechanisms: active repression and competition for binding site occupancy. *Mol. Cell. Biol.* *16*, 5346–5357. <https://doi.org/10.1128/MCB.16.10.5346>.
- Malu, S., De Ioannes, P., Kozlov, M., Greene, M., Francis, D., Hanna, M., Pena, J., Escalante, C.R., Kurosawa, A., Erdjument-Bromage, H., et al. (2012). Artemis C-terminal region facilitates V(D)J recombination through its interactions with DNA Ligase IV and DNA-PKcs. *J. Exp. Med.* *209*, 955–963. <https://doi.org/10.1084/jem.20111437>.

- 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. <https://doi.org/10.1038/ni1067>.
- Mansour, M.A., Hyodo, T., Akter, K.A., Kokuryo, T., Uehara, K., Nagino, M., and Senga, T. (2015). SATB1 and SATB2 play opposing roles in c-Myc expression and progression of colorectal cancer. *Oncotarget* *7*, 4993–5006. <https://doi.org/10.18632/oncotarget.6651>.
- Mao, L., Yang, C., Wang, J., Li, W., Wen, R., Chen, J., and Zheng, J. (2013). SATB1 is overexpressed in metastatic prostate cancer and promotes prostate cancer cell growth and invasion. *J. Transl. Med.* *11*, 111. <https://doi.org/10.1186/1479-5876-11-111>.
- Marchalot, A., Ashi, M.O., Lambert, J.-M., Carrion, C., Lecardeur, S., Srouf, N., Delpy, L., and Le Pennec, S. (2020). Uncoupling Splicing From Transcription Using Antisense Oligonucleotides Reveals a Dual Role for I Exon Donor Splice Sites in Antibody Class Switching. *Front. Immunol.* *11*. <https://doi.org/10.3389/fimmu.2020.00780>.
- Mårtensson, I.-L., and Ceredig, R. (2000). Role of the surrogate light chain and the pre-B-cell receptor in mouse B-cell development: Mouse B-cell development. *Immunology* *101*, 435–441. <https://doi.org/10.1046/j.1365-2567.2000.00151.x>.
- 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. [https://doi.org/10.1016/S1044-5323\(02\)00066-0](https://doi.org/10.1016/S1044-5323(02)00066-0).
- 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. [https://doi.org/10.1016/S1074-7613\(01\)00129-7](https://doi.org/10.1016/S1074-7613(01)00129-7).
- Martin, O., Thomas, M., Marquet, M., Garot, A., Brousse, M., Bender, S., Carrion, C., Choi, J.E., Vuong, B.Q., Gearhart, P.J., et al. (2022). The IgH  $\mu$ -MAR regions promote UNG-dependent error-prone repair to optimize somatic hypermutation. *bioRxiv* 2022.08.15.503996. <https://doi.org/10.1101/2022.08.15.503996>.
- 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. <https://doi.org/10.4049/jimmunol.1800203>.
- 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. <https://doi.org/10.1038/nature06431>.
- 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. <https://doi.org/10.1084/jem.20131512>.
- 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. <https://doi.org/10.1016/j.dnarep.2015.03.004>.

- McMurry, M.T., and Krangel, M.S. (2000). A Role for Histone Acetylation in the Developmental Regulation of V(D)J Recombination. *Science* 287, 495–498. <https://doi.org/10.1126/science.287.5452.495>.
- Mebius, R.E., and Kraal, G. (2005). Structure and function of the spleen. *Nat. Rev. Immunol.* 5, 606–616. <https://doi.org/10.1038/nri1669>.
- Melamed, D., Benschop, R.J., Cambier, J.C., and Nemazee, D. (1998). Developmental Regulation of B Lymphocyte Immune Tolerance Compartmentalizes Clonal Selection from Receptor Selection. *Cell* 92, 173–182. [https://doi.org/10.1016/S0092-8674\(00\)80912-5](https://doi.org/10.1016/S0092-8674(00)80912-5).
- 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. [https://doi.org/10.1016/0167-5699\(93\)90060-X](https://doi.org/10.1016/0167-5699(93)90060-X).
- Meng, W.-J., Yan, H., Zhou, B., Zhang, W., Kong, X.-H., Wang, R., Zhan, L., Li, Y., Zhou, Z.-G., and Sun, X.-F. (2012). Correlation of SATB1 overexpression with the progression of human rectal cancer. *Int. J. Colorectal Dis.* 27, 143–150. <https://doi.org/10.1007/s00384-011-1302-9>.
- Merrell, K.T., Benschop, R.J., Gauld, S.B., Aviszus, K., Decote-Ricardo, D., Wysocki, L.J., and Cambier, J.C. (2006). Identification of Anergic B Cells within a Wild-Type Repertoire. *Immunity* 25, 953–962. <https://doi.org/10.1016/j.immuni.2006.10.017>.
- Mikkola, I., Heavey, B., Horcher, M., and Busslinger, M. (2002). Reversion of B Cell Commitment upon Loss of *Pax5* Expression. *Science* 297, 110–113. <https://doi.org/10.1126/science.1067518>.
- Mitchison, N.A. (2004). T-cell–B-cell cooperation. *Nat. Rev. Immunol.* 4, 308–312. <https://doi.org/10.1038/nri1334>.
- Monroe, J.G., and Dorshkind, K. (2007). Fate Decisions Regulating Bone Marrow and Peripheral B Lymphocyte Development. In *Advances in Immunology*, (Elsevier), pp. 1–50.
- Montecino-Rodriguez, E., and Dorshkind, K. (2006). New perspectives in B-1 B cell development and function. *Trends Immunol.* 27, 428–433. <https://doi.org/10.1016/j.it.2006.07.005>.
- 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.* 100, 11577–11582. <https://doi.org/10.1073/pnas.1932643100>.
- 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. <https://doi.org/10.1084/jem.192.12.1833>.
- Nakagomi, K., Kohwi, Y., Dickinson, L.A., and Kohwi-Shigematsu, T. (1994). A novel DNA-binding motif in the nuclear matrix attachment DNA-binding protein SATB1. *Mol. Cell. Biol.* 14, 1852–1860. <https://doi.org/10.1128/MCB.14.3.1852>.
- Nakayama, Y., Mian, I.S., Kohwi-Shigematsu, T., and Ogawa, T. (2005). A nuclear targeting determinant for SATB1, a genome organizer in the T cell lineage. *Cell Cycle Georget. Tex* 4, 1099–1106. .
- Nakayasu, H., and Berezney, R. (1991). Nuclear matrices: identification of the major nuclear matrix proteins. *Proc. Natl. Acad. Sci. U. S. A.* 88, 10312–10316. <https://doi.org/10.1073/pnas.88.22.10312>.

- Nepveu, A. (2001). Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* 270, 1–15. [https://doi.org/10.1016/S0378-1119\(01\)00485-1](https://doi.org/10.1016/S0378-1119(01)00485-1).
- Nera, K.-P., and Lassila, O. (2006). Pax5 – a Critical Inhibitor of Plasma Cell Fate. *Scand. J. Immunol.* 64, 190–199. <https://doi.org/10.1111/j.1365-3083.2006.01809.x>.
- Neuberger, M.S., Noia, J.M.D., 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. <https://doi.org/10.1038/nri1553>.
- Nie, H., Yao, X., Maika, S.D., and Tucker, P.W. (2008). SATB1 is required for CD8 coreceptor reversal. *Mol. Immunol.* 46, 207–211. <https://doi.org/10.1016/j.molimm.2008.07.007>.
- Niirö, H., and Clark, E.A. (2002). Regulation of B-cell fate by antigen-receptor signals. *Nat. Rev. Immunol.* 2, 945–956. <https://doi.org/10.1038/nri955>.
- Niles, M.J., Matsuuchi, L., and Koshland, M.E. (1995). Polymer IgM assembly and secretion in lymphoid and nonlymphoid cell lines: evidence that J chain is required for pentamer IgM synthesis. *Proc. Natl. Acad. Sci.* 92, 2884–2888. <https://doi.org/10.1073/pnas.92.7.2884>.
- Nixon, J.C., Rajaiya, J., and Webb, C.F. (2004a). 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. <https://doi.org/10.1074/jbc.M403028200>.
- Nixon, J.C., Rajaiya, J.B., Ayers, N., Evetts, S., and Webb, C.F. (2004b). The transcription factor, Bright, is not expressed in all human B lymphocyte subpopulations. *Cell. Immunol.* 228, 42–53. <https://doi.org/10.1016/j.cellimm.2004.03.004>.
- 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.* 181, 6913–6922. <https://doi.org/10.4049/jimmunol.181.10.6913>.
- Nodin, B., Hedner, C., Uhlén, M., and Jirström, K. (2012a). Expression of the global regulator SATB1 is an independent factor of poor prognosis in high grade epithelial ovarian cancer. *J. Ovarian Res.* 5, 24. <https://doi.org/10.1186/1757-2215-5-24>.
- Nodin, B., Johannesson, H., Wangefjord, S., O'Connor, D.P., Lindquist, K.E., Uhlén, M., Jirström, K., and Eberhard, J. (2012b). Molecular correlates and prognostic significance of SATB1 expression in colorectal cancer. *Diagn. Pathol.* 7, 115. <https://doi.org/10.1186/1746-1596-7-115>.
- 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. .
- Nussenzweig, A., Chen, C., da Costa Soares, V., Sanchez, M., Sokol, K., Nussenzweig, M.C., and Li, G.C. (1996). Requirement for Ku80 in growth and immunoglobulin V(D)J recombination. *Nature* 382, 551–555. <https://doi.org/10.1038/382551a0>.
- Nutt, S.L., Urbánek, P., Rolink, A., and Busslinger, M. (1997). Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus. *Genes Dev.* 11, 476–491. <https://doi.org/10.1101/gad.11.4.476>.

- 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. <https://doi.org/10.1038/nri3795>.
- Oettinger, M.A., Schatz, D.G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V(D)J Recombination. *Science* *248*, 1517–1523. <https://doi.org/10.1126/science.2360047>.
- Oguz Erdogan, A.S., Ozdemirler, N., Oyken, M., Alper, M., and Erson-Bensan, A.E. (2014). ARID3B expression in primary breast cancers and breast cancer-derived cell lines. *Cell. Oncol.* *37*, 289–296. <https://doi.org/10.1007/s13402-014-0185-5>.
- Okada, T., and Cyster, J.G. (2006). B cell migration and interactions in the early phase of antibody responses. *Curr. Opin. Immunol.* *18*, 278–285. <https://doi.org/10.1016/j.coi.2006.02.005>.
- 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. <https://doi.org/10.1016/j.molimm.2011.09.008>.
- Oracki, S.A., Walker, J.A., Hibbs, M.L., Corcoran, L.M., and Tarlinton, D.M. (2010). Plasma cell development and survival: Plasma cell development and survival. *Immunol. Rev.* *237*, 140–159. <https://doi.org/10.1111/j.1600-065X.2010.00940.x>.
- O’Riordan, M., and Grosschedl, R. (1999). Coordinate Regulation of B Cell Differentiation by the Transcription Factors EBF and E2A. *Immunity* *11*, 21–31. [https://doi.org/10.1016/S1074-7613\(00\)80078-3](https://doi.org/10.1016/S1074-7613(00)80078-3).
- Ouk, C., Roland, L., Gachard, N., Poulain, S., Oblet, C., Rizzo, D., Saintamand, A., Lemasson, Q., Carrion, C., Thomas, M., et al. (2021). Continuous MYD88 Activation Is Associated With Expansion and Then Transformation of IgM Differentiating Plasma Cells. *Front. Immunol.* *12*. <https://doi.org/10.3389/fimmu.2021.641692>.
- Padlan, E.A. (1994). Anatomy of the antibody molecule. *Mol. Immunol.* *31*, 169–217. [https://doi.org/10.1016/0161-5890\(94\)90001-9](https://doi.org/10.1016/0161-5890(94)90001-9).
- Park, B.S., Song, D.H., Kim, H.M., Choi, B.-S., Lee, H., and Lee, J.-O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* *458*, 1191–1195. <https://doi.org/10.1038/nature07830>.
- 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.* *81*, 2650–2654. <https://doi.org/10.1073/pnas.81.9.2650>.
- Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R.S.K., 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.* *95*, 11816–11821. <https://doi.org/10.1073/pnas.95.20.11816>.
- Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S.K., Küppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* *412*, 341–346. <https://doi.org/10.1038/35085588>.
- Patta, I., Madhok, A., Khare, S., Gottimukkala, K.P., Verma, A., Giri, S., Dandewad, V., Seshadri, V., Lal, G., Misra-Sen, J., et al. (2020). Dynamic regulation of chromatin organizer SATB1 via TCR-induced



alternative promoter switch during T-cell development. *Nucleic Acids Res.* *48*, 5873–5890. <https://doi.org/10.1093/nar/gkaa321>.

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. <https://doi.org/10.1016/j.molcel.2006.03.010>.

Pavri, R., and Nussenzweig, M.C. (2011). AID Targeting in Antibody Diversity. In *Advances in Immunology*, (Elsevier), pp. 1–26.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., San-Martin, B.R., 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. <https://doi.org/10.1016/j.cell.2010.09.017>.

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. <https://doi.org/10.4049/jimmunol.176.11.6839>.

Peeper, D.S., Shvarts, A., Brummelkamp, T., Douma, S., Koh, E.Y., Daley, G.Q., and Bernards, R. (2002). A functional screen identifies hDRIL1 as an oncogene that rescues RAS-induced senescence. *Nat. Cell Biol.* *4*, 148–153. <https://doi.org/10.1038/ncb742>.

Pefanis, E., and Basu, U. (2015). RNA Exosome Regulates AID DNA Mutator Activity in the B Cell Genome. In *Advances in Immunology*, (Elsevier), pp. 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. <https://doi.org/10.1038/nature13580>.

Perlot, T., and Alt, F.W. (2008). Chapter 1 Cis-Regulatory Elements and Epigenetic Changes Control Genomic Rearrangements of the IgH Locus. In *Advances in Immunology*, (Elsevier), pp. 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.* *102*, 14362–14367. <https://doi.org/10.1073/pnas.0507090102>.

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.* *105*, 3843–3848. <https://doi.org/10.1073/pnas.0712291105>.

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. <https://doi.org/10.1126/science.1218692>.

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. <https://doi.org/10.1074/jbc.M200305200>.

Phan, R.T., and Dalla-Favera, R. (2004). The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* *432*, 635–639. <https://doi.org/10.1038/nature03147>.

- Pillai, S., and Cariappa, A. (2009). The follicular versus marginal zone B lymphocyte cell fate decision. *Nat. Rev. Immunol.* 9, 767–777. <https://doi.org/10.1038/nri2656>.
- Pillai, S., Cariappa, A., and Moran, S.T. (2005). MARGINAL ZONE B CELLS. *Annu. Rev. Immunol.* 23, 161–196. <https://doi.org/10.1146/annurev.immunol.23.021704.115728>.
- 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. <https://doi.org/10.1038/nrm.2016.111>.
- Popowski, M., Templeton, T.D., Lee, B.-K., Rhee, C., Li, H., Miner, C., Dekker, J.D., Orlanski, S., Bergman, Y., Iyer, V.R., et al. (2014). Bright/Arid3A Acts as a Barrier to Somatic Cell Reprogramming through Direct Regulation of Oct4, Sox2, and Nanog. *Stem Cell Rep.* 2, 26–35. <https://doi.org/10.1016/j.stemcr.2013.12.002>.
- Porter, R.R. (1973). Structural Studies of Immunoglobulins. *Science* 180, 713–716. <https://doi.org/10.1126/science.180.4087.713>.
- Pratama, E., Tian, X., Lestari, W., Iseki, S., Ichwan, S.J.A., and Ikeda, M.-A. (2015). Critical role of ARID3B in the expression of pro-apoptotic p53-target genes and apoptosis. *Biochem. Biophys. Res. Commun.* 468, 248–254. <https://doi.org/10.1016/j.bbrc.2015.10.121>.
- Predeus, A.V., Gopalakrishnan, S., Huang, Y., Tang, J., Feeney, A.J., Oltz, E.M., and Artyomov, M.N. (2014). Targeted Chromatin Profiling Reveals Novel Enhancers in Ig H and Ig L Chain Loci. *J. Immunol.* 192, 1064–1070. <https://doi.org/10.4049/jimmunol.1302800>.
- Prieto, J.M.B., and Felipe, M.J.B. (2017). Development, phenotype, and function of non-conventional B cells. *Comp. Immunol. Microbiol. Infect. Dis.* 54, 38–44. <https://doi.org/10.1016/j.cimid.2017.08.002>.
- Pucella, J.N., and Chaudhuri, J. (2017). AID Invited to the G4 Summit. *Mol. Cell* 67, 355–357. <https://doi.org/10.1016/j.molcel.2017.07.020>.
- Puissegur, M.P., Eichner, R., Quelen, C., Coyaud, E., Mari, B., Lebrigand, K., Broccardo, C., Nguyen-Khac, F., Bousquet, M., and Brousset, P. (2012). B-cell regulator of immunoglobulin heavy-chain transcription (Bright)/ARID3a is a direct target of the oncomir microRNA-125b in progenitor B-cells. *Leukemia* 26, 2224–2232. <https://doi.org/10.1038/leu.2012.95>.
- Purbey, P.K., Singh, S., Kumar, P.P., Mehta, S., Ganesh, K.N., Mitra, D., and Galande, S. (2008). PDZ domain-mediated dimerization and homeodomain-directed specificity are required for high-affinity DNA binding by SATB1. *Nucleic Acids Res.* 36, 2107–2122. <https://doi.org/10.1093/nar/gkm1151>.
- Purbey, P.K., Singh, S., Notani, D., Kumar, P.P., Limaye, A.S., and Galande, S. (2009). Acetylation-Dependent Interaction of SATB1 and CtBP1 Mediates Transcriptional Repression by SATB1. *Mol. Cell Biol.* 29, 1321–1337. <https://doi.org/10.1128/MCB.00822-08>.
- 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. <https://doi.org/10.1016/j.molcel.2017.06.034>.
- 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. [https://doi.org/10.1016/S1074-7613\(00\)80595-6](https://doi.org/10.1016/S1074-7613(00)80595-6).

- 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. <https://doi.org/10.1016/j.molcel.2004.10.011>.
- Radic, M.Z., and Zouali, M. (1996). Receptor Editing, Immune Diversification, and Self-Tolerance. *Immunity* 5, 505–511. [https://doi.org/10.1016/S1074-7613\(00\)80266-6](https://doi.org/10.1016/S1074-7613(00)80266-6).
- Rajaiya, J., Nixon, J.C., Ayers, N., Desgranges, Z.P., Roy, A.L., and Webb, C.F. (2006). Induction of Immunoglobulin Heavy-Chain Transcription through the Transcription Factor Bright Requires TFII-I. *Mol. Cell. Biol.* 26, 4758–4768. <https://doi.org/10.1128/MCB.02009-05>.
- Ramakrishnan, M., Liu, W.-M., DiCroce, P.A., Posner, A., Zheng, J., Kohwi-Shigematsu, T., and Krontiris, T.G. (2000). Modulated Binding of SATB1, a Matrix Attachment Region Protein, to the AT-Rich Sequence Flanking the Major Breakpoint Region of BCL2. *Mol. Cell. Biol.* 20, 868–877. .
- Ramdzan, Z.M., Pal, R., Kaur, S., Leduy, L., Bérubé, G., Davoudi, S., Vadnais, C., and Nepveu, A. (2015). The function of *cux1* in oxidative dna damage repair is needed to prevent premature senescence of mouse embryo fibroblasts. *Oncotarget* 6, 3613–3626. <https://doi.org/10.18632/oncotarget.2919>.
- 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. <https://doi.org/10.1038/ni1478>.
- Ratliff, M.L., Mishra, M., Frank, M.B., Guthridge, J.M., and Webb, C.F. (2016). The Transcription Factor ARID3a Is Important for In Vitro Differentiation of Human Hematopoietic Progenitors. *J. Immunol.* 196, 614–623. <https://doi.org/10.4049/jimmunol.1500355>.
- Ratliff, M.L., Garton, J., Garman, L., Barron, M.D., Georgescu, C., White, K.A., Chakravarty, E., Wren, J.D., Montgomery, C.G., James, J.A., et al. (2019). ARID3a gene profiles are strongly associated with human interferon alpha production. *J. Autoimmun.* 96, 158–167. <https://doi.org/10.1016/j.jaut.2018.09.013>.
- Ratliff, M.L., Shankar, M., Guthridge, J.M., James, J.A., and Webb, C.F. (2020). TLR engagement induces ARID3a in human blood hematopoietic progenitors and modulates IFN $\alpha$  production. *Cell. Immunol.* 357, 104201. <https://doi.org/10.1016/j.cellimm.2020.104201>.
- Rebhandl, S., Huemer, M., Greil, R., and Geisberger, R. (2015). AID/APOBEC deaminases and cancer. *Oncoscience* 2, 320–333. <https://doi.org/10.18632/oncoscience.155>.
- 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. [https://doi.org/10.1016/S0092-8674\(00\)00079-9](https://doi.org/10.1016/S0092-8674(00)00079-9).
- 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. (2008). Ubiquitylated PCNA plays a role in somatic hypermutation and class-switch recombination and is required for meiotic progression. *Proc. Natl. Acad. Sci.* 105, 16248–16253. <https://doi.org/10.1073/pnas.0808182105>.
- 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.* 66, 981–993. <https://doi.org/10.1007/s00018-009-8736-z>.

Rogier, M., Moritz, J., Robert, I., Lescale, C., Heyer, V., Abello, A., Martin, O., Capitani, K., Thomas, M., Thomas-Claudepierre, A.-S., et al. (2021). Fam72a enforces error-prone DNA repair during antibody diversification. *Nature* 600, 329–333. <https://doi.org/10.1038/s41586-021-04093-y>.

Rolink, A.G., Tschopp, J., Schneider, P., and Melchers, F. (2002). BAFF is a survival and maturation factor for mouse B cells. *Eur. J. Immunol.* 32, 2004. [https://doi.org/10.1002/1521-4141\(200207\)32:7<2004::AID-IMMU2004>3.0.CO;2-5](https://doi.org/10.1002/1521-4141(200207)32:7<2004::AID-IMMU2004>3.0.CO;2-5).

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. <https://doi.org/10.15252/embj.201693958>.

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. <https://doi.org/10.1073/pnas.0505449102>.

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. <https://doi.org/10.1084/jem.20130072>.

Roy, A.L., Sen, R., and Roeder, R.G. (2011). Enhancer-promoter communication and transcriptional regulation of Igh. *Trends Immunol.* 32, 532–539. <https://doi.org/10.1016/j.it.2011.06.012>.

Roy, L., Samyesudhas, S.J., Carrasco, M., Li, J., Joseph, S., Dahl, R., and Dahl, K.D.C. (2014). ARID3B increases ovarian tumor burden and is associated with a cancer stem cell gene signature. *Oncotarget* 5, 8355–8366. <https://doi.org/10.18632/oncotarget.2247>.

Saadat, K., Lestari, W., Pratama, E., Ma, T., Iseki, S., Tatsumi, M., and Ikeda, M.-A. (2021). Distinct and overlapping roles of ARID3A and ARID3B in regulating E2F-dependent transcription via direct binding to E2F target genes. *Int. J. Oncol.* 58. <https://doi.org/10.3892/ijo.2021.5192>.

Saadat, K.A.S.M., Bozgeyik, E., Arman, K., Bozgeyik, I., and Ikeda, M.-A. (2020). ARID3A-mediated modulation of TP73 and TP73-AS1 in osteosarcoma cells. *Gene Rep.* 20, 100683. <https://doi.org/10.1016/j.genrep.2020.100683>.

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.* 106, 11294–11299. <https://doi.org/10.1073/pnas.0903854106>.

Sakai, E., Bottaro, A., Davidson, L., Sleckman, B.P., and Alt, F.W. (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.* 96, 1526–1531. <https://doi.org/10.1073/pnas.96.4.1526>.

Santaguida, M., Ding, Q., Bérubé, G., Truscott, M., Whyte, P., and Nepveu, A. (2001). Phosphorylation of the CCAAT Displacement Protein (CDP)/Cux Transcription Factor by Cyclin A-Cdk1 Modulates Its DNA Binding Activity in G2. *J. Biol. Chem.* 276, 45780–45790. <https://doi.org/10.1074/jbc.M107978200>.

Saribasak, H., and Gearhart, P.J. (2012). Does DNA repair occur during somatic hypermutation? *Semin. Immunol.* 24, 287–292. <https://doi.org/10.1016/j.smim.2012.05.002>.

- Saribasak, H., Maul, R.W., Cao, Z., McClure, R.L., Yang, W., McNeill, D.R., Wilson, D.M., and Gearhart, P.J. (2011). XRCC1 suppresses somatic hypermutation and promotes alternative nonhomologous end joining in Igh genes. *J. Exp. Med.* *208*, 2209–2216. <https://doi.org/10.1084/jem.20111135>.
- 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. <https://doi.org/10.1016/j.immuni.2013.05.014>.
- 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. <https://doi.org/10.1101/gad.1815709>.
- Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* *11*, 251–263. <https://doi.org/10.1038/nri2941>.
- Schmidt, C., Kim, D., Ippolito, G.C., Naqvi, H.R., Probst, L., Mathur, S., Rosas-Acosta, G., Wilson, V.G., Oldham, A.L., Poenie, M., et al. (2009). Signalling of the BCR is regulated by a lipid rafts-localised transcription factor, Bright. *EMBO J.* *28*, 711–724. <https://doi.org/10.1038/emboj.2009.20>.
- Schroeder, H.W., and Cavacini, L. (2010). Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* *125*, S41–S52. <https://doi.org/10.1016/j.jaci.2009.09.046>.
- Sciammas, R., Shaffer, A.L., Schatz, J.H., Zhao, H., Staudt, L.M., and Singh, H. (2006). Graded Expression of Interferon Regulatory Factor-4 Coordinates Isotype Switching with Plasma Cell Differentiation. *Immunity* *25*, 225–236. <https://doi.org/10.1016/j.immuni.2006.07.009>.
- Selinger, C.I., Cooper, W.A., Al-Sohaily, S., N.Mladenova, D., Pangon, L., W. Kennedy, C., C. McCaughan, B., Stirzaker, C., and R.J. Kohonen-Corish, M. (2011). Loss of Special AT-Rich Binding Protein 1 Expression is a Marker of Poor Survival in Lung Cancer. *J. Thorac. Oncol.* *6*, 1179–1189. <https://doi.org/10.1097/JTO.0b013e31821b4ce0>.
- Sen, R., and Oltz, E. (2006). Genetic and epigenetic regulation of IgH gene assembly. *Curr. Opin. Immunol.* *18*, 237–242. <https://doi.org/10.1016/j.coi.2006.03.008>.
- Seo, J., Lozano, M.M., and Dudley, J.P. (2005). Nuclear Matrix Binding Regulates SATB1-mediated Transcriptional Repression. *J. Biol. Chem.* *280*, 24600–24609. <https://doi.org/10.1074/jbc.M414076200>.
- 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*, 2321–2327. <https://doi.org/10.1002/j.1460-2075.1993.tb05886.x>.
- 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 Overexpress the Transcription Factor Bright. *J. Immunol.* *178*, 2996–3006. <https://doi.org/10.4049/jimmunol.178.5.2996>.
- Sheppard, E.C., Morrish, R.B., Dillon, M.J., Leyland, R., and Chahwan, R. (2018). Epigenomic Modifications Mediating Antibody Maturation. *Front. Immunol.* *9*. <https://doi.org/10.3389/fimmu.2018.00355>.
- Shinkai, Y. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* *68*, 855–867. [https://doi.org/10.1016/0092-8674\(92\)90029-C](https://doi.org/10.1016/0092-8674(92)90029-C).

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. <https://doi.org/10.1182/blood.V98.13.3658>.

Singh, A.K., Tamrakar, A., Jaiswal, A., Kanayama, N., and Kodgire, P. (2020). SRSF1-3, a splicing and somatic hypermutation regulator, controls transcription of IgV genes via chromatin regulators SATB2, UBN1 and histone variant H3.3. *Mol. Immunol.* 119, 69–82. <https://doi.org/10.1016/j.molimm.2020.01.005>.

Sitnicka, E., Bryder, D., Theilgaard-Mönch, K., Buza-Vidas, N., Adolfsson, J., and Jacobsen, S.E.W. (2002). Key Role of flt3 Ligand in Regulation of the Common Lymphoid Progenitor but Not in Maintenance of the Hematopoietic Stem Cell Pool. *Immunity* 17, 463–472. [https://doi.org/10.1016/S1074-7613\(02\)00419-3](https://doi.org/10.1016/S1074-7613(02)00419-3).

Smith, K.G.C., Hewitson, T.D., Nossal, G.J.V., and Tarlinton, D.M. (1996). The phenotype and fate of the antibody-forming cells of the splenic foci. *Eur. J. Immunol.* 26, 444–448. <https://doi.org/10.1002/eji.1830260226>.

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. <https://doi.org/10.1084/jem.20031498>.

Song, G., Liu, K., Yang, X., Mu, B., Yang, J., He, L., Hu, X., Li, Q., Zhao, Y., Cai, X., et al. (2017). SATB1 plays an oncogenic role in esophageal cancer by up-regulation of FN1 and PDGFRB. *Oncotarget* 8, 17771–17784. <https://doi.org/10.18632/oncotarget.14849>.

Stadanlick, J.E., and Cancro, M.P. (2008). BAFF and the plasticity of peripheral B cell tolerance. *Curr. Opin. Immunol.* 20, 158–161. <https://doi.org/10.1016/j.coi.2008.03.015>.

Stavnezer, J., and Schrader, C.E. (2014). IgH Chain Class Switch Recombination: Mechanism and Regulation. *J. Immunol.* 193, 5370–5378. <https://doi.org/10.4049/jimmunol.1401849>.

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.* 111, 9217–9222. <https://doi.org/10.1073/pnas.1405590111>.

Stephen, T.L., Payne, K.K., Chaurio, R.A., Allegranza, M.J., Zhu, H., Perez-Sanz, J., Perales-Puchalt, A., Nguyen, J.M., Vara-Ailor, A.E., Eruslanov, E.B., et al. (2017). SATB1 Expression Governs Epigenetic Repression of PD-1 in Tumor-Reactive T Cells. *Immunity* 46, 51–64. <https://doi.org/10.1016/j.immuni.2016.12.015>.

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. <https://doi.org/10.1128/mcb.9.2.711-718.1989>.

Sun, J., Yi, S., Qiu, L., Fu, W., Wang, A., Liu, F., Wang, L., Wang, T., Chen, H., Wang, L., et al. (2018). SATB1 Defines a Subtype of Cutaneous CD30+ Lymphoproliferative Disorders Associated with a T-Helper 17 Cytokine Profile. *J. Invest. Dermatol.* 138, 1795–1804. <https://doi.org/10.1016/j.jid.2018.02.028>.

Szemes, M., Gyorgy, A., Paweletz, C., Dobi, A., and Agoston, D.V. (2006). Isolation and Characterization of SATB2, a Novel AT-rich DNA Binding Protein Expressed in Development- and Cell-Specific Manner in the Rat Brain. *Neurochem. Res.* *31*, 237–246. <https://doi.org/10.1007/s11064-005-9012-8>.

Takebe, A., Era, T., Okada, M., Martin Jakt, L., Kuroda, Y., and Nishikawa, S.-I. (2006). Microarray analysis of PDGFR $\alpha$ + populations in ES cell differentiation culture identifies genes involved in differentiation of mesoderm and mesenchyme including ARID3b that is essential for development of embryonic mesenchymal cells. *Dev. Biol.* *293*, 25–37. <https://doi.org/10.1016/j.ydbio.2005.12.016>.

Takeda, K., and Akira, S. (2004). TLR signaling pathways. *Semin. Immunol.* *16*, 3–9. <https://doi.org/10.1016/j.smim.2003.10.003>.

Tanaka, Y., Sotome, T., Inoue, A., Mukozu, T., Kuwabara, T., Mikami, T., Kowhi-Shigematsu, T., and Kondo, M. (2017). SATB1 Conditional Knockout Results in Sjögren’s Syndrome in Mice. *J. Immunol.* *199*, 4016–4022. <https://doi.org/10.4049/jimmunol.1700550>.

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. <https://doi.org/10.1074/jbc.M311177200>.

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.* *178*, 7511–7515. <https://doi.org/10.4049/jimmunol.178.12.7511>.

Tellier, J., and Nutt, S.L. (2019). Plasma cells: The programming of an antibody-secreting machine. *Eur. J. Immunol.* *49*, 30–37. <https://doi.org/10.1002/eji.201847517>.

The Immunological Genome Project Consortium, Heng, T.S.P., Painter, M.W., Elpek, K., Lukacs-Kornek, V., Mauermann, N., Turley, S.J., Koller, D., Kim, F.S., Wagers, A.J., et al. (2008). The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* *9*, 1091–1094. <https://doi.org/10.1038/ni1008-1091>.

Thompson, N., Isenberg, D.A., Jury, E.C., and Ciurtin, C. (2016). Exploring BAFF: its expression, receptors and contribution to the immunopathogenesis of Sjögren’s syndrome. *Rheumatology* *55*, 1548–1555. <https://doi.org/10.1093/rheumatology/kev420>.

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. <https://doi.org/10.1016/j.molimm.2011.08.025>.

Tong (2010). SATB1 is an independent prognostic marker for gastric cancer in a Chinese population. *Oncol. Rep.* [https://doi.org/10.3892/or\\_00000945](https://doi.org/10.3892/or_00000945).

Tu, W., Luo, M., Wang, Z., Yan, W., Xia, Y., Deng, H., He, J., Han, P., and Tian, D. (2012). Upregulation of SATB1 promotes tumor growth and metastasis in liver cancer. *Liver Int.* *32*, 1064–1078. <https://doi.org/10.1111/j.1478-3231.2012.02815.x>.

Tumas-Brundage, K., and Manser, T. (1997). The Transcriptional Promoter Regulates Hypermutation of the Antibody Heavy Chain Locus. *J. Exp. Med.* *185*, 239–250. <https://doi.org/10.1084/jem.185.2.239>.

- Tunyaplin, C., Shaffer, A.L., Angelin-Duclos, C.D., Yu, X., Staudt, L.M., and Calame, K.L. (2004). Direct Repression of *prdm1* by Bcl-6 Inhibits Plasmacytic Differentiation. *J. Immunol.* *173*, 1158–1165. <https://doi.org/10.4049/jimmunol.173.2.1158>.
- Uribe, V., Badía-Careaga, C., Casanova, J.C., Domínguez, J.N., de la Pompa, J.L., and Sanz-Ezquerro, J.J. (2014). *Arid3b* is essential for second heart field cell deployment and heart patterning. *Development* *141*, 4168–4181. <https://doi.org/10.1242/dev.109918>.
- Vadnais, C., Davoudi, S., Afshin, M., Harada, R., Dudley, R., Clermont, P.-L., Drobetsky, E., and Nepveu, A. (2012). CUX1 transcription factor is required for optimal ATM/ATR-mediated responses to DNA damage. *Nucleic Acids Res.* *40*, 4483–4495. <https://doi.org/10.1093/nar/gks041>.
- Vadnais, C., Awan, A.A., Harada, R., Clermont, P.-L., Leduy, L., Bérubé, G., and Nepveu, A. (2013). Long-range transcriptional regulation by the p110 CUX1 homeodomain protein on the ENCODE array. *BMC Genomics* *14*. <https://doi.org/10.1186/1471-2164-14-258>.
- Victoria, G.D., and Nussenzweig, M.C. (2012). Germinal Centers. *Annu. Rev. Immunol.* *30*, 429–457. <https://doi.org/10.1146/annurev-immunol-020711-075032>.
- 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 with a Photoactivatable Fluorescent Reporter. *Cell* *143*, 592–605. <https://doi.org/10.1016/j.cell.2010.10.032>.
- Vikstrom, I., Carotta, S., Lüthje, K., Peperzak, V., Jost, P.J., Glaser, S., Busslinger, M., Bouillet, P., Strasser, A., Nutt, S.L., et al. (2010). Mcl-1 Is Essential for Germinal Center Formation and B Cell Memory. *Science* *330*, 1095–1099. <https://doi.org/10.1126/science.1191793>.
- Villa, A., Santagata, S., Bozzi, F., Giliani, S., Frattini, A., Imberti, L., Gatta, L.B., Ochs, H.D., Schwarz, K., Notarangelo, L.D., et al. (1998). Partial V(D)J Recombination Activity Leads to Omenn Syndrome. *Cell* *93*, 885–896. [https://doi.org/10.1016/S0092-8674\(00\)81448-8](https://doi.org/10.1016/S0092-8674(00)81448-8).
- Wang, Y., and Bhattacharya, D. (2014). Adjuvant-specific regulation of long-term antibody responses by ZBTB20. *J. Exp. Med.* *211*, 841–856. <https://doi.org/10.1084/jem.20131821>.
- Wang, B., Zhang, L., Dai, T., Qin, Z., Lu, H., Zhang, L., and Zhou, F. (2021). Liquid–liquid phase separation in human health and diseases. *Signal Transduct. Target. Ther.* *6*. <https://doi.org/10.1038/s41392-021-00678-1>.
- Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009a). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.* *206*, 1817–1830. <https://doi.org/10.1084/jem.20081678>.
- Wang, S., Zhou, J., Wang, X.-Y., Hao, J.-M., Chen, J.-Z., Zhang, X.-M., Jin, H., Liu, L., Zhang, Y.-F., Liu, J., et al. (2009b). Down-regulated expression of SATB2 is associated with metastasis and poor prognosis in colorectal cancer. *J. Pathol.* *219*, 114–122. <https://doi.org/10.1002/path.2575>.
- Wang, Y., Su, M., Zhou, L.L., Tu, P., Zhang, X., Jiang, X., and Zhou, Y. (2011). Deficiency of SATB1 expression in Sézary cells causes apoptosis resistance by regulating FasL/CD95L transcription. *Blood* *117*, 3826–3835. <https://doi.org/10.1182/blood-2010-07-294819>.
- Wang, Y., Gu, X., Zhang, G., Wang, L., Wang, T., Zhao, Y., Zhang, X., Zhou, Y., Kadin, M., and Tu, P. (2014). SATB1 overexpression promotes malignant T-cell proliferation in cutaneous CD30+



lymphoproliferative disease by repressing p21. *Blood* 123, 3452–3461. <https://doi.org/10.1182/blood-2013-10-534693>.

Wang, Z., Yang, X., Chu, X., Zhang, J., Zhou, H., Shen, Y., and Long, J. (2012). The structural basis for the oligomerization of the N-terminal domain of SATB1. *Nucleic Acids Res.* 40, 4193–4202. <https://doi.org/10.1093/nar/gkr1284>.

Ward, J.M., Ratliff, M.L., Dozmorov, M.G., Wiley, G., Guthridge, J.M., Gaffney, P.M., James, J.A., and Webb, C.F. (2016). Human effector B lymphocytes express ARID3a and secrete interferon alpha. *J. Autoimmun.* 75, 130–140. <https://doi.org/10.1016/j.jaut.2016.08.003>.

Webb, C.F., Das, C., Eaton, S., Calame, K., and Tucker, P.W. (1991). Novel protein-DNA interactions associated with increased immunoglobulin transcription in response to antigen plus interleukin-5. *Mol. Cell. Biol.* 11, 5197–5205. <https://doi.org/10.1128/mcb.11.10.5197-5205.1991>.

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., Yamashita, Y., Ayers, N., Evetts, S., Paulin, Y., Conley, M.E., and Smith, E.A. (2000). The Transcription Factor *Bright* Associates with Bruton's Tyrosine Kinase, the Defective Protein in Immunodeficiency Disease. *J. Immunol.* 165, 6956–6965. <https://doi.org/10.4049/jimmunol.165.12.6956>.

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. (2011). The ARID Family Transcription Factor Bright Is Required for both Hematopoietic Stem Cell and B Lineage Development. *Mol. Cell. Biol.* 31, 1041–1053. <https://doi.org/10.1128/MCB.01448-10>.

de Weers, M., Verschuren, M.C.M., Kraakman, M.E.M., Mensink, R.G.J., Schuurman, R.K.B., van Dongen, J.J.M., and Hendriks, R.W. (1993). The Bruton's tyrosine kinase gene is expressed throughout B cell differentiation, from early precursor B cell stages preceding immunoglobulin gene rearrangement up to mature B cell stages. *Eur. J. Immunol.* 23, 3109–3114. <https://doi.org/10.1002/eji.1830231210>.

Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J., and Hood, L. (1980). The joining of V and J gene segments creates antibody diversity. *Nature* 283, 497–499. <https://doi.org/10.1038/283497a0>.

Wen, J., Huang, S., Rogers, H., Dickinson, L.A., Kohwi-Shigematsu, T., and Noguchi, C.T. (2005). SATB1 family protein expressed during early erythroid differentiation modifies globin gene expression. *Blood* 105, 3330–3339. <https://doi.org/10.1182/blood-2004-08-2988>.

van Wijnen, A.J., van Gurp, M.F., de Ridder, M.C., Tufarelli, C., Last, T.J., Birnbaum, M., Vaughan, P.S., Giordano, A., Krek, W., Neufeld, E.J., et al. (1996). CDP/cut is the DNA-binding subunit of histone gene transcription factor HiNF-D: a mechanism for gene regulation at the G1/S phase cell cycle transition point independent of transcription factor E2F. *Proc. Natl. Acad. Sci.* 93, 11516–11521. <https://doi.org/10.1073/pnas.93.21.11516>.

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. <https://doi.org/10.1038/ni.2572>.

- Wilsker, D., Probst, L., Wain, H., Maltais, L., Tucker, P., and Moran, E. (2005). Nomenclature of the ARID family of DNA-binding proteins. *Genomics* 86, 242–251. <https://doi.org/10.1016/j.ygeno.2005.03.013>.
- Wöhner, M., Tagoh, H., Bilic, I., Jaritz, M., Poliakova, D.K., Fischer, M., and Busslinger, M. (2016). Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. *J. Exp. Med.* 213, 1201–1221. <https://doi.org/10.1084/jem.20152002>.
- 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. [https://doi.org/10.1016/S1074-7613\(03\)00261-9](https://doi.org/10.1016/S1074-7613(03)00261-9).
- 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.* 15, 676–684. <https://doi.org/10.1038/cmi.2017.133>.
- Xi, X. (2012). AT-rich sequence binding protein 1: Contribution to tumor progression and metastasis of human ovarian carcinoma. *Oncol. Lett.* <https://doi.org/10.3892/ol.2012.571>.
- Xiang, Y., Zhou, X., Hewitt, S.L., Skok, J.A., and Garrard, W.T. (2011). A Multifunctional Element in the Mouse *Igk* Locus That Specifies Repertoire and *Ig* Loci Subnuclear Location. *J. Immunol.* 186, 5356–5366. <https://doi.org/10.4049/jimmunol.1003794>.
- Xiang, Y., Park, S.-K., and Garrard, W.T. (2013). *Vk* Gene Repertoire and Locus Contraction Are Specified by Critical DNase I Hypersensitive Sites within the *Vk*-*Jk* Intervening Region. *J. Immunol.* 190, 1819–1826. <https://doi.org/10.4049/jimmunol.1203127>.
- Xiao, T., Fu, L., and Jie, Z. (2017). SATB1 overexpression correlates with gastrointestinal neoplasms invasion and metastasis: a meta-analysis for Chinese population. *Oncotarget* 8, 48282–48290. <https://doi.org/10.18632/oncotarget.18548>.
- Xu, X., Deobagkar-Lele, M., Bull, K.R., Crockford, T.L., Mead, A.J., Cribbs, A.P., Sims, D., Anzilotti, C., and Cornall, R.J. (2020). An ontogenetic switch drives the positive and negative selection of B cells. *Proc. Natl. Acad. Sci.* 117, 3718–3727. <https://doi.org/10.1073/pnas.1915247117>.
- Xu, Y., Davidson, L., Alt, F.W., and Baltimore, D. (1996). Deletion of the *Igk* Light Chain Intronic Enhancer/Matrix Attachment Region Impairs but Does Not Abolish *Vk**Jk* Rearrangement. *Immunity* 4, 377–385. [https://doi.org/10.1016/S1074-7613\(00\)80251-4](https://doi.org/10.1016/S1074-7613(00)80251-4).
- 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. <https://doi.org/10.1038/nsmb.1884>.
- Yamaguchi, H., Tateno, M., and Yamasaki, K. (2006). Solution Structure and DNA-binding Mode of the Matrix Attachment Region-binding Domain of the Transcription Factor SATB1 That Regulates the T-cell Maturation. *J. Biol. Chem.* 281, 5319–5327. <https://doi.org/10.1074/jbc.M510933200>.
- Yamasaki, K., Akiba, T., Yamasaki, T., and Harata, K. (2007). Structural basis for recognition of the matrix attachment region of DNA by transcription factor SATB1. *Nucleic Acids Res.* 35, 5073–5084. <https://doi.org/10.1093/nar/gkm504>.
- Yasuda, K., Kitagawa, Y., Kawakami, R., Isaka, Y., Watanabe, H., Kondoh, G., Kohwi-Shigematsu, T., Sakaguchi, S., and Hirota, K. (2019). *Satb1* regulates the effector program of encephalitogenic tissue Th17 cells in chronic inflammation. *Nat. Commun.* 10. <https://doi.org/10.1038/s41467-019-08404-w>.

- 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. <https://doi.org/10.1038/nature01084>.
- Ye, M., and Graf, T. (2007). Early decisions in lymphoid development. *Curr. Opin. Immunol.* *19*, 123–128. <https://doi.org/10.1016/j.coi.2007.02.007>.
- Yewdell, W.T., Kim, Y., Chowdhury, P., Lau, C.M., Smolkin, R.M., Belcheva, K.T., Fernandez, K.C., Cols, M., Yen, W.-F., Vaidyanathan, B., et al. (2020). A Hyper-IgM Syndrome Mutation in Activation-Induced Cytidine Deaminase Disrupts G-Quadruplex Binding and Genome-wide Chromatin Localization. *Immunity* *53*, 952–970.e11. <https://doi.org/10.1016/j.immuni.2020.10.003>.
- Yi, M., Wu, P., Trevorrow, K.W., Clafin, L., and Garrard, W.T. (1999). Evidence that the Igkappa 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. <https://doi.org/10.1038/ni.2688>.
- Yoon, S.O., and Chikaraishi, D.M. (1994). Isolation of two E-box binding factors that interact with the rat tyrosine hydroxylase enhancer. *J. Biol. Chem.* *269*, 18453–18462. .
- Yoshida, T., Yao-Ming Ng, S., Zuniga-Pflucker, J.C., and Georgopoulos, K. (2006). Early hematopoietic lineage restrictions directed by Ikaros. *Nat. Immunol.* *7*, 382–391. <https://doi.org/10.1038/ni1314>.
- Yoshinaga, S.K., Whoriskey, J.S., Khare, S.D., Sarmiento, U., Guo, J., Horan, T., Shih, G., Zhang, M., Coccia, M.A., Kohno, T., et al. (1999). T-cell co-stimulation through B7RP-1 and ICOS. *Nature* *402*, 827–832. <https://doi.org/10.1038/45582>.
- 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. <https://doi.org/10.1038/ni919>.
- Zarate, Y.A., and Fish, J.L. (2017). *SATB2* -associated syndrome: Mechanisms, phenotype, and practical recommendations. *Am. J. Med. Genet. A.* *173*, 327–337. <https://doi.org/10.1002/ajmg.a.38022>.
- Zarate, Y.A., Bosanko, K.A., Caffrey, A.R., Bernstein, J.A., Martin, D.M., Williams, M.S., Berry-Kravis, E.M., Mark, P.R., Manning, M.A., Bhamhani, V., et al. (2019). Mutation update for the *SATB2* gene. *Hum. Mutat.* <https://doi.org/10.1002/humu.23771>.
- Zelenka, T., Klonizakis, A., Tsoukatou, D., Franzenburg, S., Tzerpos, P., Papamatheakis, D.-A., Tzonevrakis, I.-R., Nikolaou, C., Plewczynski, D., and Spilianakis, C. (2021a). The 3D enhancer network of the developing T cell genome is controlled by SATB1. *bioRxiv* 2021.07.09.451769. <https://doi.org/10.1101/2021.07.09.451769>.
- Zelenka, T., Tzerpos, P., Panagopoulos, G., Tsolis, K., Papamatheakis, D.-A., Papadakis, V.M., Stanek, D., and Spilianakis, C. (2021b). SATB1 undergoes isoform-specific phase transitions in T cells. *bioRxiv*.
- Zhang, X., Yoon, H.S., Chapdelaine-Williams, A.M., Kyritsis, N., and Alt, F.W. (2021). Physiological role of the 3'IgH CBEs super-anchor in antibody class switching. *Proc. Natl. Acad. Sci.* *118*. <https://doi.org/10.1073/pnas.2024392118>.

- Zhang, Y., Tian, X., Ji, H., Guan, X., Xu, W., Dong, B., Zhao, M., Wei, M., Ye, C., Sun, Y., et al. (2014a). Expression of SATB1 Promotes the Growth and Metastasis of Colorectal Cancer. *PLoS ONE* 9, e100413. <https://doi.org/10.1371/journal.pone.0100413>.
- Zhang, Y., Ye, Q., He, J., Chen, P., Wan, J., Li, J., Yang, Y., and Li, X. (2020). Recurrence-Associated Multi-RNA Signature to Predict Disease-Free Survival for Ovarian Cancer Patients. *BioMed Res. Int.* 2020, 1–19. <https://doi.org/10.1155/2020/1618527>.
- Zhang, Z.Z., Pannunzio, N.R., Han, L., Hsieh, C.-L., Yu, K., and Lieber, M.R. (2014b). 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. <https://doi.org/10.1016/j.celrep.2014.06.021>.
- Zhao, X.-D., Ji, W., Zhang, W., He, L.-X., Yang, J., Liang, H.-J., and Wang, L.-L. (2010). Overexpression of SATB1 in Laryngeal Squamous Cell Carcinoma. *ORL* 72, 1–5. <https://doi.org/10.1159/000264777>.
- Zhou, L.-Q., Wu, J., Wang, W.-T., Yu, W., Zhao, G.-N., Zhang, P., Xiong, J., Li, M., Xue, Z., Wang, X., et al. (2012). The AT-rich DNA-binding Protein SATB2 Promotes Expression and Physical Association of Human G $\gamma$ - and A $\gamma$ -Globin Genes. *J. Biol. Chem.* 287, 30641–30652. <https://doi.org/10.1074/jbc.M112.355271>.
- Zhou, Y., Li, Y.-S., Bandi, S.R., Tang, L., Shinton, S.A., Hayakawa, K., and Hardy, R.R. (2015). Lin28b promotes fetal B lymphopoiesis through the transcription factor Arid3a. *J. Exp. Med.* 212, 569–580. <https://doi.org/10.1084/jem.20141510>.
- Zong, R.-T., Das, C., and Tucker, P.W. (2000). Regulation of matrix attachment region-dependent, lymphocyte-restricted transcription through differential localization within promyelocytic leukemia nuclear bodies. *EMBO J.* 19, 4123–4133. <https://doi.org/10.1093/emboj/19.15.4123>.



## Annexes

---

Annexe 1. The IgH E $\mu$ -MAR regions promote UNG-dependent error-prone repair to optimize somatic hypermutation .....	127
Annexe 2. Fam72a enforces error-prone DNA repair during antibody diversification.....	128
Annexe 3. Continuous MYD88 activation is associated with expansion and then transformation of IgM differentiating plasma cells .....	129



**Annexe 1. The IgH E $\mu$ -MAR regions promote UNG-dependent error-prone repair to optimize somatic hypermutation**

Ophélie Martin, Morgane Thomas, Marie Marquet, Armand Garot, Mylène Brousse, Sébastien Bender, Claire Carrion, Jee Eun Choi, Bao Q. Vuong, Patricia J. Gearhart, Robert W. Maul, Sandrine Le Noir and Eric Pinaud.

BioRxiv doi: <https://doi.org/10.1101/2022.08.15.503996>





## The IgH $E\mu$ -MAR regions promote UNG-dependent error-prone repair to optimize somatic hypermutation

**Running Title:**  $E\mu$ -MAR regions optimize hypermutation

Ophélie Martin<sup>1,5</sup>, Morgane Thomas<sup>1, ‡</sup>, Marie Marquet<sup>1, ‡</sup>, Armand Garot<sup>1</sup>, Mylène Brousse<sup>1</sup>, Sébastien Bender<sup>1,2</sup>, Claire Carrion<sup>1</sup>, Jee Eun Choi<sup>3</sup>, Bao Q. Vuong<sup>3</sup>, Patricia J. Gearhart<sup>4</sup>, Robert W. Maul<sup>4</sup>, Sandrine Le Noir<sup>1\*</sup> and Eric Pinaud<sup>1\*</sup>

<sup>1</sup>*Centre National de la Recherche Scientifique (CNRS) UMR7276, Institut National de la Santé et de la Recherche Médicale (Inserm) UMR1262-Contrôle de la Réponse Immune B et des Lymphoproliférations, Université de Limoges, Limoges, France.*

<sup>2</sup>*Centre Hospitalier Universitaire Dupuytren, Limoges, France*

<sup>3</sup>*The Graduate Center, The City University of New York, New York, NY 10016, USA*

<sup>4</sup>*Laboratory of Molecular Biology and Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA*

<sup>5</sup>*Current address: Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton, United Kingdom*

<sup>‡</sup>*M.T. and M.M. 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. or SLN

## Abstract

Two scaffold/matrix attachment regions (5'- and 3'- $MARS_{E\mu}$ ) 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  $MARS_{E\mu}$  is still unclear and their involvement in somatic hypermutation (SHM) has never been deeply evaluated. By analysing a mouse model devoid of  $MARS_{E\mu}$ , we observed an inverted substitution pattern: SHM being decreased upstream from  $cE\mu$  and increased downstream of it. Strikingly, the SHM defect induced by  $MARS_{E\mu}$ -deletion was accompanied by an increase of sense transcription of the IgH V region, excluding a direct transcription-coupled effect. Interestingly, by breeding to DNA repair-deficient backgrounds, we showed that the SHM defect, observed upstream from  $cE\mu$  in this model, was not due to a decrease in AID deamination but rather the consequence of a defect in base excision repair-associated unfaithful repair process. Our study pointed out an unexpected “fence” function of  $MARS_{E\mu}$  regions in limiting the error-prone repair machinery to the variable region of Ig gene loci.

## Introduction

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 & 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 & 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 & 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 (Perlot *et al*, 2005; Afshar *et al*, 2006) 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 (Bachl & Wabl, 1996; Azuma *et al*, 1993; Giusti & Manser, 1993; Motoyama *et al*, 1994; Lin *et al*, 1998; 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 VB1-8i region, *E $\mu$*  deletion still resulted in a high level 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; this minor effect was likely due to the reduced inflow of peripheral and, consequently, GC B cells in this model (Perlot *et al*, 2005). Strikingly, the role of

$MARS_{E\mu}$  was also elusive and somewhat controversial. Whereas their endogenous deletion, analysed in mouse chimeras by the RAG-2 complementation assay, demonstrated that  $MARS_{E\mu}$  are dispensable for VDJ recombination and IgH expression (Sakai *et al*, 1999b), the ambiguous function of  $MARS_{E\mu}$  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 ( $Ig\kappa$ ), 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 KO mouse model that accumulated premature light chain rearrangements with a mild SHM defect (Yi *et al*, 1999), an effect comparable to one observed at  $IgH$  locus in hybridoma cells devoid of  $MARS_{E\mu}$  (Ronai *et al*, 2005). At that time, while these studies instigated a variety of hypotheses accounting for MARs in modulating SHM (Franklin & Blanden, 2005), these were contradicted by a study comparing  $3'E\kappa$ - and  $MAR_{Ei\kappa}$ - function in mouse 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  $MARS_{E\mu}$  and bred it into DNA repair-deficient backgrounds. In our models devoid of  $MARS_{E\mu}$  and their *wt* counterparts, we proceeded to side by side comparison of total SHM, transcription patterns, AID targeting and error prone repair events leading to SHM, in regions located upstream and downstream from the intronic enhancer. Our study showed that the absence of  $MARS_{E\mu}$  allows some of the error-prone repair machinery to get access to the region downstream from the  $E\mu$  enhancer. We propose that  $MARS_{E\mu}$  act as physiological barriers for error-prone repair in activated B cells. As a rational hypothesis, our study suggests that the conservation of nuclear matrix attachment regions in Ig genes serve to optimize SHM events within the variable regions.

## Results

### Normal B cell development and Ig production in the absence of $MARs_{E\mu}$

We generated a mouse mutant line carrying an endogenous deletion of both the 5' and 3' *IgH* matrix attachment regions that flank the  $J_H$ - $C_H$  intronic  $cE\mu$  enhancer. Although generated with slightly different targeting vector backbone and homology arms, the resulting *IgH* allele, so-called  $MARs_{E\mu}^{\Delta}$  (Fig. 1A), is similar to that generated by Sakai *et al.* (Sakai *et al.*, 1999b). Bone marrow subsets of B cell precursors were analysed in *wt* and homozygous  $MARs_{E\mu}^{\Delta/\Delta}$  deficient mice. When compared to age-matched *wt* animals,  $MARs_{E\mu}^{\Delta/\Delta}$  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),  $MARs_{E\mu}$  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  $MARs_{E\mu}^{\Delta/\Delta}$  mutants (Supplementary Table 1). In agreement with the normal inflow of mature B cells in  $MARs_{E\mu}^{\Delta/\Delta}$  animals, Peyer's patches were efficiently colonized by naive and GC B cells. Numbers and proportions of GC B cells were even significantly increased in homozygous mutants (Fig. 1B left panels and Supplementary Table 1). The similar proportion of proliferating KI67<sup>+</sup> GC B cells in *wt* and  $MARs_{E\mu}^{\Delta/\Delta}$  mice implied that this increase was not due to over-proliferation of Peyer's patch B cells (Fig. 1B right panels). Finally, levels of serum Ig isotypes were unaffected in  $MARs_{E\mu}^{\Delta/\Delta}$  animals (Fig. 1C). This normal B cell homeostasis in homozygous mutants confirmed that  $MARs_{E\mu}$  are dispensable for B cell ontogeny and antibody production. This statement is in agreement with previous studies of an analogous MAR region in the *Ig $\kappa$*  locus (Sakai *et al.*, 1999b; Yi *et al.*, 1999).

### $MARs_{E\mu}$ deletion inverts SHM distribution on both sides of the $E\mu$ enhancer region

To assess whether  $MARs_{E\mu}$  deletion could affect *IgH* somatic hypermutation, we first quantified mutations within the 500-bp regions downstream from the variable exons rearranged to  $J_H3$  and  $J_H4$  segments (Fig. 2A) in Peyer's patch GC B cells sorted from *wt* and  $MARs_{E\mu}^{\Delta/\Delta}$  (overall data reported in Fig. 2 left, data from individual animals reported in Supplementary Figure S1A and B and Supplementary Table 2A and B). For this we used two complementary sequencing methods: the first one, based on classical Sanger approach and GS junior technology, allowed to discriminate and exclude unmutated and clonally related sequences from the calculation of SHM frequency, as initially described (Rada *et al.*, 1991). The second method used Ion proton deep sequencing coupled to

DeMinEr filtering. Since using DNA templates including non-mutated alleles, this second approach underestimated the SHM frequency; but since including AID-deficient control samples as a reference, the method provided highly reproducible and reliable quantification of SHM in a DNA sample extracted from GC B cells (Martin *et al*, 2018). Interestingly, by using Sanger approach,  $MARS_{E\mu}^{\Delta/\Delta}$  GC B cells displayed significant differences in the distribution of mutations: an increased proportion of unmutated sequences (less than 10% in *wt* compared to 38% in  $MARS_{E\mu}^{\Delta/\Delta}$ ) (from 30.6 to 45.5%, overall data collected from several mice, data from independent mice in Supplementary Fig. S1A). Another effect of  $MARS_{E\mu}$  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 ~24% (from 18.9 to 28.3%) while in mutants they were barely present (~2%) (Fig. 2B left and Supplementary Fig S1A). When comparing only the mutated sequences, mutation frequency was decreased at least by two fold in  $MARS_{E\mu}^{\Delta/\Delta}$  mutants, with 7.6 mutations per 1000 bp (in average) compared to 14.9 in *wt* (in average) (Fig. 2B left and Supplementary Figure S1A). By using next generation sequencing (NGS), the decreased SHM frequency was also highly significant (Fig. 2C, 12.4‰ vs 8.5‰,  $p=0.008$ , individual mice in Supplementary Table S2A). To monitor SHM upon antigen challenge, we analyzed mutations by Sanger and NGS in a large number of GC B cells sorted from spleen of SRBC-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 (Sanger method, excluding unmutated clones) or 3.4 (NGS bulk method) mutations per 1000bp in *wt* cells to respectively 4.6 or 2.3 in  $MARS_{E\mu}^{\Delta/\Delta}$  cells (Fig. 2B middle, and 2C, Supplementary Fig.S1B, Table S2B). Although not statistically significant, this showed that mutations accumulated by GC B cells devoid of  $MARS_{E\mu}$  were already decreased only eight days after antigen challenge. 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. 2B). This confirmed that the intronic region was less efficiently targeted by SHM in  $MARS_{E\mu}$  deficient mice. An identical SHM defect was also observed in mice harbouring deletion of the entire  $E\mu$  region (core enhancer and flanking MARs; (Marquet *et al*, 2014)) (Supplementary Fig. S1A top right); this data indicated that the SHM failure was the only consequence of MARs deletion. This hypothesis is completely consistent with a previous study showing that SHM efficiency was not affected by the endogenous deletion of the  $cE\mu$  enhancer alone (Perlot *et al*, 2005). The comparison between those three models is certainly relevant since all knock outs were created in murine germlines with similar mixed genetic backgrounds.

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 (Lennon & Perry, 1985; Cockerill *et al*, 1987), we also quantified SHM immediately downstream from this exon in a 600bp region described as mutated in GC B cells (Fig

2A) (Nagaoka *et al*, 2002; Petersen *et al*, 2001). Unlike the intronic regions downstream from the rearranged *VDJ* exon, the overall mutation frequency downstream from  $I\mu$  was strongly increased in GC B cells devoid of  $MAR_{S_{E\mu}}$  region and reached 6.9 (in average) mutations per 1000 bp compared to 3.3 (in average) in *wt* cells (Fig. 2B right). This suggests that the region downstream from the  $cE\mu$  was more efficiently targeted in the absence of its *MARs*. This was supported by the very low proportion of unmutated sequences (less than 3% in Fig. 2B right and Supplementary S1C) and the increased proportion of highly mutated sequences (more than 8%) in  $MAR_{S_{E\mu}}$ -deficient GC B cells (Fig. 2B right and data from individual mice reported in Supplementary Fig. S1C). This data was efficiently confirmed by NGS analysis that estimated that mutation frequency was increased by 3.5 fold in the  $MAR_{S_{E\mu}}$ -deficient GC B compared to *wt* mice (1.7‰ vs 5.8‰) (Fig 2C right and Supplementary Table S2C).

Analysis of mutation distribution in *wt* and in  $MAR_{S_{E\mu}}^{\Delta\Delta}$  did not show any difference between models (Fig 2D), indicating that, while affecting SHM efficiency, the absence of  $MAR_{S_{E\mu}}$  region did not influence DNA sequence hotspot or preferences for SHM within the  $J_H4$  intron.

Importantly, our mouse model clearly assigns a specific function for endogenous  $MAR_{S_{E\mu}}$  on SHM at the *IgH* locus, in accord with the requirement of similar regions for efficient SHM previously pointed out in the endogenous *Igκ* 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). While our study suggests that  $MAR_{S_{E\mu}}$  optimizes SHM upstream from the  $cE\mu$  enhancer; the presence of such regulatory regions does not prevent the SHM machinery to get access to downstream regions as reported in a recent study (Heltzel *et al*, 2022). This hypothesis is mostly supported by the increased SHM frequency downstream from the  $cE\mu$  enhancer in the absence of  $MAR_{S_{E\mu}}$ , a finding consistent with previous works describing increased  $S\mu$  internal deletions in hybridomas devoid of *MARs* regions (Sakai *et al*, 1999a). We could speculate that one physiological function of  $MAR_{S_{E\mu}}$  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. Such a "locked" target conformation could provide an optimal environment for somatic mutations by trapping the transcription machinery and its co-factors including AID and error-prone repair factors. 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).



### ***MARS<sub>E $\mu$</sub>* deletion modifies transcription patterns on both sides of the *E $\mu$* enhancer region**

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 total amounts of total *IgH* primary transcripts by using multiple q-PCR probes located respectively downstream from *J<sub>H4</sub>* and *J<sub>H3</sub>*: the previously described probe A (Tinguely *et al*, 2012) (Fig. 3A and Supplementary FigS2) complemented by probes A' and C (Supplementary Figures S2 and S3A). 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<sub>E $\mu$</sub>* -deficient cells in both GC and *in vitro*-stimulated samples by using probe A (Fig. 3B) as well as with probes A' and C (Supplementary Fig. S3A), although an upward trend could be noticed in LPS-activated samples. The intriguing discrepancy between the mutation phenotype observed in *MARS<sub>E $\mu$</sub>* -deficient GC 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<sub>H4</sub>* segment (S1 and S2) and within the *cE $\mu$*  enhancer (S3) (Fig. 3C and Supplementary Fig. S2 and S3A). 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<sub>H2</sub>* (AS0), *J<sub>H3</sub>* (AS1) and *J<sub>H4</sub>* (AS2 and AS3) (Fig. 3E and Supplementary Fig. S2 and S3B). For both sense and antisense, quantification of transcripts was possible with the same probes A, A' and C. For strand-specific quantification assays with a given probe, the baseline level was either provided by a control reaction (P-) measuring endogenous priming since devoid of primer or by one strand-specific template that cannot be detected by the probe (T-) as reported previously (Zhao *et al*, 2009; Bolland *et al*, 2004). To note, strand-specific transcripts were optimally detected when primers and probes were closer (sense transcripts with primer S1/probe A or antisense transcripts with primer AS2/probe A). Side by side comparison of *wt* and *MARS<sub>E $\mu$</sub>* -deficient activated B cells samples revealed several interesting differences. When quantified with optimal primer S1/probe A tandem, sense transcripts were significantly increased in the absence of *MARS<sub>E $\mu$</sub>*  (Fig. 3C). In GC B cells, a two fold increase was noticed by using S1 template (Fig. 3C, left bar graph,  $p=0.019$ ). In *in vitro*-activated cells, an increase of sense transcription was also observed upon *MARS<sub>E $\mu$</sub>* -deletion, this effect became significant for long transcripts that reach the *cE $\mu$*  (Fig. 3C, right bar graphs,  $p=0.004$  for S3/probe A). By using A' and C probes, sense transcripts were hardly detectable in GC samples (Supplementary Fig. S3A; middle

bar graphs); although a significant increase was noticed with S3 and S2/probe C tandems in LPS-activated samples upon  $MARs_{E\mu}$ -deletion (Supplementary Fig. S3A; right bar graphs). As a potential consequence of the increased transcription of the  $VDJ$  unit in observed upon  $MARs_{E\mu}$  deletion, we measured by flow cytometry the level of intracellular  $Ig\mu$  chain in Peyer's patch naive ( $B220^+/GL7^{neg}$ ) or germinal centre B cells ( $B220^+/GL7^+$ ) of *wt* and  $MARs_{E\mu}^{\Delta/\Delta}$  mice. In both cell types, the significant increase of intracellular  $Ig\mu$  chain observed in the absence of  $MARs_{E\mu}$  region (Fig. 3D) corroborate our sense-transcription data. This indicated that the absence of  $MARs_{E\mu}$  certainly did not hamper RNA pol II machinery to progress 3' to the  $VDJ$  unit and might even facilitate this process in activated cells.

Globally less abundant than their sense counterparts, quantification of antisense transcripts running downstream from  $J_H$  segments showed quite different patterns (Fig. 3E and Supplementary Fig. S3B). While quite similar levels were detected in LPS-activated samples (Fig. 3E right and Supplementary Fig. S3B), intronic antisense transcripts were about 2 fold less abundant in  $MARs_{E\mu}$ -deficient GC B cells when detection was allowed by optimal probe/primer combination (Fig. 3E, left,  $p=0.025$ , AS2/probe A and Supplementary Fig. S3B left,  $p=0.041$  for AS3/probe A').

The obvious unbalanced sense/antisense transcription ratio could result from either weak transcription efficiency or instability of antisense products. Nevertheless, Perlot *et al.* identified by RACE assays, in normal GC B cells, multiple antisense-transcript initiation start sites downstream from every  $J_H$  region and raised the question of specific enhancers. Our current data refines this previous study by identifying  $MARs_{E\mu}$  as potential boosters of antisense transcripts that, given their proximity to the enhancer, could achieve some regulatory function like eRNA or PROMPT/uaRNA (Li *et al.*, 2016). Highlighting a correlation between mutation efficacy and strand-specific transcription pattern upstream from  $E\mu$ , our data support the idea that some level antisense transcription downstream from the  $VDJ$  exon could prepare to SHM (Perlot *et al.*, 2008). Seemingly transient, specific to cell subsets and occurring upstream from an enhancer, such antisense transcripts could be substrates for RNA exosome and lead to optimized SHM targeting as proposed by Basu and colleagues (Lim *et al.*, 2017; Laffleur *et al.*, 2017, 2021).

Since a strong increase of mutations was observed within the  $S\mu$  region in the absence of  $MARs_{E\mu}$ , we also sought to correlate SHM and transcription on the other side of  $cE\mu$  by quantifying total transcripts (probe B) running in this region (Fig. 3A). In this case and according to what could be expected, transcription was significantly increased in  $MARs_{E\mu}$ -deficient GC B cells from Peyer's patches (Fig.3B left,  $p=0.04$ ); a similar trend, although not significant, was observed in LPS activated B cells (Fig 3B left). Accordingly, we also observed a modest but reproducible increase of CSR  $C\gamma3$  and  $C\gamma1$  in  $MARs_{E\mu}$ -deficient B cells stimulated *in vitro* respectively by LPS or by LPS + IL4 cocktail (Supplementary Fig. S4). A similar modest CSR effect associated to an increase of  $S\mu$  internal deletions has been previously reported in hybridomas carrying the same  $MARs_{E\mu}$ -deletion (Sakai *et al.*,

1999a). This indicated that the absence of  $MARs_{E\mu}$  lead to a global increase in transcription of the donor S region and consequently favours SHM targeting.

The significant changes in transcription patterns upstream and downstream from  $cE\mu$  observed in our models put forward the hypothesis that  $MARs_{E\mu}$  act as physiological barriers in activated B cells, limiting sense transcription of the  $VDJ$  unit up to the intronic enhancer. For transcription running through the  $S\mu$  region, our data is in agreement with a repressive function of  $MARs_{E\mu}$  in activated B cells, in order to limit SHM targeting of this area. However, our data also suggest that  $MARs_{E\mu}$  act as transcriptional repressors of the  $VDJ$  unit in both naïve and activated cells; a statement in contradiction with our hypothesis that  $MARs_{E\mu}$  facilitates SHM upstream from  $cE\mu$ . To settle such a discrepancy in our  $MARs_{E\mu}$ -deficient B cells, we first questioned AID deamination efficiency and second error-prone repair pathways processing in SHM targeted regions: upstream and downstream from  $cE\mu$ .

#### **$MARs_{E\mu}$ deletion impairs error-prone repair pathway upstream from the $E\mu$ enhancer region**

One critical experiment needed to challenge the function of  $MARs_{E\mu}$  as physiological barrier for SHM machinery was to first assess whether IgH AID targeting could be impaired in the absence of  $MARs_{E\mu}$ . To proceed, we bred our  $MARs_{E\mu}$ -KO mice in a genetic background deficient for both base excision repair ( $Ung^{\Delta/\Delta}$ ) and mismatch repair ( $Msh2^{\Delta/\Delta}$ ) in order to evaluate, on an unbiased manner, the DNA footprint of AID deamination upstream and downstream from  $cE\mu$  (Fig 4 A). As expected and according to the literature (Rada *et al*, 2004; Shen *et al*, 2006; Liu *et al*, 2008), models deficient for both BER and MMR displayed only transitions at C/G pairs reflecting cytidine deamination on respectively the template and non-template strands. By looking at deamination frequencies between control ( $Ung^{\Delta/\Delta} Msh2^{\Delta/\Delta}$ ) and mutant animals ( $Ung^{\Delta/\Delta} Msh2^{\Delta/\Delta} MARs_{E\mu}^{\Delta/\Delta}$ ), our data showed that AID activity upstream from  $cE\mu$  was not impeded upon  $MARs_{E\mu}$ -deletion; while differences were not statistically significant (evaluated on n=3 to 4 mice of each genotype), cytidine deamination even tended to be increased in B cells devoid of  $MARs_{E\mu}$ , on both sides of  $cE\mu$  (Fig 4B and Table S3A). When compared to control animals ( $Ung^{\Delta/\Delta} Msh2^{\Delta/\Delta}$ ), nucleotide substitution patterns were unchanged in the absence of  $MARs_{E\mu}$  (Fig S5A), proving identical strand-specific cytidine deamination: roughly 2/3 on the template strand (C to T substitutions) and 1/3 on the non-template strand (G to A substitutions). Besides imbalanced transcription upstream from  $E\mu$ , this data indicates that  $MARs_{E\mu}$ -deletion does not impact the choice of any DNA strand for AID targeting within intronic regions.

This notable increased AID deamination footprint prompted by  $MARs_{E\mu}$ -deletion was in total agreement with the increased transcription observed in the corresponding regions of activated B cells. The obvious discrepancy between efficient C to U deamination events and the strong SHM targeting defect within the same  $J_H$  intron region unravel the origin of the SHM defect in  $MARs_{E\mu}$ -

deficient mice as a default of the mutagenic process occurring downstream from the normally-introduced U-G mismatches in DNA.

This prompted us to investigate whether  $MARS_{E\mu}$ -deletion could provoke skewed mutation patterns within SHM-targeted regions. In the intron region downstream from  $J_H4$ , mutation frequency at each of the four bases in *wt* and  $MARS_{E\mu}$ -deficient backgrounds (Fig 4C and Supplementary Fig S5B) revealed a global significant decrease of mutations at all bases except for substitutions occurring at C in the absence of  $MARS_{E\mu}$ . Similarly, beyond significant differences for C>A, G>C, T>A and T>G events, individual mutation patterns unveiled a global decrease that did not offer any clear hypothesis regarding the mechanism impeding SHM upon  $MARS_{E\mu}$ -deletion (Supplementary Fig S5B).

To solve this paradox, we bred our  $MARS_{E\mu}$ -KO mice into base excision repair deficient background ( $Ung^{\Delta/\Delta}$ ) and analysed SHM in the same region. Strikingly, in the absence of UNG, SHM frequency within the  $J_H4$  intron region was identical upon the presence ( $Ung^{\Delta/\Delta}$  control mice) or the absence of  $MARS_{E\mu}$  ( $Ung^{\Delta/\Delta} MARS_{E\mu}^{\Delta/\Delta}$  mice) (Fig.4D and Supplementary Table S3B). Beyond the expected increase of G/C transitions, a typical hallmark of UNG-deficient background, substitution frequencies at all four bases were also identical in  $Ung^{\Delta/\Delta} MARS_{E\mu}^{\Delta/\Delta}$  mice (Fig 4D) the same was true when looking at individual substitution events (Supplementary Fig S5C). The fact that the SHM deficiency induced by  $MARS_{E\mu}$  deletion was no more observed in UNG-deficient background (Fig 4C and 4D) strongly imply the involvement of BER pathway in the initial mutagenic defect. This same data also proved that SHM events occurring independently of UNG (*e.g.* altogether obtained by replication across U and/or processed by MMR pathway) took place normally within the  $J_H$  intron in the absence of  $MARS_{E\mu}$ . Given this, a rational hypothesis to explain the origin of the SHM defect in our model was that abasic sites generated by UNG upstream from  $cE\mu$  are processed differently upon the absence of  $MARS_{E\mu}$ . Our data suggests that U:G mismatches processed by UNG are accurately repaired in the absence of  $MARS_{E\mu}$  while these are normally subject to error-prone repair; sustaining for a specific function of  $MARS_{E\mu}$  in recruiting mutagenic BER-associated factors.

In contrast to what observed in the  $J_H$  intron, substitution frequencies and mutation patterns downstream from  $cE\mu$  evidenced a different function for such regulatory regions. In B cells capable of BER and MMR, the absence of  $MARS_{E\mu}$  significantly boosted mutations at all bases by at least two fold (Fig 4F), this was true for any kind of substitution (Supplementary Fig S5E). Substitution patterns collected in mutant animals devoid of BER and MMR highlighted a global “overtargeting” of the  $S\mu$  region induced by  $MARS_{E\mu}$  deletion (Fig. 4E, Supplementary Fig. S5D and Supplementary Table S3C). This was in line with the general increase in both  $S\mu$  germline transcription observed in this model. In models impaired for BER, our data showed that UNG-deficiency combined to deletion of  $MARS_{E\mu}$  ( $Ung^{\Delta/\Delta} MARS_{E\mu}^{\Delta/\Delta}$  mice) maintained the SHM burden downstream from  $cE\mu$  significantly higher than

what observed for UNG alone (*Ung*<sup>ΔΔ</sup> control mice) (Fig 4G, Supplementary Fig S5F and Supplementary Table S3D). Such a comparison suggests that error-prone repair factors could more readily access to abasic sites generated in the S region when *MARs*<sub>Eμ</sub> are missing. In this way, our data support the idea that *MARs*<sub>Eμ</sub> act as physiological barrier that optimize SHM upstream from the Eμ region and rationalize the fact that MARs are evolutionary conserved downstream from Ig gene V regions (Yi *et al*, 1999); and moreover conserved structures in mammals (Scheuermann & Garrard, 1999).

## Concluding remarks

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). Beyond the major enhancer regions, *e.g.* *cE $\mu$*  and the *3'RR*, our current study identifies *MAR<sub>S $E\mu$</sub>* , also conserved upon any rearrangement, as part of these most critical *IgH* elements. Taking advantage of the deletion of *MAR<sub>S $E\mu$</sub>*  in the mouse, our current study rationalizes some of the molecular mechanisms that physiologically enhance SHM, through the ability of MAR to delimit some error-prone repair processes upstream from the *E $\mu$*  intronic enhancer.

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 & Gearhart, 1990; Shen & Storb, 2004; Wright *et al*, 2008). The hypothesis that *MAR<sub>S $E\mu$</sub>*  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 & Storb, 2004). A model proposed by Alt and colleagues (Meng *et al*, 2014) would be that the optimal chromatin environment for AID-induced mutations would be provided by convergent transcription as the result of fine balance between sense and antisense events. It is now broadly admitted that RNA pol II stalling is involved in SHM-targeting (Kodgire *et al*, 2013; Maul *et al*, 2014). While DIVAC regions within Ig genes were initially described as SHM regulatory regions (Kohler *et al*, 2012), a recently study proposes that such elements facilitate RNA pol II before AID targeting (Tarsalainen *et al*, 2022). The necessity of antisense transcription for mutations is still debated, as possible byproducts of RNA Pol II collision, antisense or regulatory transcripts in such regions remain transient and difficult to detect in a *wt* context; probably because processed by RNA exosome or other RNase activities (Basu *et al*, 2011; Pefanis *et al*, 2014). Our model clearly demonstrates that, while modifying sense and antisense transcription pattern, *MAR<sub>S $E\mu$</sub>*  deletion does not impede AID footprint but rather some of repair mechanisms acting downstream from the U-G mismatch. Our data also indicate that only BER-dependent error-prone repair is impeded by *MAR<sub>S $E\mu$</sub>*  deletion and suggests that efficient MMR-dependent error-prone repair in this region does not require such a barrier. The question of MARs binding factors and their respective dynamic association to such regulatory regions needs to be further investigated. The literature already suggest that some of them, like the Special AT-rich binding factor 1 (SATB1), could act as accessory factors in BER (Kaur *et al*, 2016). Recent findings showing the critical for the UNG2-interacting protein FAM72A to promote error prone processing of U-G mismatch in Ig genes (Feng *et al*, 2021; Rogier *et al*, 2021) raises the question of its specific recruitment to AID-targeted regions; our current study suggests that *MAR<sub>S $E\mu$</sub>*  could potentially interact with UNG2 or its associated error-

prone factors. Another future challenge remains to define whether some components of the nuclear matrix, nuclear filaments or proteins anchored in the envelope, could be involved in the anchorage of SHM targets.

## 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.*, 1999b). 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. *wt*,  $MARS_{E\mu}^{\Delta/\Delta}$ ,  $E\mu^{\Delta/\Delta}$  (Marquet *et al.*, 2014),  $Msh2^{\Delta/\Delta}$ ,  $Ung^{\Delta/\Delta}$  (a kind gift of Dr S. Storck) and  $Aicda^{-/-}$  (a kind gift of Pr. T. Honjo) homozygous mice were used for our experiments and maintained at 21–23°C with a 12-h light/dark cycle. Procedures were reviewed and approved by the Ministère de l'Éducation Nationale de l'Enseignement Supérieur et Recherche autorisation APAFIS#16639-2018090612249522v2.

**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 5' probe (0.803 kpb *SacI-SphI* fragment) and 3' probe (0.8 kpb *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).

**Total serum Ig quantification by ELISA.** Sera were collected at 8 weeks of age from non-immunized *wt* and  $MARS_{E\mu}^{\Delta/\Delta}$  mice and analyzed for the presence of different Ig classes and subclasses by ELISA as previously described (Pinaud *et al.*, 2001).

**SRBC Immunisation.** Mice were challenged by intraperitoneal injection with 200 $\mu$ L 50% sheep red blood cell suspension and sacrificed 8 days later to collect GC B cells in the spleen.

**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,  $\alpha$ CD117-PE,  $\alpha$ 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 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 PNA, GL7,  $\alpha$ B220, and  $\alpha$ FAS reagents and sorted based on distinct gates defined as germinal centre B cells ( $B220^+ / GL7^+$  or  $B220^+ / PNA^{High} / Fas^+$ ).

**Cell culture.** Splenocytes were collected, after red blood cells lysis,  $CD43^+$  cells were depleted using anti- $CD43$  MicroBeads (Miltenyi Biotec).  $CD43^-$  splenic B cells were cultured for 3 days at a density of  $1 \times 10^6$  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  $1 \mu\text{g}/\text{ml}$  LPS (Invivogen) alone (for transcription assays) or plus  $20 \text{ng}/\text{ml}$  IL4 (PeproTech) (for CSR 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 GS Junior (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<sup>®</sup> High-Fidelity DNA Polymerase (New England Biolabs) according to the following program: DNA was denatured 40 s at  $98^\circ\text{C}$  and then submitted to 38 cycles consisting of  $98^\circ\text{C}$  for 10 s,  $68^\circ\text{C}$  for 30 s and  $72^\circ\text{C}$  for 30 s, and 1 cycle at  $72^\circ\text{C}$  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 *IgHJ<sub>4</sub>*-downstream intron or *Smu* and only full length sequences were kept for mutation analysis. For *IgHJ<sub>4</sub>*, 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 frequencies 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<sup>™</sup> Plus gDNA Fragment Library Preparation (Cat. no. 4471269, Life Technologies). Briefly, PCR products (100ng) were fragmented by enzymatic digestion (Ion Shear<sup>™</sup> 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 (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 (10 $\mu$ M) (sequence available in Supplementary Fig. S2.) using superscript III enzyme (Invitrogen). As control, we performed a reverse transcription without primer to determine the threshold (referred as P $\bar{}$  in bar graphs). Each real-time qPCR reaction was performed in duplicate on 10 ng of RNA equivalent, using TaqMan Universal (except for S $\mu$  quantification we used 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 $\mu$*  : Smu-Fw (5'-ACCCAGGCTAAGAAGGCAAT-3'), Smu-Rev (5'-CCTTCCTTCTGCGTATCCAT-3'). 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<sup>-( $\Delta\Delta CT$ )</sup> method. The *wt* mice templates used as calibrators were S2 for sense transcripts, AS0 or AS2 for antisense transcripts.

**Statistical analysis.** If not specified in the figure legend, Mann Whitney two-tailed tests were used for statistical analysis using GraphPad Prism software (\*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001).

**Data availability.** NGS data are available at the European Nucleotide Archive (PRJEB52221).

## **Acknowledgments**

We thank BISCEm and the animal core facility team for help with mouse work on both practical and regulatory aspects. We are grateful to Drs. Fred Alt for providing  $MARs_{E\mu}$  targeting construct and Zéliha Oruc-Ratinaud regarding the animal health status. We thank Drs. Dominik Schenten, Sébastien Storck, Christophe Sirac, Laurent Delpy, Brice Laffleur, Alexis Saintamand and Jeanne Moreau for discussions and helpful comments. OM and MT were supported by PhD fellowship of the french Ministère de l'Enseignement Supérieur, de la Recherche et de de l'Innovation. This work was supported by Région Nouvelle Aquitaine, La Ligue Contre le Cancer (comités 87, 23 to EP and SLN); the Fondation ARC pour la recherche sur le cancer (PJA 20181207918 to EP and PhD continuation fellowship to MT), Institut CARNOT CALYM, INCa-Cancéropôle GSO Emergence (to EP).

## **Author contribution**

OM, MT, MM, AG, MB, SB, CC, JEC, EP and SLN performed experiments. EP and SLN conceived and supervised the study. MM developed the experimental model. OM, BQV, PJG, RWM, EP and SLN wrote the manuscript.

## **Additional information**

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

## References

- Afshar R, Pierce S, Bolland DJ, Corcoran A & Oltz EM (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 LE & Loh DY (1993) Mutations of the chloramphenicol acetyl transferase transgene driven by the immunoglobulin promoter and intron enhancer. *Int Immunol* 5: 121–130
- Bachl J & 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 CR, Wesemann DR, *et al* (2011) The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144: 353–363
- Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, Fraser PJ & Corcoran AE (2004) Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* 5: 630–637
- Cockerill PN, Yuen MH & Garrard WT (1987) The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *J Biol Chem* 262: 5394–5397
- Feng Y, Li C, Stewart JA, Barbulescu P, Seija Desivo N, Álvarez-Quilón A, Pezo RC, Perera MLW, Chan K, Tong AHY, *et al* (2021) FAM72A antagonizes UNG2 to promote mutagenic repair during antibody maturation. *Nature* 600: 324–328
- Franklin A & Blanden RV (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 & Rajewsky K (1998) Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* 9: 105–114
- Giusti AM & 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 & 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 GT, Riddell A, Neuberger MS & 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
- Heltzel JHM, Maul RW, Yang W & Gearhart PJ (2022) Promoter Proximity Defines Mutation Window for VH and VK Genes Rearranged to Different J Genes. *J Immunol*: ji2101002
- Inlay MA, Gao HH, Odegard VH, Lin T, Schatz DG & Xu Y (2006) Roles of the Ig kappa light chain intronic and 3' enhancers in Ikg somatic hypermutation. *J Immunol* 177: 1146–1151
- Kaplan MH, Zong RT, Herrscher RF, Scheuermann RH & Tucker PW (2001) Transcriptional activation by a matrix associating region-binding protein. contextual requirements for the function of bright. *J Biol Chem* 276: 21325–21330

- Kaur S, Coulombe Y, Ramdzan ZM, Leduy L, Masson J-Y & Nepveu A (2016) Special AT-rich Sequence-binding Protein 1 (SATB1) Functions as an Accessory Factor in Base Excision Repair. *J Biol Chem* 291: 22769–22780
- Kobayashi M, Sabouri Z, Sabouri S, Kitawaki Y, Pommier Y, Abe T, Kiyonari H & Honjo T (2011) Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proc Natl Acad Sci USA* 108: 19305–19310
- Kodgire P, Mukkavar P, Ratnam S, Martin TE & 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 KM, McDonald JJ, Duke JL, Arakawa H, Tan S, Kleinstein SH, Buerstedde J-M & Schatz DG (2012) Identification of core DNA elements that target somatic hypermutation. *J Immunol* 189: 5314–5326
- Kohwi-Shigematsu T, Maass K & Bode J (1997) A thymocyte factor SATB1 suppresses transcription of stably integrated matrix-attachment region-linked reporter genes. *Biochemistry* 36: 12005–12010
- Laffleur B, Basu U & Lim J (2017) RNA Exosome and Non-coding RNA-Coupled Mechanisms in AID-Mediated Genomic Alterations. *J Mol Biol*
- Laffleur B, Lim J, Zhang W, Chen Y, Pefanis E, Bizarro J, Batista CR, Wu L, Economides AN, Wang J, *et al* (2021) Noncoding RNA processing by DIS3 regulates chromosomal architecture and somatic hypermutation in B cells. *Nat Genet* 53: 230–242
- Lebecque SG & Gearhart PJ (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 GG & Perry RP (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 DA & Eckhardt LA (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 & Rosenfeld MG (2016) Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet* 17: 207–223
- Lim J, Giri PK, Kazadi D, Laffleur B, Zhang W, Grinstein V, Pefanis E, Brown LM, Ladewig E, Martin O, *et al* (2017) Nuclear Proximity of Mtr4 to RNA Exosome Restricts DNA Mutational Asymmetry. *Cell* 169: 523-537.e15
- Lin MM, Green NS, Zhang W & Scharff MD (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
- Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH & Schatz DG (2008) Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451: 841–845
- Marquet M, Garot A, Bender S, Carrion C, Rouaud P, Lecardeur S, Denizot Y, Cogné M & 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 OA, Garot A, Le Noir S, Aldigier J-C, Cogné M, Pinaud E & 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 RW, Cao Z, Venkataraman L, Giorgetti CA, Press JL, Denizot Y, Du H, Sen R & Gearhart PJ (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 RW, Saribasak H, Cao Z & Gearhart PJ (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 RM, Amor C, Wasserman CR, 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 & 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 & 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
- Pefanis E, Wang J, Rothschild G, Lim J, Chao J, Rabadan R, Economides AN & Basu U (2014) Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514: 389–393
- Perlot T & Alt FW (2008) Cis-regulatory elements and epigenetic changes control genomic rearrangements of the IgH locus. *Adv Immunol* 99: 1–32
- Perlot T, Alt FW, Bassing CH, Suh H & Pinaud E (2005) Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci USA* 102: 14362–14367
- Perlot T, Li G & Alt FW (2008) Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proc Natl Acad Sci USA* 105: 3843–3848
- Petersen S, Casellas R, Reina-San-Martin B, Chen HT, Difilippantonio MJ, Wilson PC, Hanitsch L, Celeste A, Muramatsu M, Pilch DR, *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 AA, Le Morvan C, Drouet M, Nalesso V, Le Bert M & Cogné M (2001) Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* 15: 187–199
- Pommier Y, Sun Y, Huang S-YN & Nitiss JL (2016) Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* 17: 703–721
- Rada C, Di Noia JM & Neuberger MS (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
- Rada C, Gupta SK, Gherardi E & Milstein C (1991) Mutation and selection during the secondary response to 2-phenylloxazalone. *Proc Natl Acad Sci U S A* 88: 5508–5512

- Rogier M, Moritz J, Robert I, Lescale C, Heyer V, Abello A, Martin O, Capitani K, Thomas M, Thomas-Claudepierre A-S, *et al* (2021) Fam72a enforces error-prone DNA repair during antibody diversification. *Nature* 600: 329–333
- Ronai D, Berru M & Shulman MJ (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 MD, Fan M, Shulman MJ & Scharff MD (2005) Complex regulation of somatic hypermutation by cis-acting sequences in the endogenous IgH gene in hybridoma cells. *Proc Natl Acad Sci USA* 102: 11829–11834
- Rouaud P, Vincent-Fabert C, Saintamand A, Fiancette R, Marquet M, Robert I, Reina-San-Martin B, Pinaud E, Cogné M & 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 & Alt FW (1999a) The Ig heavy chain intronic enhancer core region is necessary and sufficient to promote efficient class switch recombination. *Int Immunol* 11: 1709–1713
- Sakai E, Bottaro A, Davidson L, Sleckman BP & Alt FW (1999b) 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 USA* 96: 1526–1531
- Scheuermann RH & Garrard WT (1999) MARs of antigen receptor and co-receptor genes. *Crit Rev Eukaryot Gene Expr* 9: 295–310
- Shen HM & Storb U (2004) Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc Natl Acad Sci USA* 101: 12997–13002
- Shen HM, Tanaka A, Bozek G, Nicolae D & Storb U (2006) Somatic hypermutation and class switch recombination in Msh6(-/-)Ung(-/-) double-knockout mice. *J Immunol* 177: 5386–5392
- Tarsalain A, Maman Y, Meng F-L, Kyläniemi MK, Soikkeli A, Budzyńska P, McDonald JJ, Šenigl F, Alt FW, Schatz DG, *et al* (2022) Ig Enhancers Increase RNA Polymerase II Stalling at Somatic Hypermutation Target Sequences. *J Immunol* 208: 143–154
- Teves SS & 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 & Delpy L (2012) Cross talk between immunoglobulin heavy-chain transcription and RNA surveillance during B cell development. *Mol Cell Biol* 32: 107–117
- Wang Z, Goldstein A, Zong RT, Lin D, Neufeld EJ, Scheuermann RH & Tucker PW (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 EJ, Ronai D, Berru M, Tsui FW & Shulman MJ (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 BE, Schmidt KH, Davis N, Hunt AT & Minnick MF (2008) II. Correlations between secondary structure stability and mutation frequency during somatic hypermutation. *Mol Immunol* 45: 3600–3608

Xue K, Rada C & Neuberger MS (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 KW, Claflin L & Garrard WT (1999) Evidence that the Igkappa gene MAR regulates the probability of premature V-J joining and somatic hypermutation. *J Immunol* 162: 6029–6039

Zhao Y, Dunn-Walters DK, Barone F & Spencer J (2009) Antisense transcripts of V(D)J rearrangements; artifacts caused by false priming? *Mol Immunol* 46: 2357–2362



## Figure legends

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

(A) Schematic representation of *wt* and  $MAR_{E\mu}\Delta$  alleles (top). Targeting construct and Southern blot performed on recombinant ES cells with  $Neo^R$  insertion. Hybridization with the 5' probe detected 4 kpb and 5 kpb *Sacl* genomic fragments respectively for *wt* and recombined alleles. Hybridization with the 3' probe detected 8 kpb and 6 kpb *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  $MAR_{E\mu}^{\Delta/\Delta}$  animals by flow cytometry: dot plots showed percentage of naïve ( $B220^+/GL7^-$ ) and GC ( $B220^+/GL7^+$ ) B cells (left panels) and, for each subset, the percentage of dividing cells ( $Ki67^+$ ) was indicated on cell count histogram plots (right panels). Experiments were performed twice with a minimum of 3 mice per group. (C) Immunoglobulin isotype secretion in sera from *wt* and  $MAR_{E\mu}^{\Delta/\Delta}$  mice determined by ELISA (n=9 to 12 mice, mean $\pm$ SEM).

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

(A) Location of *IgH* regions (thick purple lines) tested for SHM, arrows represent primers used for PCR amplification. (B) Pie charts represent distribution of mutated sequences (proportional to the area in each slice, data obtained by Sanger and GS Junior sequencing method) quantified in *wt* and  $MAR_{E\mu}^{\Delta/\Delta}$  mice in individually recombined *IgH* alleles. For each genotype number of individual clones is indicated in the center (after removal of clonally related sequences based on VDJ junction) and overall mutation frequencies (mutation per 1000 bp in mutated clones) are indicated below. Left: SHM downstream from  $J_H3$  and  $J_H4$  segments in Peyer's patch sorted GC B cells, data obtained after cloning and sequencing by classical Sanger method. Middle: SHM downstream from  $J_H4$  segments in spleen GC B cells sorted from SRBC-immunized mice, data obtained by NGS (GS Junior). Right: SHM downstream from *cE $\mu$*  region from Peyer's patch GC sorted B cells, data obtained by both classical Sanger method and NGS (GS Junior). (C) Graphical representation of SHM frequency in *wt* and  $MAR_{E\mu}^{\Delta/\Delta}$  mice, quantified by NGS (Ion Proton) submitted to DeMinEr filtering, a pipeline that identifies substitution frequency at each nucleotide based on an *Aicda* <sup>$\Delta/\Delta$</sup>  control sample (Martin *et al.*, 2018). Since no indication in sequence distribution is available using this method, data were represented as scattered plots, each point refers to a mutation frequency from one individual mice, overall mutation frequencies are indicated above. Mean  $\pm$  SEM are represented. (D) Mutation distribution along the  $J_H4$  intron in *wt* (top) and in  $MAR_{E\mu}^{\Delta/\Delta}$  (bottom).

**Figure 3.  $MARS_{E\mu}$  deletion impairs strand-specific transcription upstream from  $E\mu$  region**

(A) *IgH* locus with the location of q-PCR probes (A and B) used for transcripts quantification. (B) Total primary transcripts quantified with A and B q-PCR probes in Peyer's patch GC B cells (dark colors) and *in vitro*-activated B cells (light colors) from *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice. (C) Detection of sense transcripts (dotted arrows) in murine *IgH* locus (not to scale). Arrows indicate primers (S1, S2, S3) downstream from  $J_H3$  and  $J_H4$  used for strand-specific reverse transcription. Primary sense transcripts were quantified with A q-PCR probe (indicated by a black bar) in Peyer's patch GC B cells and *in vitro*-activated B cells from *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice. Dots indicate antisense transcript start sites according to Perlot *et al.* (Perlot *et al.*, 2008). Baseline levels were defined in a control retrotranscription reaction performed without primers (P-). Bar graphs show the three (S1, S2 and S3) relative sense transcripts quantity (mean $\pm$ SEM) of two to three independent experiments. (D) Intracellular IgM 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. Bar graphs indicate data from individual mice ( $n=6$  mice in 2 independent experiments, mean $\pm$ SEM); a representative example of cell count overlay is associated. (E) Detection of antisense transcripts (dotted arrows) in murine *IgH* locus (not to scale). Arrows indicate primers (AS1, AS2, AS3) downstream from  $J_H3$  and  $J_H4$  used for strand-specific reverse transcription. Primary antisense transcripts were quantified with A q-PCR probe (indicated by a black bar) in Peyer's patch GC B (dark colours) cells and *in vitro*-activated (light colors) B cells from *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice. Dots indicate antisense transcripts start sites according to Perlot *et al.* (Perlot *et al.*, 2008). Baseline levels were defined using a control retrotranscription reaction performed without primers (P-) or using a strand-specific template that cannot be detected with A q-PCR probe (T-). Bar graphs show mean $\pm$ SEM of two to three independent experiments.

**Figure 4.  $MARS_{E\mu}$  deletion impedes error-prone repair pathways upstream from  $E\mu$  region**

Comparison of *IgH* SHM events occurring in Peyer's patch GC B cells sorted from *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice models, bred in genetic backgrounds deficient for base excision repair (*Ung* KO) and mismatch repair (*Msh2* KO). Data were obtained by NGS (Ion Proton) combined to DeMinEr filtering (Martin *et al.*, 2018). In each region, analyzed and represented as a panel, bar graphs report overall mutation frequencies (left) and detailed mutation frequencies at all bases (right). (A) Location of *IgH* regions (thick purple lines) tested for SHM, arrows represent primers used for PCR amplification. (B) SHM downstream from  $J_H4$  in double-deficient  $Ung^{\Delta\Delta} Msh2^{\Delta\Delta}$  background. (C) SHM downstream from  $J_H4$  in DNA repair proficient ( $Ung^{+/+} Msh2^{+/+}$ ) background. (D) SHM downstream from  $J_H4$  in  $Ung^{\Delta\Delta}$  background. (E) SHM downstream from  $cE\mu$  in double-deficient  $Ung^{\Delta\Delta} Msh2^{\Delta\Delta}$  background. (F) SHM

downstream from  $cE\mu$  in DNA repair proficient ( $Ung^{+/+} Msh2^{+/+}$ ) background. **(G)** SHM downstream from  $cE\mu$  in  $Ung^{\Delta\Delta}$  background. Bar graphs show mean $\pm$ SEM of two to three independent experiments.

**Supplementary Figure S1.  $MARs_{E\mu}$  deletion inverts SHM distribution on both sides of the  $E\mu$  enhancer region**

**(A)** SHM 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; after cloning and sequencing by classical Sanger method. For each genotype, pie charts represent distribution of mutated sequences (proportional to the area in each slice, data obtained by Sanger and GS Junior sequencing method) in individually recombined *IgH* alleles. Number of individual clones is reported in the center (after removal of clonally related sequences based on VDJ junction). Each pie chart represent SHM obtained from an individual experiment. Under each pie chart, SHM frequency, sequencing strategy and sample type (individual mice or pool) is indicated. Mean SHM frequency and p values are reported. **(B)** Equivalent data representation than reported in **A** for SHM downstream from  $J_H4$  segments in splenic GC B cells sorted from SRBC-immunized *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  mice. Mean SHM frequency and p values are reported. **(C)** Equivalent data representation than reported in **(A)** for SHM within the intron 5' to  $S\mu$  region in Peyer's patch GC B cells sorted from *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  mice

**Supplementary Figure S2. Annotated nucleotide map of the *IgH-J\_H1* to  $E\mu$  germline region from of 129 *wt* mice**

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

**Supplementary Figure S. Sense and antisense transcripts quantified with *IgH J\_H3* and  $J_H 4$  exons with additional TaqMan probes**

**(A)** Murine *IgH* locus (not to scale) indicating location of primers (S1, S2, S3; black arrows) within introns downstream from  $J_H3$  and  $J_H4$  used for strand-specific reverse transcription to detect sense transcripts (dotted arrows). Black bars (A' and C) indicate location of q-PCR probes. Total primary transcripts and primary sense transcripts were quantified with A' and C probes in Peyer's patch GC B cells (filled bar graphs) and *in vitro*-activated B cells (bar graphs) from *wt* and  $MARs_{E\mu}^{\Delta\Delta}$  mice. Dots indicated antisense transcripts start sites according to Perlot et al. (Perlot et al, 2008). **(B)** Murine *IgH* locus (not to scale) indicating location of primers (AS0, AS1, AS2, AS3; black arrows) within introns

downstream from  $J_H2$ ,  $J_H3$  and  $J_H4$  used for strand-specific reverse transcription to detect antisense transcripts (dotted arrows). Black bars (A' and C) indicate location of q-PCR probes. Primary antisense transcripts quantified with A' and C probes in Peyer's patch GC B cells (filled bar graphs) and *in vitro*-activated B cells (emptied bar graphs) from *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice. Dots indicated antisense transcripts start sites according to published data (Perlot *et al*, 2008). Baseline level was either provided using a control retrotranscription reaction performed without primers (P-) or using one strand-specific template that cannot be detected with the current probe (T-). Bar graphs show mean $\pm$ SEM of two to three independent experiments.

#### **Supplementary Figure S4. Comparison of *in-vitro* Ig class switching in *wt* and $MARS_{E\mu}^{\Delta\Delta}$ mice**

Percentage of IgG3 and IgG1 positive cells measured by flow cytometry after respectively LPS or LPS + IL4 stimulation for 3 days of splenic B cells sorted from *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice.

#### **Supplementary Figure S5. Base substitution patterns in BER- and MMR-deficient backgrounds**

Comparison of SHM-related base substitution patterns, reported as frequencies, at *IgH* in Peyer's patch GC B cells sorted from *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice models, bred in genetic backgrounds deficient for base excision repair (*Ung* KO) and mismatch repair (*Msh2* KO). Data were obtained by NGS (Ion Proton) combined to DeMinEr filtering (Martin *et al*, 2018). (A) Substitution pattern downstream from  $J_H4$  in double-deficient  $Ung^{\Delta\Delta} Msh2^{\Delta\Delta}$  background. (B) Substitution pattern downstream from  $J_H4$  in DNA repair proficient ( $Ung^{\Delta\Delta} Msh2^{\Delta\Delta}$ ) background. (C) Substitution pattern downstream from  $J_H4$  in  $Ung^{\Delta\Delta}$  background. (D) Substitution pattern downstream from  $cE\mu$  in double-deficient  $Ung^{\Delta\Delta} Msh2^{\Delta\Delta}$  background. (E) Substitution pattern downstream from  $cE\mu$  in DNA repair proficient ( $Ung^{\Delta\Delta} Msh2^{\Delta\Delta}$ ) background. (F) Substitution pattern downstream from  $cE\mu$  in  $Ung^{\Delta\Delta}$  background. Bar graph show mean $\pm$ SEM of two to three independent experiments.

#### **Table S1. $MARS_{E\mu}$ deletion led to normal B-lineage cell development**

Bone Marrow and peripheral B cell subsets counts in *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice. Absolute numbers are reported as mean $\pm$  SEM. Significance was assessed with Student T test. P value is indicated when difference is significant.

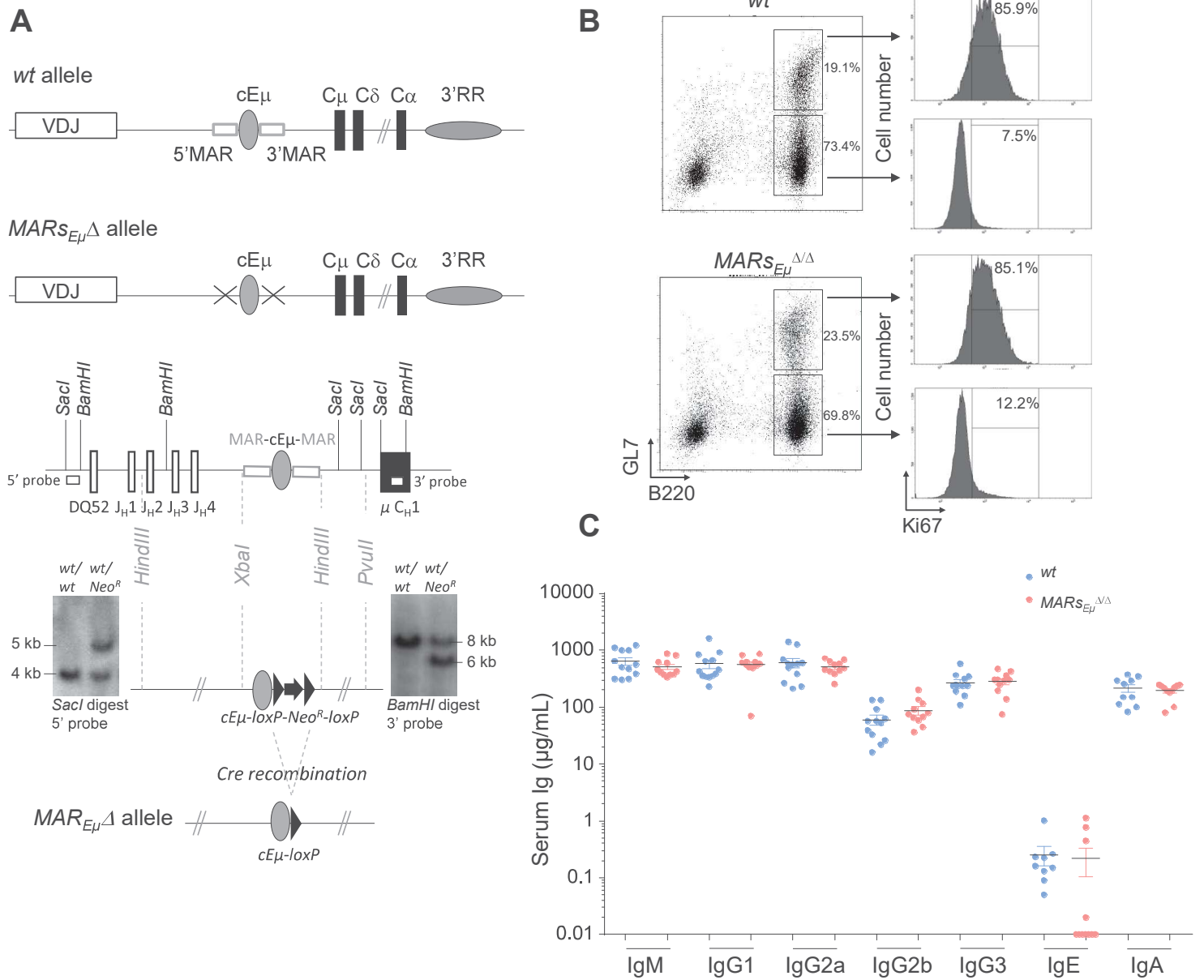
#### **Table S2. SHM data (NGS) from individual mice in DNA repair proficient background**

Total number of mutations, total number of bp analyzed and mutation frequencies for *wt* and  $MARS_{E\mu}^{\Delta\Delta}$  mice. (A) Data from intron 3' to  $J_H4$  in Peyer's patches GC B cells, (B) Data from spleen GC B cells from SRBC-immunized mice, (C) Data from intron 5' to  $S\mu$  in Peyer's patches GC B cells.

**Table S3. SHM data (NGS) from individual mice in genetic backgrounds deficient for base excision repair (*Ung*-deficient) and mismatch repair (*Msh2*-deficient)**

Total number of mutations, total number of bp analyzed and mutation frequencies. **(A)** Data from intron 3' to *J<sub>H</sub>4* in Peyer's patches GC B cells of *Ung*<sup>ΔΔ</sup> *Msh2*<sup>ΔΔ</sup> and *Ung*<sup>ΔΔ</sup> *Msh2*<sup>ΔΔ</sup> *MARS<sub>Eμ</sub>*<sup>ΔΔ</sup> mice, **(B)** Data from *Ung*<sup>ΔΔ</sup> and *Ung*<sup>ΔΔ</sup> *MARS<sub>Eμ</sub>*<sup>ΔΔ</sup> mice, **(C)** Data from intron 5' to *Sμ* in Peyer's patches GC B cells of *Ung*<sup>ΔΔ</sup> *Msh2*<sup>ΔΔ</sup> and *Ung*<sup>ΔΔ</sup> *Msh2*<sup>ΔΔ</sup> *MARS<sub>Eμ</sub>*<sup>ΔΔ</sup> mice, **(D)** Data from *Ung*<sup>ΔΔ</sup> and *Ung*<sup>ΔΔ</sup> *MARS<sub>Eμ</sub>*<sup>ΔΔ</sup> mice.

Figure 1



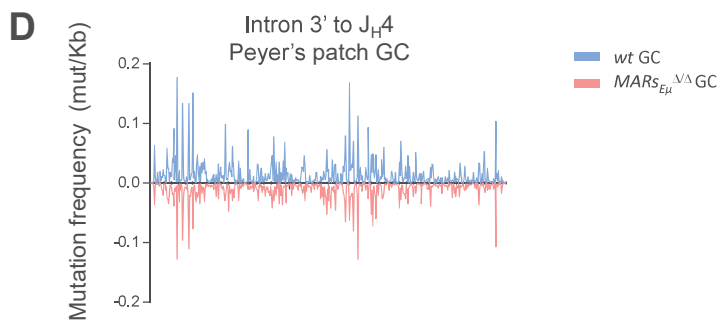
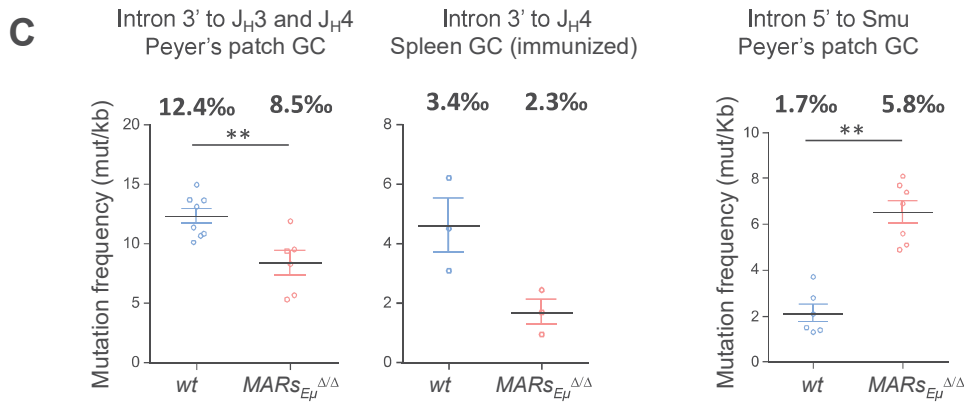
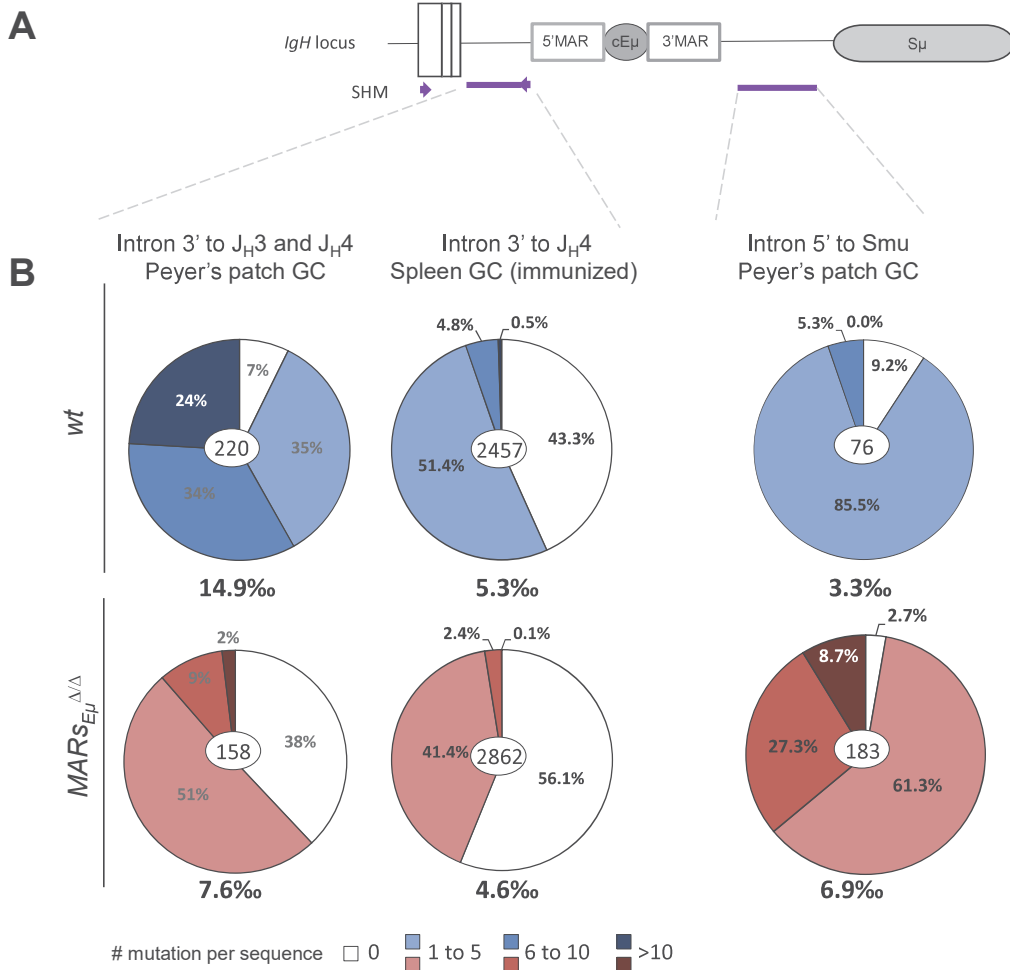
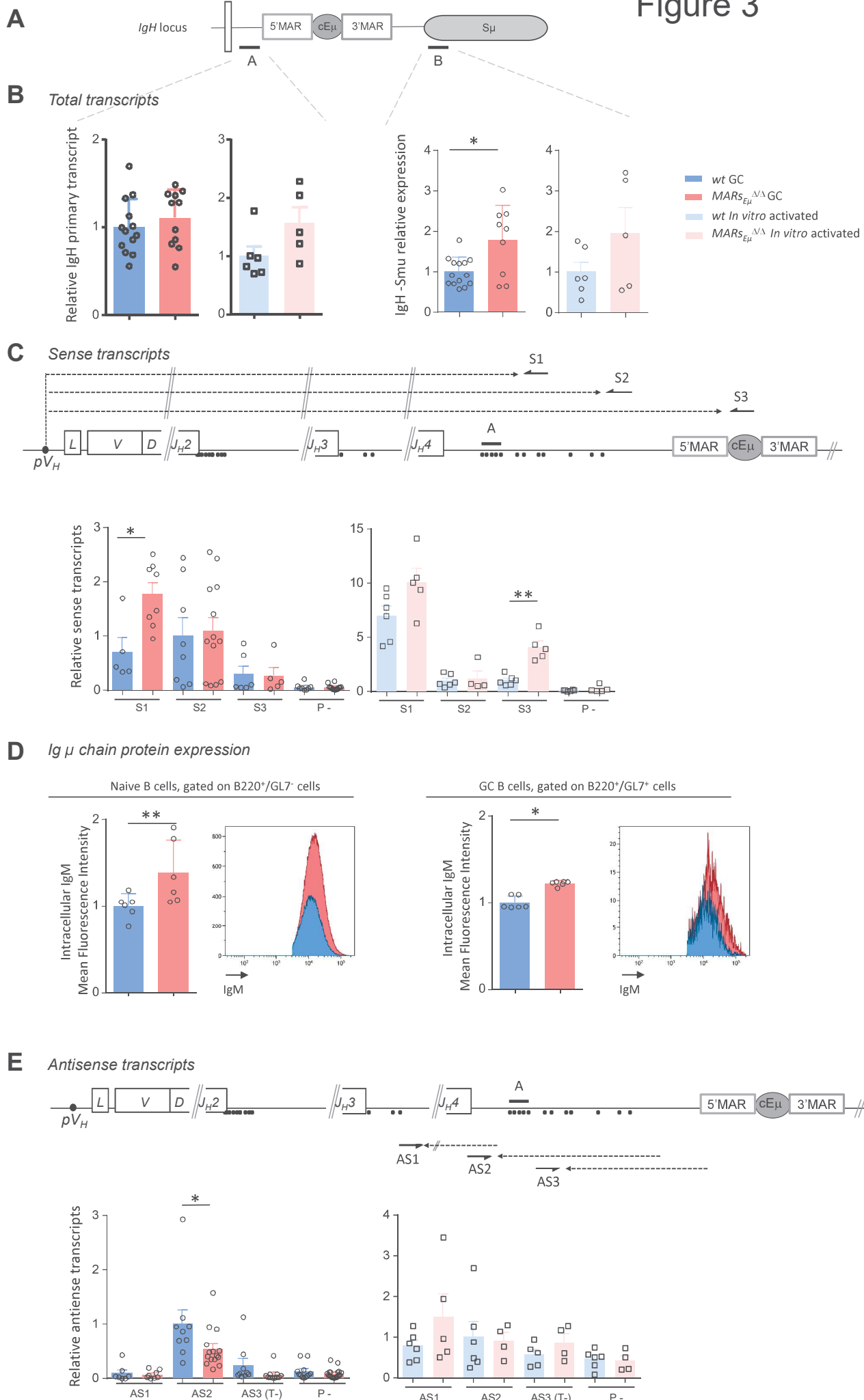


Figure 3





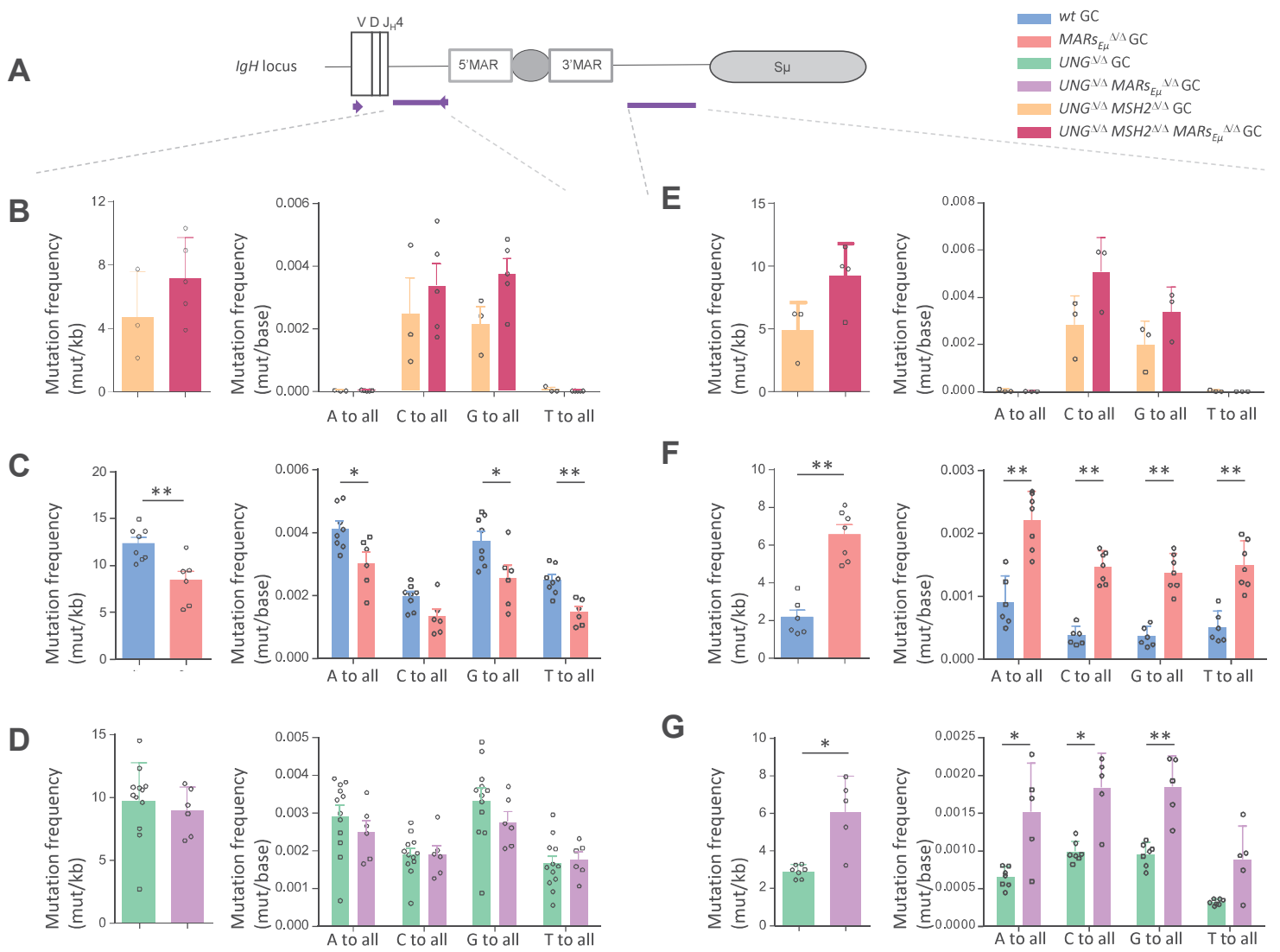


Figure 4

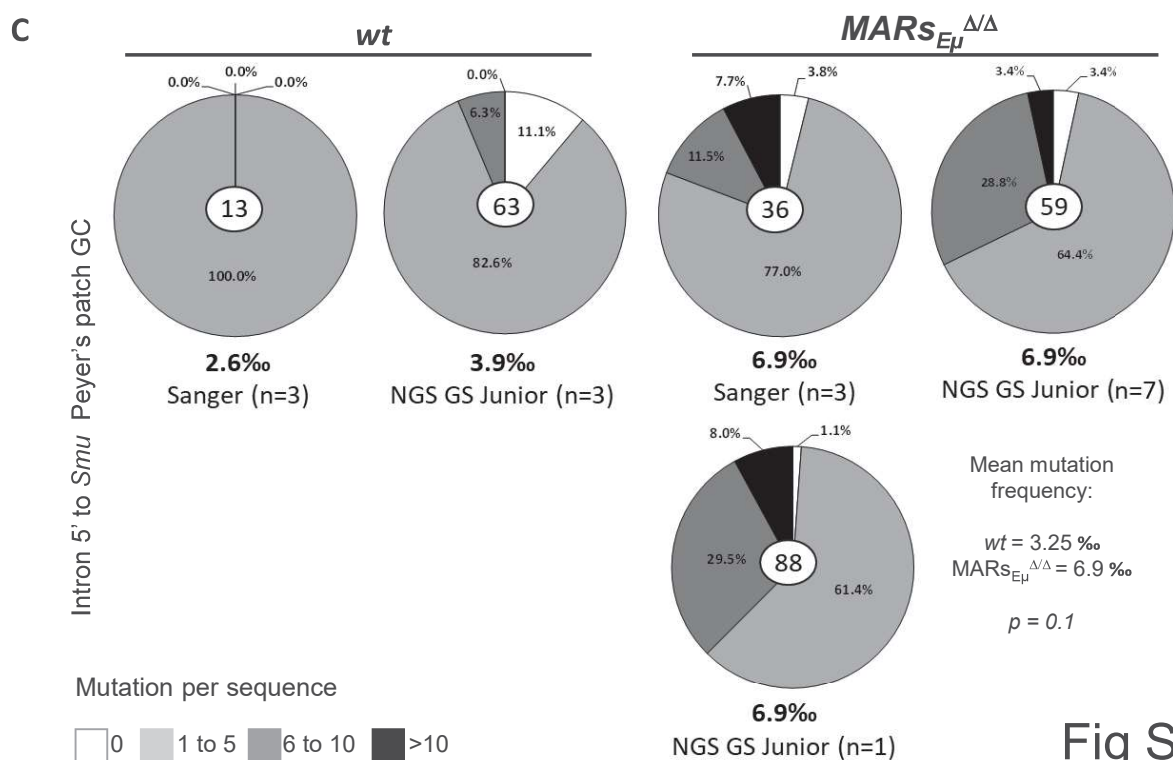
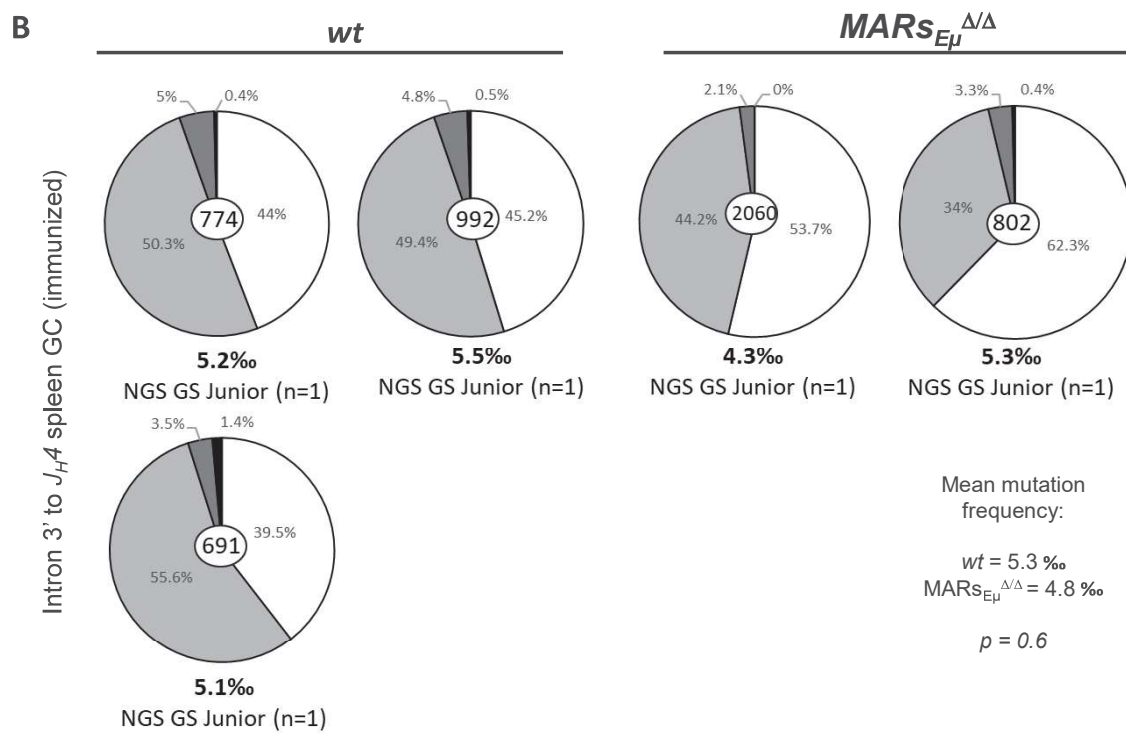
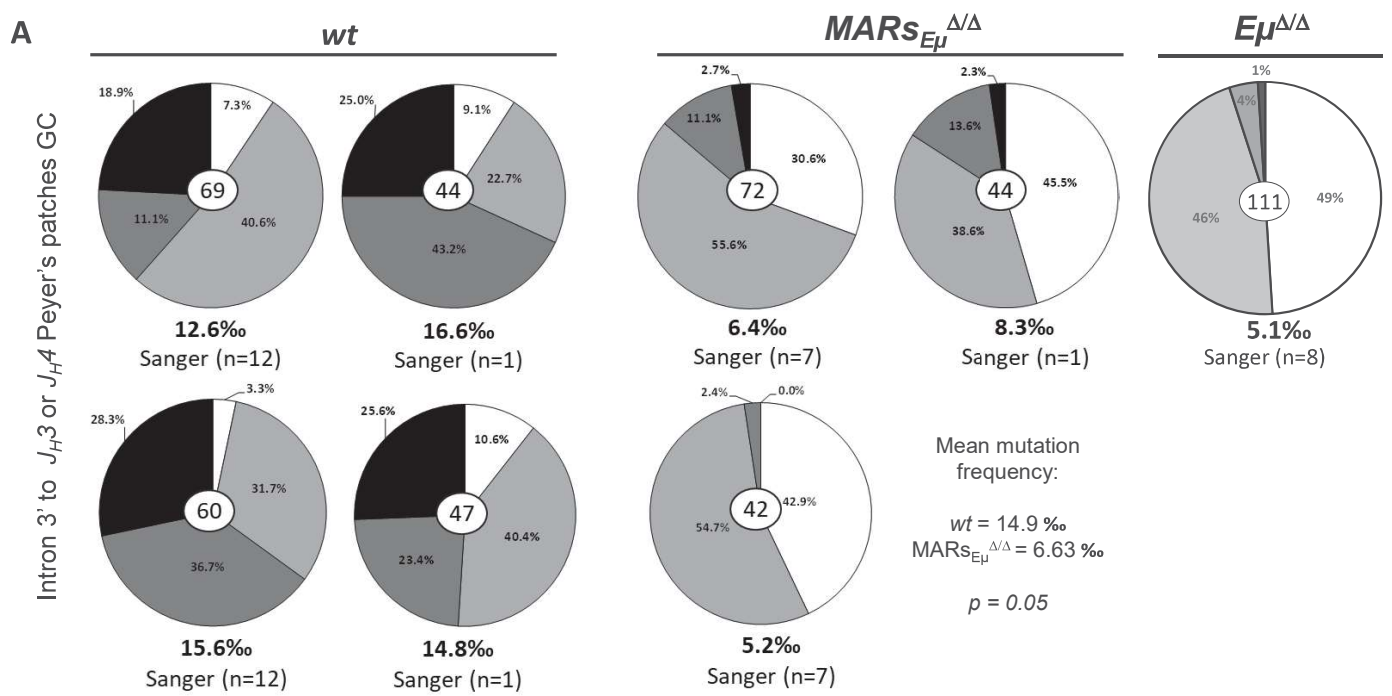


Fig S1

**J<sub>H</sub>1**  
**ctactggtacttcgatgtctggggcgcagggaccacggtcaccgtctcctcag**gtaagctggccttttttctttctgcac  
 attccattctgaaatgggaaaagatattctcagatctccccatggtcaggccatctgccacactctgcatgctgcagaag  
 cttttctgtaaggatagggctttcactcccaggaaaagaggcagtcagaggctagctgcctgtggaacagtgacaatca  
 tggaaaataggcatttacattggttaggctacatgggtagatgggtttttgtacaccactaaaggggtctatgatagtg  
 tg

**J<sub>H</sub>2**  
**actactttgactactggggccaaggcaccactctcacagtcctcctcag**gtgagtccttacaacctctctcttctattca  
 gcttaaatagattttactgcatttgggtgggggggaaatgtgtgtatctgaatttcagggtcatgaaggactagggacacc  
 ttgggagtcagaaaggggtcattggggagccctggctgagcagacagacatcctcagctcccagacttcatggccagaga  
 tttatagggatcctggccagcattgcccgtaggtccctctcttctatgctttctttgtccctcactggcctccatctga  
 gataatcctggagccctagccaaggatcatttattgtcaggggtctaatcattggtgtcacaatgtg

**J<sub>H</sub>3**  
**cctggtttgcttactggggccaagggactctggctcactgtctctcag**gtgagtcctaacttctccattctaaatgca  
 tgttggggggattctgagccttcaggaccaagattctctgcaaacgggaatcaagattcaaccctttgtcccaaagtt  
 gagacatgggtctgggtcagggactctctgcctgctggctctgtggtgacattagaactgaagtatgatgaaggatctgc  
 cagaactgaagcttgaagctgagggcagaatcttgtccagggctctatcggactcttgtgagaattaggggctgacagtt  
 gatggtgacaatttcaggggtcagtgactgtctgggtttctctgaggtgaggcttggaatataggtcaccttgaagacttaa  
 gaggggtccagggggcttctgcacagggcagggaaacagaatgtggaacaatgacttgaatgggttattcttgtgtgacac  
 caggaattggcataatgtctgagttgccaggggtgattctagtcagactctgggggtttttgtcgggtatagaggaaaa  
 atccactattgtg

**J<sub>H</sub>4**  
**attactatgctatggactactggggtaaggaacctcagtcaccgtctcctcag**gtaagaatggcctctccaggtcttt  
 atttttaacctttgttatggagttttctgagcattgcagactaatcttgggatatttgtccctgagggagccggctgaga  
 gaagttgggaaataaactgtctagggatctcagagccttttaggacagattatctccacatctttgaaaaactaagaatc  
 tgtgtgatggtggtggtgagtcctggatgatgggatagggactttggaggctcatttggggagatgctaaacaat  
 cctatggctggagggatagttggggctgtagttggagattttcagtttttagaataaaagtattagctgcggaatatac  
 ttcaggaccacctctgtgacagcatttatacagtatccgatgcatagggacaaagagtggagtggggcactttcttttag  
 atttgtgaggaatggtccacactagattggtttaaacttcatttgttggaggagagctgtcttagtgattgagtgcaag  
 ggagaaaggcatctagcctcggtctcaaaagggtagttgctg

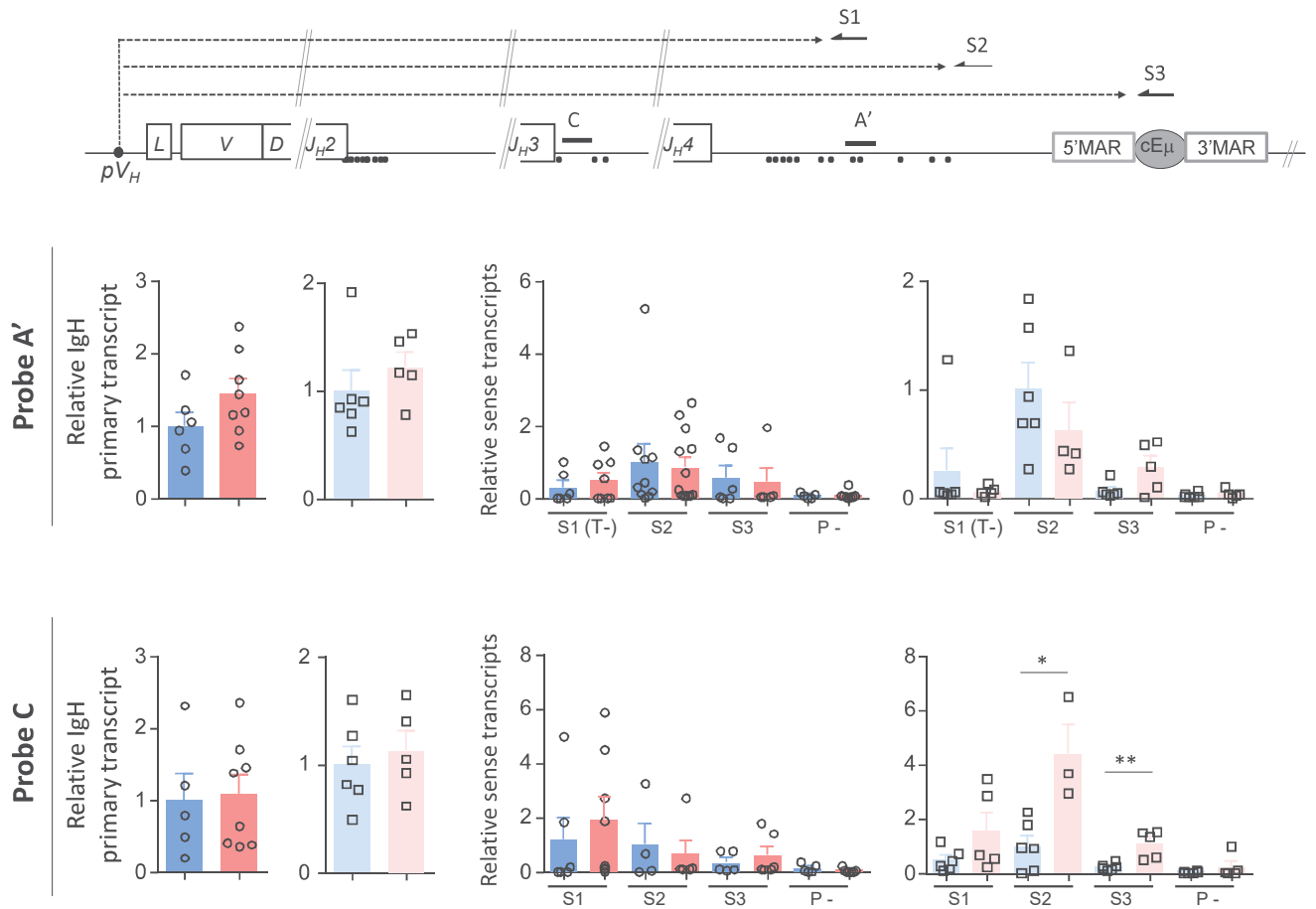
5' MAR  
 tctagagaggtctgggtggagcctgcaaaagtccagctttcaaaggaacacagagaagtatgtgtatggaatattagaagat  
 gttgcttttactcttaagttggttcctaggaaaaatagttaaatactgtgactttaaagtgtgagaggggttttcaagta  
 ctcatTTTTTaaatgtccaaaatTTTTgtcaatcaatttgaggcttgggttgtgtagaactgacattacttaaagttt  
 aaccgaggaatgggagtgaggctctctcataccctattcagaactgacttttaacaataataaattaagttttaaataat  
 tTTTaaatgaattgagcaatggtgagttggagtc

**Core E<sub>μ</sub>**  
**aagatggccgatcagaaccagaacacctgcagcagctggcaggaagcaggtcatgtggcaaggcta**tttggggaagggga  
 aaataaaaccactaggtaaacttgtagctgtggtttgaagaagtgggtttgaaacactctgtccagccccaccaaaccg  
 aaagtccaggtgagcaaaaaccacctgggtaattttgcatttctaaataagttgag

3' MAR  
 gattcagccgaaactggagaggtcctcttttaacttattgagttcaaccttttaatttttagcttgagtagttctagttt  
 ccccaacttaagtttatcgacttcaaaatgtattttagaattcattttcaaaattaggttatgtaagaaattgaaggac  
 tttagtgcttttaatttctaataatatttagaaaacttcttaaaattactctattattcttccctctgattattgggtctc  
 cattcaattcttttccaataaccgaagcatttacagtgactttgttcatgatcttttttagttggtttgtttgccttac  
 tattaagactttgacattctggtcaaaacggcttcacaaatctttttcaagaccactttctgagtagttcatttttaggag  
 aaatacttttttttaaatgaatgcaattatctaga

Fig S2

### A Sense transcripts



### B Antisense transcripts

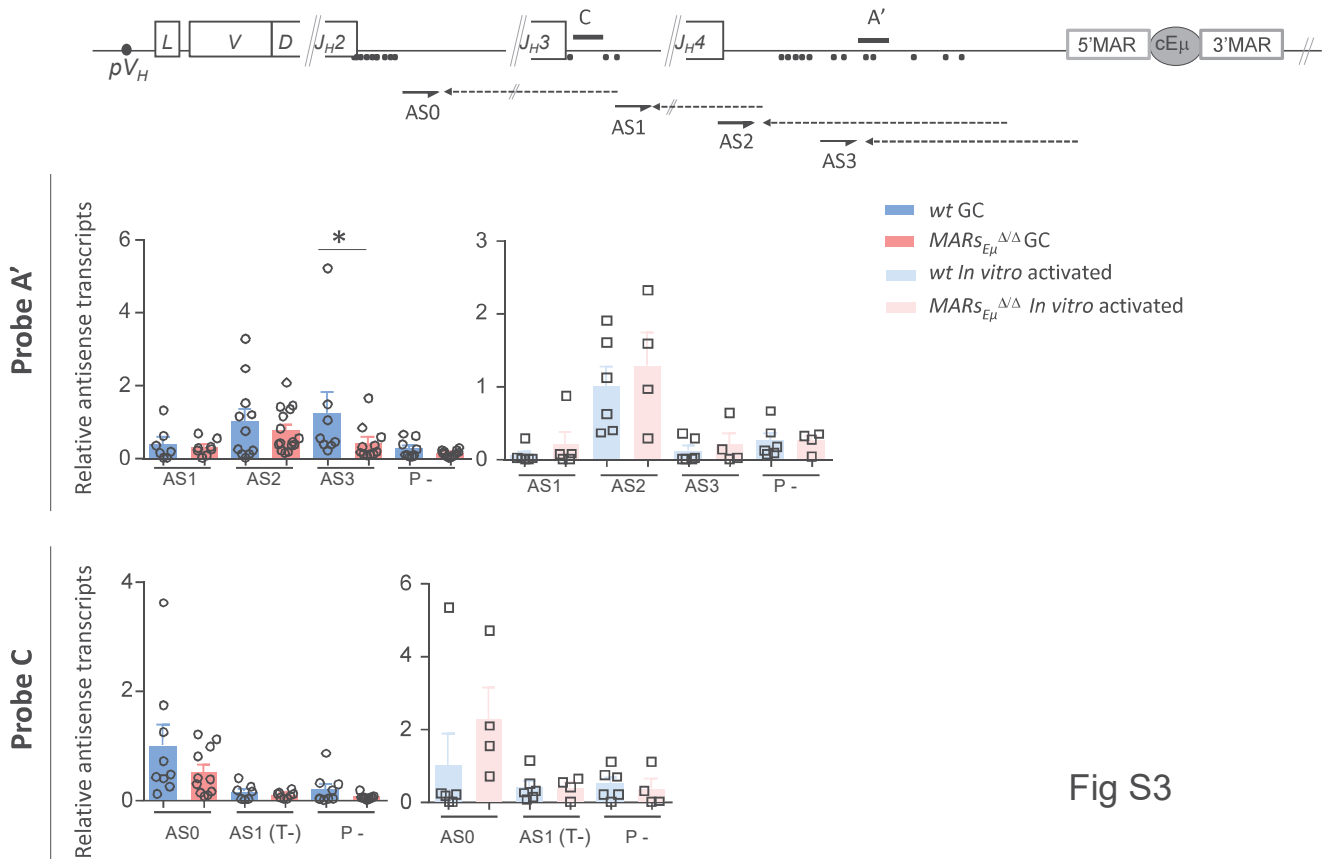


Fig S3

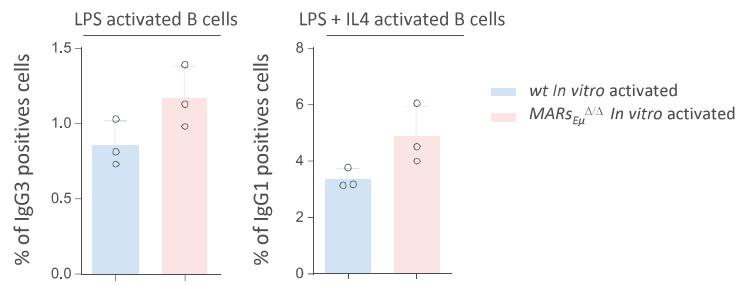


Figure S4

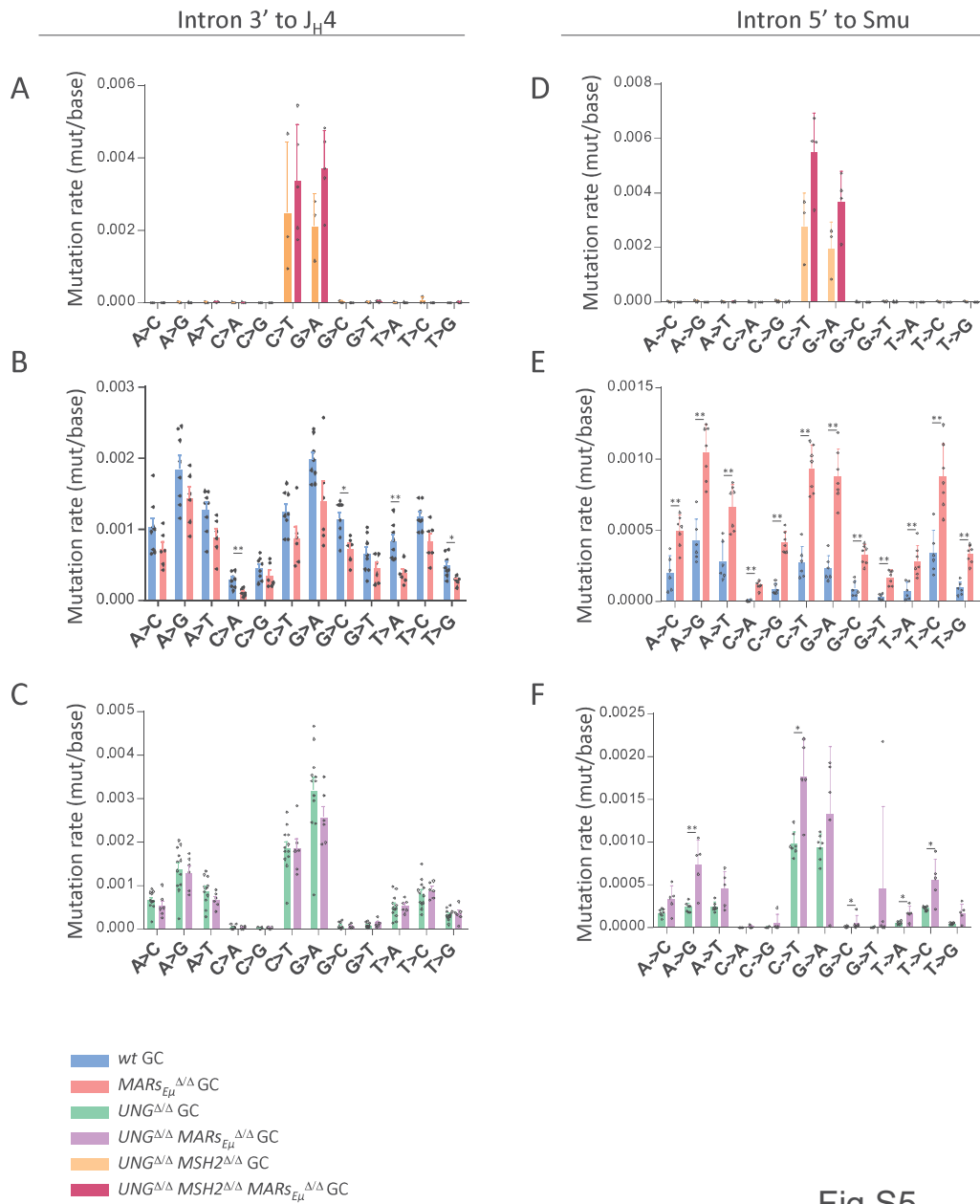


Fig S5



## **Annexe 2. Fam72a enforces error-prone DNA repair during antibody diversification**

Mélanie Rogier, Jacques Moritz, Isabelle Robert, Chloé Lescale, Vincent Heyer, Arthur Abello, Ophélie Martin, Katia Capitani, Morgane Thomas, Anne-Sophie Thomas-Claudepierre, Brice Laffleur, Florence Jouan, Eric Pinaud, Karin Tarte, Michel Cogné, Silvestro G. Conticello, Evi Soutoglou, Ludovic Deriano & Bernardo Reina-San-Martin.

Nature. 2021 Dec;600(7888):329-333. doi: 10.1038/s41586-021-04093-y. Epub 2021 Nov 24.





# *Fam72a* enforces error-prone DNA repair during antibody diversification

<https://doi.org/10.1038/s41586-021-04093-y>

Received: 17 December 2020

Accepted: 4 October 2021

Published online: 24 November 2021

 Check for updates

Mélanie Rogier<sup>1,2,3,4,12</sup>, Jacques Moritz<sup>1,2,3,4,12</sup>, Isabelle Robert<sup>1,2,3,4</sup>, Chloé Lescale<sup>5</sup>, Vincent Heyer<sup>1,2,3,4</sup>, Arthur Abello<sup>1,2,3,4</sup>, Ophélie Martin<sup>6</sup>, Katia Capitani<sup>7,8</sup>, Morgane Thomas<sup>9</sup>, Anne-Sophie Thomas-Claudepierre<sup>1,2,3,4</sup>, Brice Laffleur<sup>10</sup>, Florence Jouan<sup>10</sup>, Eric Pinaud<sup>9</sup>, Karin Tarte<sup>10</sup>, Michel Cogne<sup>9,10</sup>, Silvestro G. Conticello<sup>7,11</sup>, Evi Soutoglou<sup>6</sup>, Ludovic Deriano<sup>5</sup> & Bernardo Reina-San-Martin<sup>1,2,3,4</sup>✉

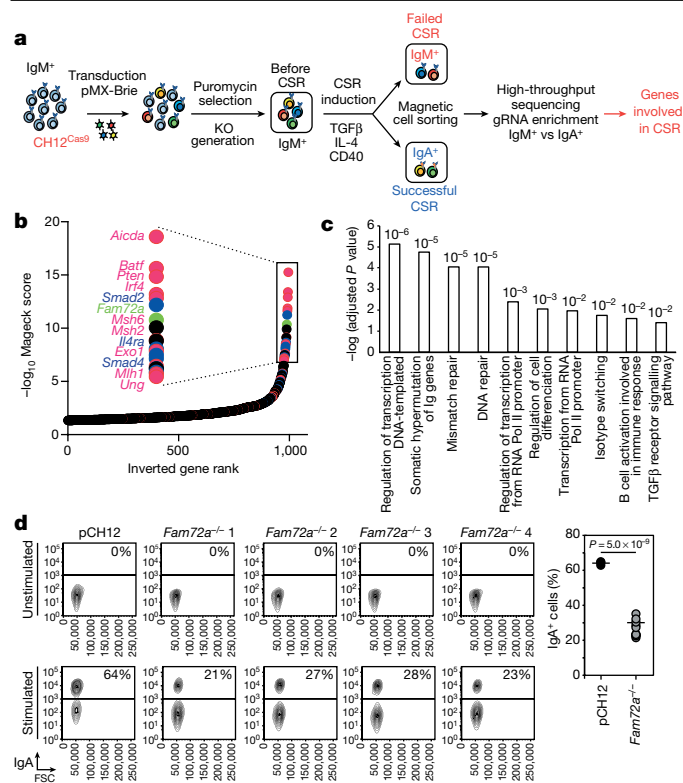
Efficient humoral responses rely on DNA damage, mutagenesis and error-prone DNA repair. Diversification of B cell receptors through somatic hypermutation and class-switch recombination are initiated by cytidine deamination in DNA mediated by activation-induced cytidine deaminase (AID)<sup>1</sup> and by the subsequent excision of the resulting uracils by uracil DNA glycosylase (UNG) and by mismatch repair proteins<sup>1–3</sup>. Although uracils arising in DNA are accurately repaired<sup>1–4</sup>, how these pathways are co-opted to generate mutations and double-strand DNA breaks in the context of somatic hypermutation and class-switch recombination is unknown<sup>1–3</sup>. Here we performed a genome-wide CRISPR–Cas9 knockout screen for genes involved in class-switch recombination and identified FAM72A, a protein that interacts with the nuclear isoform of UNG (UNG2)<sup>5</sup> and is overexpressed in several cancers<sup>5</sup>. We show that the FAM72A–UNG2 interaction controls the levels of UNG2 and that class-switch recombination is defective in *Fam72a*<sup>−/−</sup> B cells due to the upregulation of UNG2. Moreover, we show that somatic hypermutation is reduced in *Fam72a*<sup>−/−</sup> B cells and that its pattern is skewed upon upregulation of UNG2. Our results are consistent with a model in which FAM72A interacts with UNG2 to control its physiological level by triggering its degradation, regulating the level of uracil excision and thus the balance between error-prone and error-free DNA repair. Our findings have potential implications for tumorigenesis, as reduced levels of UNG2 mediated by overexpression of *Fam72a* would shift the balance towards mutagenic DNA repair, rendering cells more prone to acquire mutations.

During immune responses, mammalian B cells undergo somatic hypermutation (SHM) and class-switch recombination (CSR) to diversify the B cell receptor repertoire<sup>1–3</sup>. SHM modifies antibody affinity by mutating the variable region of both heavy (IgH) and light (IgL) chains<sup>1–3</sup>. CSR modulates antibody effector functions by replacing the isotype expressed from IgM/IgD to IgG, IgA or IgE, through a recombination event taking place at the *Igh* locus<sup>1–3</sup>. SHM and CSR are triggered by activation-induced cytidine deaminase (AID), which deaminates cytosines into uracils in DNA<sup>1–3</sup>. AID-generated U:G mismatches are processed mainly by the base-excision repair enzyme UNG<sup>3,6</sup> but also by mismatch repair<sup>7</sup> proteins to introduce mutations or double-strand DNA breaks (DSBs)<sup>1–3</sup>. During SHM, DNA replication over U:G mismatches generates transition mutations at G:C base pairs. The excision of uracil by UNG, followed by replication over abasic sites introduces

both transition and transversion mutations at C:G base pairs. In addition, single-stranded DNA surrounding the U:G mismatch or abasic site can be excised by both mismatch repair and base-excision repair to introduce transition and transversion mutations at A:T base pairs. During CSR, cytidine deamination and DNA cleavage on opposite strands generate DSBs<sup>1–3</sup>. While uracils in DNA are usually accurately repaired<sup>1–3</sup>, the mechanisms by which AID-induced uracils are instead processed in an error-prone way during SHM and CSR are largely unknown<sup>1–3</sup>.

To gain insight into the molecular mechanisms of CSR, we conducted a genome-wide CRISPR–Cas9 knockout screen (Fig. 1a) using the IgM<sup>+</sup> murine B cell line CH12 (ref. <sup>8</sup>), which we engineered to stably express Cas9 (CH12<sup>Cas9</sup>) Extended Data Fig. 1a) and the Brie<sup>9</sup> guide RNA (gRNA) library (pMX-mBrie) (Extended Data Fig. 1b). CH12<sup>Cas9</sup> cells were transduced, selected with puromycin and then stimulated to undergo CSR for

<sup>1</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch, France. <sup>2</sup>Institut National de la Santé et de la Recherche Médicale (INSERM), U1258, Illkirch, France. <sup>3</sup>Centre National de la Recherche Scientifique (CNRS), UMR7104, Illkirch, France. <sup>4</sup>Université de Strasbourg, Illkirch, France. <sup>5</sup>Genome Integrity, Immunity and Cancer Unit, Equipe Labellisée Ligue Contre Le Cancer, INSERM U1223, Institut Pasteur, Paris, France. <sup>6</sup>Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Brighton, UK. <sup>7</sup>Core Research Laboratory, ISPRO, Firenze, Italy. <sup>8</sup>Department of Medical Biotechnologies, University of Siena, Siena, Italy. <sup>9</sup>Centre National de la Recherche Scientifique (CNRS), UMR7276, Institut National de la Santé et de la Recherche Médicale (INSERM), UMR1262-Contrôle de la Réponse Immune B et Lymphoproliférations, Université de Limoges, Limoges, France. <sup>10</sup>Institut national de la santé et de la recherche médicale (INSERM), UMR1236, Université Rennes 1, Etablissement Français du Sang Bretagne, Rennes, France. <sup>11</sup>Institute of Clinical Physiology, National Research Council, Pisa, Italy. <sup>12</sup>These authors contributed equally: Mélanie Rogier, Jacques Moritz. ✉e-mail: reinab@igbmc.fr



**Fig. 1 | A genome-wide CRISPR-Cas9 knockout screen for genes involved in CSR identifies *Fam72a*.** **a**, Schematic overview. KO, knockout. **b**, Gene ranking ( $-\log_{10}$  Mageck score), plotted against the logarithmic value to the base 2 of the ratio of IgM over IgA read counts (LFC). Genes known to be involved in CSR are depicted in red; TGF $\beta$ , IL-4, CD40 and B cell receptor signalling pathways are depicted in blue; and genes with no known function in CSR are in black. **c**, Gene ontology analysis of genes identified in the screen. *P* values are indicated. RNA Pol II, RNA polymerase II. **d**, FACS analysis of IgA expression (versus forward scatter (FSC)) in wild-type (pCH12) and four independent *Fam72a*<sup>-/-</sup> CH12 cell clones stimulated or not with TGF $\beta$ , anti-CD40 antibody and IL-4. The percentage of IgA<sup>+</sup> cells from three independent experiments is shown on the right (one-tailed Student's *t*-test). The horizontal line denotes mean values.

72 h. Cells that failed to undergo CSR (which remained IgM<sup>+</sup>) or that succeed (which became IgA<sup>+</sup>) were sorted (Fig. 1a, Extended Data Fig. 1c). Genes, whose knockout affect the efficiency of CSR, were identified by scoring the enrichment of gRNA in the IgM<sup>+</sup> versus IgA<sup>+</sup> populations through deep sequencing<sup>10</sup> (Fig. 1a). The screen identified 654 candidate genes (Fig. 1b, Supplementary Table 1) and was successful as it identified genes known to be required for CSR (*Aicda* (which encodes AID), *Ung* and *Trp53bp1*), together with transcription factors that drive AID expression (*Batf*, *Pten* and *Irf4*), non-homologous end-joining factors (*Xrcc5* and *Fam35a*), mismatch repair proteins (*Msh2*, *Msh6*, *Pms1*, *Pms2* and *Mlh1*), members of the TGF $\beta$ , IL-4 and CD40 signalling pathways (*Tgfb1*, *Tgfb2*, *Il4ra*, *Il2rg* and *Cd40*) and other genes known to be implicated in CSR (*Mediator*, *Trim28* and *Zmynd8*, among other genes). Consistent with this, gene ontology analysis revealed a significant enrichment in pathways that are relevant for CSR and SHM (Fig. 1c). A number of genes with no known function in CSR were identified, including *Fam72a*.

We focused our analysis on *Fam72a*, whose expression is enhanced in B cells undergoing CSR (Extended Data Fig. 1d). *Fam72a* encodes a protein that binds to the nuclear isoform of UNG (UNG2) in a tryptophan 125-dependent manner<sup>5</sup> (Extended Data Fig. 2a, Supplementary Fig. 1) and is overexpressed in multiple cancers<sup>5</sup>. To determine whether *Fam72a* is required for CSR, we generated *Fam72a*<sup>-/-</sup> CH12 cells (Extended Data Fig. 2b, c). We found that the efficiency of CSR was

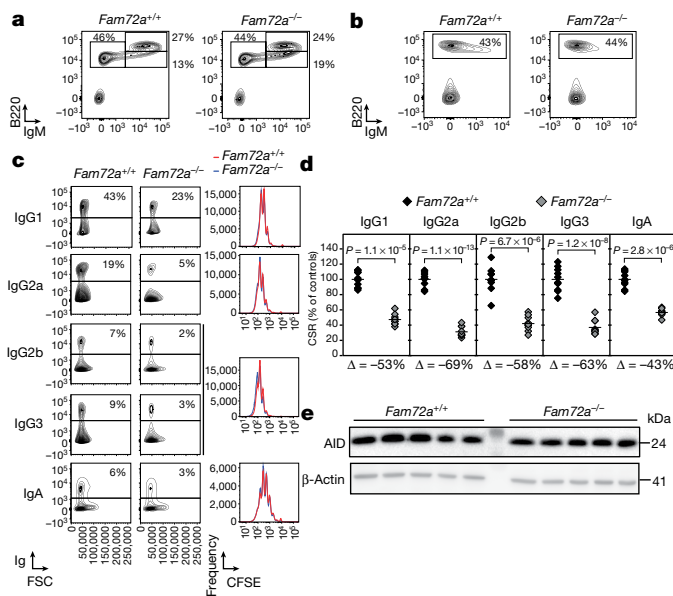
reduced by 50–60% in *Fam72a*<sup>-/-</sup> CH12 clones (Fig. 1d). The defect in CSR was not due to impaired expression of AID (Extended Data Fig. 2d), defects in proliferation (Extended Data Fig. 2e) or switch region transcription (Extended Data Fig. 2f). We concluded that *Fam72a* is required for efficient CSR in CH12 cells.

To investigate the physiological role of *Fam72a* in CSR and SHM in vivo, we obtained a *Fam72a*<sup>-/-</sup> mouse model<sup>11</sup> (Extended Data Fig. 3a). *Fam72a*<sup>-/-</sup> mice displayed normal development of B cells (Fig. 2a, b, Extended Data Fig. 3b, c), suggesting that FAM72A is dispensable for V(D)J recombination.

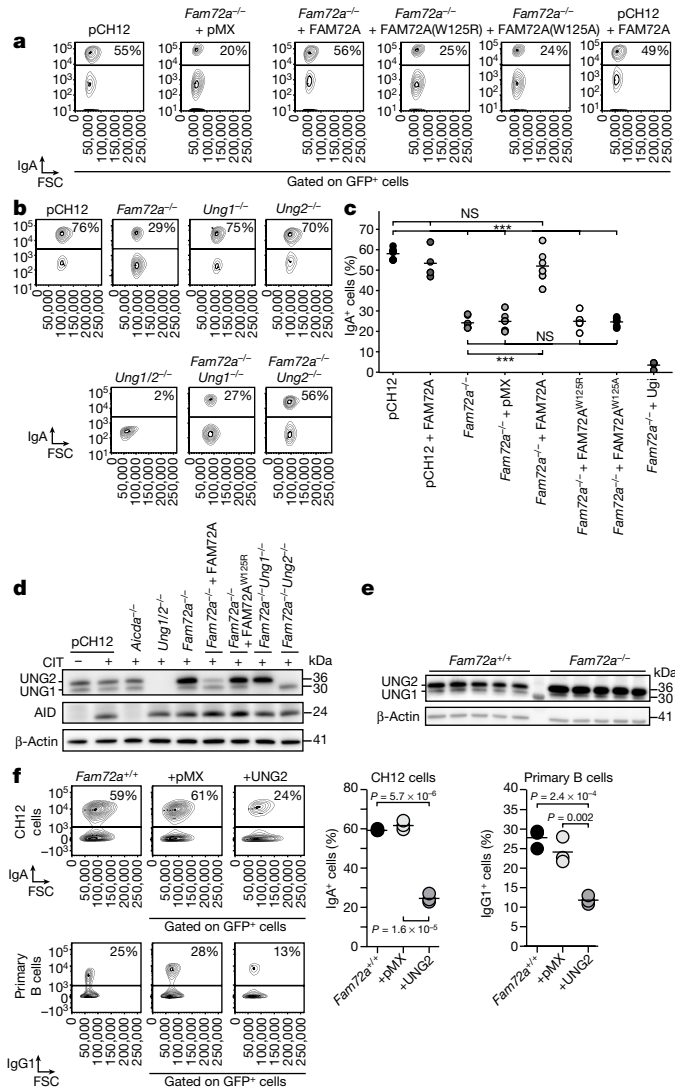
To determine whether *Fam72a*<sup>-/-</sup> B cells display a defect in CSR, we purified resting splenic B cells and induced CSR to different isotypes. *Fam72a*<sup>-/-</sup> B cells displayed a 40–70% reduction in CSR (Fig. 2c, d), which was not due to defects in proliferation (Fig. 2c), expression of AID (Fig. 2e) or switch region transcription (Extended Data Fig. 3d). We concluded that deficiency in *Fam72a* results in a B cell-intrinsic CSR defect in mouse primary B cells.

To investigate the functional relevance of the FAM72A-UNG2 interaction, we re-expressed FAM72A and the FAM72A(W125R) or FAM72A(W125A) UNG2-binding defective mutants<sup>5</sup> (Extended Data Fig. 2a) in *Fam72a*<sup>-/-</sup> CH12 B cells. Re-expression of FAM72A (Extended Data Fig. 4a) rescued CSR (Fig. 3a, c), whereas FAM72A(W125R) or FAM72A(W125A) did not (Fig. 3a, c). The residual CSR activity in *Fam72a*<sup>-/-</sup> B cells was dependent on the catalytic activity of UNG (Fig. 3c, Extended Data Fig. 4b). We concluded that the CSR defect observed is due to the lack of FAM72A, that the FAM72A-UNG2 interaction is of functional relevance for CSR and that the residual CSR activity is dependent on UNG.

The *Ung* gene encodes a mitochondrial (UNG1) and a nuclear (UNG2) isoform (Extended Data Fig. 4c), both of which are able to



**Fig. 2 | CSR is defective in B cells from *Fam72a*<sup>-/-</sup> mice.** **a**, **b**, FACS analysis of bone marrow (**a**) and spleen (**b**) cells from *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> mice. Staining antibodies and the percentage of cells within gates are indicated. **c**, FACS analysis of Ig expression in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> stimulated splenic B cells. The percentage of Ig-expressing cells is indicated. Representative plots from two experiments with five mice of each genotype are shown. Carboxyfluorescein succinimidyl ester (CFSE) dilution analysis is shown on the right. **d**, Percentage of CSR in *Fam72a*<sup>-/-</sup> B cells relative to *Fam72a*<sup>+/+</sup> B cells for the different isotypes from ten independent mice. The horizontal line denotes mean values. The percentage difference ( $\Delta$ ) is indicated below (one-tailed Student's *t*-test). **e**, Western blot analysis for AID and  $\beta$ -actin in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> stimulated splenic B cells. For gel source data, see Supplementary Fig. 1.



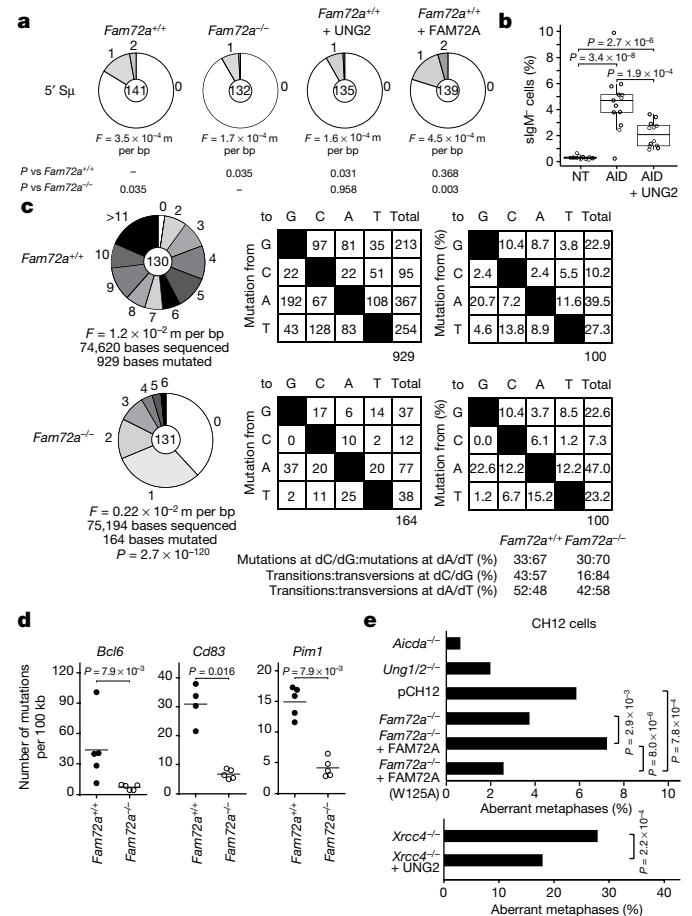
**Fig. 3 | The CSR defect in *Fam72a*<sup>-/-</sup> B cells is due to the specific upregulation of UNG2.** **a**, FACS analysis of IgA expression in stimulated *Fam72a*<sup>-/-</sup> CH12 cells transduced with an empty retrovirus (pMX) or expressing FAM72A, FAM72A(W125A) or FAM72A(W125R). The percentage of IgA-expressing cells is indicated. The parental cell line (pCH12) was included as a control. **b**, FACS analysis of IgA expression in stimulated CH12 cells with the indicated genotype. The percentage of IgA-expressing cells is indicated. **c**, The percentage of IgA<sup>+</sup> cells from three independent experiments from **a** and Extended Data Fig. 4b (two-tailed Student's *t*-test; for *P* values, see the 'Statistics and reproducibility' section in the Methods. The horizontal line denotes mean values. NS, not significant; \*\*\**P* < 0.001. **d**, Western blot analysis for UNG1, UNG2, AID and β-actin in CH12 cells with the indicated genotype transduced or not with a retrovirus expressing FAM72A or FAM72A(W125R) and stimulated with TGFβ, IL-4 and anti-CD40 antibody (CIT). **e**, Western blot analysis for UNG1, UNG2 and β-actin in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> stimulated splenic B cells. **f**, FACS analysis of Ig expression in wild-type CH12 cells and primary B cells transduced with an empty retrovirus (pMX) or expressing UNG2. Representative plots of three experiments are shown (one-tailed Student's *t*-test). The percentage of IgA-expressing cells is indicated and quantified on the right. The horizontal line denotes mean values. For gel source data, see Supplementary Fig. 1.

sustain CSR<sup>12</sup>. To examine the contribution of these isoforms relative to FAM72A, we generated *Ung1*<sup>-/-</sup>, *Ung2*<sup>-/-</sup>, *Ung1/2*<sup>-/-</sup>, *Fam72a*<sup>-/-</sup>*Ung1*<sup>-/-</sup> and *Fam72a*<sup>-/-</sup>*Ung2*<sup>-/-</sup> CH12 cell clones (Extended Data Fig. 4d). While *Ung1*<sup>-/-</sup> and *Ung2*<sup>-/-</sup> cells underwent CSR at wild-type levels (Fig. 3b, Extended Data Fig. 4e), CSR in *Ung1/2*<sup>-/-</sup> CH12 cells was completely

abolished (Fig. 3b, Extended Data Fig. 4e). While *Fam72a*<sup>-/-</sup>*Ung2*<sup>-/-</sup> rescued CSR, *Fam72a*<sup>-/-</sup>*Ung1*<sup>-/-</sup> did not (Fig. 3b, Extended Data Fig. 4e). This prompted us to analyse the protein levels of UNG1 and UNG2. We found that deficiency in *Fam72a* resulted in a significant and specific upregulation of the UNG2 isoform (Fig. 3d, Extended Data Fig. 4d), which accumulated on chromatin (Extended Data Fig. 4f). This phenotype was also observed in *Fam72a*<sup>-/-</sup> primary B cells (Fig. 3e) and was not due to changes at the mRNA level (Extended Data Fig. 4g). The upregulation of UNG2 was almost abolished by re-expressing FAM72A but not FAM72A(W125R) (Fig. 3d). In addition, we showed that UNG2 is degraded via a proteasome-dependent mechanism (Extended Data Fig. 4h, i). We concluded that FAM72A specifically regulates the protein level of the UNG2 isoform by controlling its degradation, although the exact mechanism remains to be determined. We therefore assume that the defect in CSR observed in *Fam72a*<sup>-/-</sup> and *Fam72a*<sup>-/-</sup>*Ung1*<sup>-/-</sup> cells is due to the upregulation of UNG2. Consistent with this, overexpression of murine UNG2 (mUNG2) in wild-type CH12 cells and primary splenic B cells (Extended Data Fig. 5a) suppressed CSR (Fig. 3f). This is in contrast to previous reports<sup>13–15</sup> and we speculate that this discrepancy might be related to differences in the expression levels of UNG2 and/or to the capacity of human UNG2 to reconstitute murine cells. Nevertheless, expression of the HIV-1 VPR protein, which specifically reduces the protein levels of UNG2 (ref. 16), rescued CSR (Extended Data Fig. 5b, c). Moreover, while expression of mUNG2 fused to an auxin-inducible degron<sup>17</sup> (mUNG2<sup>degron</sup>) only partially rescued CSR in *Ung1/2*<sup>-/-</sup> CH12 cells (Extended Data Fig. 5d), the addition of auxin completely restored CSR (Extended Data Fig. 5e). Together, it appears that the maintenance of the physiological levels of UNG2 is of critical importance for the efficiency of CSR. Our results suggest that FAM72A, by controlling the level of UNG2, influences the usage of error-prone versus error-free DNA repair in response to AID-induced uracils in DNA.

To determine whether deficiency in *Fam72a* has an influence in the frequency of AID-induced mutations, we sequenced the Sμ switch region in activated CH12 cells. The mutation frequency was reduced by 50% in *Fam72a*<sup>-/-</sup> CH12 cells (Fig. 4a). Consistent with this, overexpression of mUNG2 in CH12 cells also resulted in a 50% decrease in mutation frequency (Fig. 4a). Overexpression of FAM72A, which had no effect on CSR (Fig. 3a, c), did not lead to a statistically significant difference in mutation frequency (Fig. 4a). To further demonstrate that upregulation of UNG2 has a significant effect on SHM, we made use of an AID-deficient chicken DT40 B cell line, in which all of the pseudo V genes at the light chain locus have been deleted<sup>18</sup> and B cell receptor diversification happens exclusively via SHM when AID is re-expressed. As these cells are IgM<sup>+</sup>, SHM can be assessed by scoring the appearance of IgM<sup>-</sup> cells after AID expression. The percentage of IgM<sup>-</sup> cells induced by AID expression was reduced by 2.5-fold when UNG2 was co-expressed with AID (Fig. 4b). To further confirm that deficiency in *Fam72a* results in impaired SHM, we analysed the J<sub>H</sub>4 intron in germinal centre B cells isolated from the Peyer's patches of unimmunized mice (Extended Data Fig. 6a). The mutation frequency at the J<sub>H</sub>4 intron sequence was drastically reduced in *Fam72a*<sup>-/-</sup> B cells (Fig. 4c, Extended Data Fig. 6b, c). Consistent with upregulation of UNG2, we found that the percentage of transition mutations at C:G base pairs, which are generated by replication over U:G mismatches, were reduced from 43% in *Fam72a*<sup>+/+</sup> to 16% in *Fam72a*<sup>-/-</sup> B cells (Fig. 4c). Conversely, even though the mutation frequency was greatly reduced in *Fam72a*<sup>-/-</sup> cells, the relative percentage of transversion mutations at C:G base pairs, which are generated through replication over abasic sites, increased from 57% in *Fam72a*<sup>+/+</sup> to 84% in *Fam72a*<sup>-/-</sup> B cells (Fig. 4c). This phenotype is diametrically opposed to *Ung1/2*<sup>-/-</sup> B cells<sup>6</sup>, in which transition mutations at C:G base pairs reaches 95% and transversion mutations at C:G base pairs are suppressed to 5%<sup>6</sup>.

To determine whether the effect of *Fam72a* deficiency on SHM is specific to the *Ig* loci, we performed high-throughput sequencing<sup>19</sup> on three prominent AID off-targets. We found that the mutation frequency



**Fig. 4 | Deficiency in *Fam72a* shifts the balance towards error-free DNA repair.** **a**, Sμ mutation analysis in stimulated pCH12, *Fam72a*<sup>-/-</sup> and pCH12 cells transduced with a retrovirus expressing UNG2 or FAM72a. Pie charts depict the proportion of sequences with the indicated number of mutations. The mutation frequency (*F*) is shown below as mutations per base pairs sequenced (m per bp). The total number of sequences analysed is indicated in the centre (two-tailed Student's *t*-test). **b**, IgM expression in slgM<sup>+</sup> φV *Aicda/ICDA*<sup>-/-</sup> DT40 cells transduced with a retrovirus expressing AID and/or UNG2. Cells were gated on EGFP and/or mCherry. Twelve independent clones were analysed after 31 days of clonal expansion. For not transduced (NT), minimum = 0.15, maximum = 0.65, median = 0.285, quartile 1 = 0.225, quartile 3 = 0.35, upper whisker bound (UWB) = 0.54125 and lower whisker bound (LWB) = 0.03125. For AID, minimum = 2.44, maximum = 9.9, median = 4.79, quartile 1 = 3.805, quartile 3 = 5.31, UWB = 7.5675 and LWB = 1.5475. For AID + UNG2, minimum = 0.92, maximum = 3.64, median = 2.075, quartile 1 = 1.215, quartile 3 = 2.79, UWB = 5.1525 and LWB = 0 (two-tailed Student's *t*-test). **c**, Analysis of the J<sub>H</sub>4 intron mutation in germinal centre B cells from Peyer's patches depicted as in **a**. Mutation profiles are shown on the right. *P* values as determined by SHMtool ( $\chi^2$ ). Data are from five mice of each genotype. **d**, SHM analysis of AID off-targets (Mann-Whitney two-tailed test). **e**, The percentage of aberrant metaphases in stimulated CH12 cells of the indicated genotype harbouring *Igh* locus aberrations (*n* = 470 metaphases (*Aicda*<sup>-/-</sup>), *n* = 452 (*Ung1/2*<sup>-/-</sup>), *n* = 887 (pCH12), *n* = 720 (*Fam72a*<sup>-/-</sup>), *n* = 842 (*Fam72a*<sup>-/-</sup> + FAM72A) and *n* = 853 (*Fam72a*<sup>-/-</sup> + FAM72A(W125A))). The bottom panel shows the percentage of aberrant metaphases in *Xrcc4*<sup>-/-</sup> CH12 cells transduced or not with a retrovirus expressing UNG2 (*n* = 498 metaphases (*Xrcc4*<sup>-/-</sup>) and *n* = 502 (*Xrcc4*<sup>-/-</sup> + UNG2)); two independent cell clones were used, see also Supplementary Table 2; two-sided Fisher's exact test.

at these genes was reduced by 3–5-fold in *Fam72a*<sup>-/-</sup> B cells (Fig. 4d), with a clear inversion in the ratio between transition and transversion mutations at C:G base pairs (Extended Data Fig. 7a). The effect of *Fam72a* deficiency on SHM is therefore not *Ig* loci-specific, suggesting

that FAM72A-mediated regulation of UNG2 could have genome-wide repercussions.

These results support the hypothesis that in FAM72A-deficient B cells, where UNG2 is upregulated, AID-induced uracils are more efficiently excised from DNA by UNG2 and that error-free repair is enforced. Furthermore, enhanced excision of uracil would result in fewer U:G mismatches, rendering mismatch repair less effective in generating mutations at A:T base pairs surrounding the deaminated cytosines. Therefore, the level of UNG2 needs to be tightly controlled to trigger error-prone DNA repair, which is essential for antibody diversification.

The defect in CSR observed in *Fam72a*<sup>-/-</sup> B cells could be explained by enforced error-free repair due to the upregulation of UNG2, which in turn would lead to the generation of fewer DSBs. In wild-type B cells, a small fraction of AID-induced DSBs can be detected in the form of chromosome breaks or translocations<sup>20</sup>. To determine whether deficiency in *Fam72a* influences the level of DSBs, we quantified aberrant metaphases harbouring *Igh* locus-specific chromosomal breaks (Extended Data Fig. 7b) in activated *Aicda*<sup>-/-</sup>, *Ung1/2*<sup>-/-</sup>, *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> CH12 cells reconstituted (or not) with FAM72A or FAM72A(W125A). We found lower levels of aberrant metaphases in *Fam72a*<sup>-/-</sup> cells or *Fam72a*<sup>-/-</sup> cells reconstituted with FAM72A(W125A) than in wild-type and *Fam72a*<sup>-/-</sup> cells reconstituted with FAM72A (Fig. 4e, Extended Data Fig. 7b, Supplementary Table 2), suggesting that FAM72A-mediated UNG2 upregulation controls the level of AID-induced DSBs during CSR. Consistent with this, overexpression of UNG2 in *Xrcc4*<sup>-/-</sup> CH12 cells, which accumulate unrepaired DSBs<sup>21</sup>, resulted in a 40% decrease in the frequency of *Igh* DSBs (Fig. 4e). We concluded that the defect in CSR observed in the absence of *Fam72a* is probably due to the inefficient generation of AID-induced DSBs rather than to a perturbation of downstream DSB repair pathways.

Our results are consistent with a model (Extended Data Fig. 8) in which FAM72A interacts with UNG2 to control its physiological level by triggering its degradation. Consequently, deficiency in *Fam72a* leads to the specific upregulation of UNG2 and its accumulation on chromatin. This would enforce uracil excision, resulting in a reduction in the efficiency of SHM and CSR. It is possible that supraphysiological levels of UNG2 in the absence of FAM72A tilt the balance towards error-free base-excision repair by favouring the recruitment of polymerase β<sup>22–25</sup>. Although the question of how the choice between error-prone and error-free DNA repair is specifically established at *Ig* loci remains to be fully elucidated, we demonstrate here that FAM72A is an UNG2 regulator (and not an AID regulator) and an important player in antibody diversification.

A global and genome-wide shift in the balance from error-free to error-prone DNA repair in response to uracils in DNA has important implications in the onset of cancer. As *Fam72a* is overexpressed in several types of tumours and transformed cell lines<sup>5</sup> and as *Ung1/2*<sup>-/-</sup> mice develop B cell lymphomas<sup>26–28</sup>, it is possible that overexpression of *Fam72a* could suppress the levels of UNG2 leading to inefficient uracil excision, enforcing error-prone DNA repair and making these cells more susceptible to accumulate mutations and hence more prone to tumorigenesis.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-04093-y>.

1. Feng, Y., Seija, N., Di Noia, J. M. & Martin, A. AID in antibody diversification: there and back again. *Trends Immunol.* **41**, 586–600 (2020).
2. Methot, S. P. & Di Noia, J. M. Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv. Immunol.* **133**, 37–87 (2017).

3. Stratigopoulou, M., van Dam, T. P. & Guikema, J. E. J. Base excision repair in the immune system: small DNA lesions with big consequences. *Front. Immunol.* **11**, 1084 (2020).
4. Krokan, H. E. et al. Error-free versus mutagenic processing of genomic uracil—relevance to cancer. *DNA Repair* **19**, 38–47 (2014).
5. Guo, C. et al. Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase. *Cancer Res.* **68**, 6118–6126 (2008).
6. Rada, C. et al. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* **12**, 1748–1755 (2002).
7. Zanotti, K. J. & Gearhart, P. J. Antibody diversification caused by disrupted mismatch repair and promiscuous DNA polymerases. *DNA Repair* **38**, 110–116 (2016).
8. Nakamura, M. et al. High frequency class switching of an IgM<sup>+</sup> B lymphoma clone CH12F3 to IgA<sup>+</sup> cells. *Int. Immunol.* **8**, 193–201 (1996).
9. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
10. Joung, J. et al. Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).
11. KOMP strain detail sheet. *MMRRC* <[https://www.mmrrc.org/catalog/sds.php?mmrrc\\_id=47653](https://www.mmrrc.org/catalog/sds.php?mmrrc_id=47653)>
12. Sarno, A. et al. Uracil-DNA glycosylase UNG1 isoform variant supports class switch recombination and repairs nuclear genomic uracil. *Nucleic Acids Res.* **47**, 4569–4585 (2019).
13. Begum, N. A. et al. Uracil DNA glycosylase activity is dispensable for immunoglobulin class switch. *Science* **305**, 1160–1163 (2004).
14. Di Noia, J. M. et al. Dependence of antibody gene diversification on uracil excision. *J. Exp. Med.* **204**, 3209–3219 (2007).
15. Yousif, A. S., Stanlie, A., Mondal, S., Honjo, T. & Begum, N. A. Differential regulation of S-region hypermutation and class-switch recombination by noncanonical functions of uracil DNA glycosylase. *Proc. Natl Acad. Sci. USA* **111**, E1016–E1024 (2014).
16. Eldin, P. et al. Impact of HIV-1 Vpr manipulation of the DNA repair enzyme UNG2 on B lymphocyte class switch recombination. *J. Transl. Med.* **18**, 310 (2020).
17. Holland, A. J., Fachinetti, D., Han, J. S. & Cleveland, D. W. Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. *Proc. Natl Acad. Sci. USA* **109**, E3350–E3357 (2012).
18. Arakawa, H., Saribasak, H. & Buerstedde, J. M. Activation-induced cytidine deaminase initiates immunoglobulin gene conversion and hypermutation by a common intermediate. *PLoS Biol.* **2**, E179 (2004).
19. Martin, O. A. et al. 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 (2018).
20. Dev, H. et al. Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells. *Nat. Cell Biol.* **20**, 954–965 (2018).
21. Yan, C. T. et al. IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* **449**, 478–482 (2007).
22. Krokan, H. E. & Bjoras, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012583 (2013).
23. Krokan, H. E., Drablos, F. & Slupphaug, G. Uracil in DNA—occurrence, consequences and repair. *Oncogene* **21**, 8935–8948 (2002).
24. Wu, X. & Stavnezer, J. DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *J. Exp. Med.* **204**, 1677–1689 (2007).
25. Schrader, C. E., Guikema, J. E., Wu, X. & Stavnezer, J. The roles of APE1, APE2, DNA polymerase  $\beta$  and mismatch repair in creating S region DNA breaks during antibody class switch. *Phil. Trans. R. Soc. B* **364**, 645–652 (2009).
26. Nilsen, H. et al. Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene* **22**, 5381–5386 (2003).
27. Andersen, S. et al. Monoclonal B-cell hyperplasia and leukocyte imbalance precede development of B-cell malignancies in uracil-DNA glycosylase deficient mice. *DNA Repair* **4**, 1432–1441 (2005).
28. Nilsen, H., An, Q. & Lindahl, T. Mutation frequencies and AID activation state in B-cell lymphomas from Ung-deficient mice. *Oncogene* **24**, 3063–3066 (2005).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Methods

### CH12<sup>Cas9</sup> cell generation

CH12 cells were transduced with retroviral supernatants obtained by transfecting Bosc23 cells with a retrovirus (pMX-26) expressing a Cas9-mCherry-P2A-hygromycin cassette (Extended Data Fig. 1a). After 72 h of hygromycin selection (300 µg/ml; Sigma), mCherry<sup>+</sup> cells were sorted into 96-well plates using a cell-sorter FACS ARIA Fusion (Becton Dickinson) and cultured for 10 days with hygromycin (300 µg/ml; Sigma). Individual clones were then tested for their ability to express Cas9 by western blot (Extended Data Fig. 1a) and their ability to undergo efficient CSR after stimulation (Extended Data Fig. 1a). To test the functionality of Cas9, clones were transfected with gRNAs known to be effective (see Supplementary Table 3 for gRNA sequences). Three days later, PCR was performed to identify clones with Cas9-mediated deletions (Extended Data Fig. 1a). Clone #11, which had the highest Cas9 activity and displayed a CSR efficiency similar to the parental CH12 cell line, was selected to conduct the screen.

### Subcloning of the mBrie gRNA library

The mouse Brie CRISPR knockout pooled library (lentiCRISPRv2 backbone; a gift from D. Root and J. Doench (Addgene 73633)) was subcloned into the pMX-28 retroviral vector (Extended Data Fig. 1b) as previously described<sup>10</sup> to generate the pMX-mBrie gRNA library. In brief, the U6p-gRNA-scaffold cassette was amplified by PCR (see Supplementary Table 3 for primers), digested with BamHI and NotI restriction enzymes and ligated into pMX-28 (Extended Data Fig. 1b). Stbl4-electrocompetent cells (Thermo Fisher) were electroporated using Eporator Eppendorf (Program 1-1700V), plated on ampicillin agar plates and incubated at 37 °C for 15 h. Plasmid DNA was extracted and the U6-gRNA-scaffold cassette was amplified by PCR to add sequencing adaptors and barcodes (see Supplementary Table 3 for primers) and then analysed by deep sequencing to determine gRNA representation and library uniformity as previously described<sup>10</sup>.

### CRISPR-Cas9 knockout screen

CH12<sup>Cas9</sup> cells were transduced with the pMX-mBrie gRNA library in quadruplicate at a multiplicity of infection (MOI) of 0.3 and with a 300× coverage, as previously described<sup>10</sup>. Transduced cells were selected with puromycin (0.5 µg/ml; Sigma) and hygromycin (300 µg/ml; Sigma) for 15 days. Selected cells were then induced to undergo CSR with TGFβ (1 ng/ml; R&D Systems Europe), IL-4 (5 ng/ml; Peprotech) and an anti-CD40 antibody (200 ng/ml; eBioscience) for 72 h. IgM<sup>+</sup> and IgA<sup>+</sup> cells were sorted using anti-IgM and anti-IgA coupled magnetic beads (Miltenyi). Population purity was assessed by flow cytometry (Extended Data Fig. 1c) using a Fortessa flow cytometer (Becton Dickinson). Genomic DNA from 24 million cells was extracted using phenol/chloroform and subjected to PCR (one cycle at 95 °C for 2 min; 28–31 cycles at 95 °C for 15 s, 65 °C for 20 s, 72 °C for 30 s; and one cycle at 72 °C for 3 min) to amplify gRNA sequences using staggered primers having Illumina adaptors and barcodes (see Supplementary Table 3 for primers). Multiplexed samples were submitted to high-throughput sequencing (1 × 50 bp) on an Illumina HiSeq4000 sequencer at the GenomEast sequencing platform of Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC). gRNA counts were extracted from raw data using the PoolQ software from the Broad Institute. gRNA counts were normalized according as previously described<sup>10</sup>. Ranking of candidate genes was performed using MaGeCK<sup>29</sup>. Candidate genes ( $P < 0.05$ ) and having more than two effective gRNAs were selected. Gene ontology analysis of genes identified in the screen was performed with Enrichr database and statistical analysis was performed using two-tailed Fisher's exact test.

### Cell culture

Bosc23 cells were cultured in DMEM supplemented with glucose (4.5 g/l), 10% of heat-inactivated fetal calf serum (FCS),

penicillin-streptomycin (100 U/ml) and sodium pyruvate (1 mM). CH12 cells and primary B cells were cultured in RPMI supplemented with 10% FCS, HEPES (10 mM), penicillin-streptomycin (100 U/ml), sodium pyruvate (1 mM) and β-mercaptoethanol (50 µM). CH12 cells were cultured in the presence of cycloheximide (CHX; 10 µg/ml; Fisher Scientific) and MG132 (10 µM; R&D Systems) and auxin (IAA; 5–500 µM; Sigma-Aldrich). slgM<sup>+</sup> φV *Aicda*<sup>-/-</sup> DT40 cells were cultured in RPMI 1640 supplemented with 9% FCS, 1% chicken serum, β-mercaptoethanol (50 µM) and penicillin-streptomycin (100 U/ml) at 37 °C in 5% CO<sub>2</sub>. HEK293T cells were cultured in DMEM supplemented with 10% FCS and penicillin-streptomycin (100 U/ml). All cell lines used were mycoplasma negative.

### CSR assays

CH12 cells were cultured for 72 h in the presence of TGFβ (1 ng/ml; R&D Systems Europe), IL-4 (5 ng/ml; Peprotech) and an anti-CD40 antibody (200 ng/ml; eBioscience). Cells were then stained with an anti-IgA-PE antibody (1/500; Southern Biotech) to assess CSR by flow cytometry. Before analysis, DAPI was added to discriminate dead cells. Samples were analysed using a Fortessa flow cytometer (Becton Dickinson) and the FlowJo software.

### Primary B cell cultures

Splenic resting B cells were purified using anti-CD43 magnetic beads (Miltenyi), labelled with CFSE (Thermo Fisher Scientific) and cultured for 4 days with a combination of LPS (25 µg/ml; Sigma-Aldrich), IL-4 (25 µg/ml; Peprotech), anti-IgD-dextran (6 ng/ml; Fina Biosolutions), IFNγ (100 ng/ml; Peprotech), IL-5 (5 ng/ml; R&D Systems), TGFβ (3 ng/ml; R&D Systems) and retinoic acid (RA; 0.3 ng/ml; Sigma-Aldrich).

### RT-qPCR

RNA and cDNA were prepared using standard techniques. Quantitative PCR (qPCR) was performed in triplicates using Roche LightCycler 480 Probes Master mix Universal Probe Library (UPL) in combination with appropriate UPL probes (Supplementary Table 3) or QuantiTect SYBR Green Master mix (Qiagen). Transcript quantities were calculated relative to standard curves and normalized to *Hprt* or *Igb* (also known as *Cd79b*) mRNA.

### Retroviral transductions

CH12 cells were transduced with retroviral supernatants obtained by transfecting Bosc23 cells with an empty retrovirus (pMX-PIE; puromycin-IRES-EGFP) or expressing mFAM72A (pMX-FAM72A), mFAM72A(W125R) (pMX-mFAM72A(W125R)), mFAM72A(W125A) (pMX-FAM72A(W125A)), mUNG2 (pMX-mUNG2), Ugi (pMX-Ugi), HIV1-VPR (pMX-VPR), mUNG2<sup>Degron</sup> (pMX-mUNG2<sup>Degron</sup>) or TIR1 (pCH-TIR1). Transduced cells were then selected with puromycin (1 µg/ml) and/or hygromycin (300 µg/ml; Sigma) for 7 days and submitted to CSR assays. slgM<sup>+</sup> φV *Aicda*<sup>-/-</sup> DT40 cells were transduced with retroviral supernatants obtained by transfecting HEK293T cells with an empty retrovirus (pMX-PIE; puromycin-IRES-EGFP) or expressing mUNG2 (pMX-mUNG2) and with a retrovirus expressing mAID (pCH-8, mCherry and hygromycin).

### SHM analysis on DT40 cells

Transduced slgM<sup>+</sup> φV *Aicda*<sup>-/-</sup> DT40 cells were seeded (5 × 10<sup>5</sup> cells per well) in 24-well plates in 1 ml of medium. After 24 h, cells were transferred to 96-well plates in culture medium supplemented with puromycin (0.25 µg/ml) and hygromycin (300 µg/ml). After 48 h, cells were cultured under limiting dilution conditions to obtain single-cell clones. After 2 weeks, antibiotic-resistant colonies were picked and expanded for 2 additional weeks. The frequency of slgM-loss variants was then determined by flow cytometry after staining with anti-chicken IgM-PE antibody (Southern Biotech).

## Co-immunoprecipitation

Whole-cell extracts prepared from wild-type and *Fam72a*<sup>-/-</sup> CH12 cells transduced with a retrovirus expressing FAM72A-Flag or FAM72A(W125R)-Flag were immunoprecipitated with an anti-Flag antibody coupled to agarose beads (M2; Sigma) and analysed by western blot.

## Generation of CH12 knockout clones

CH12 cells were transfected by electroporation using the Neon transfection System (Thermo Fisher) with a plasmid expressing one or two gRNAs targeting a critical exon (see Supplementary Table 3 for gRNA sequences) and co-expressing the high-fidelity Cas9 nuclease (Cas9-HF1)<sup>30</sup> coupled to EGFP. Twenty-four hours after transfection, individual EGFP-positive cells were sorted into 96-well plates using a cell-sorter FACS ARIA Fusion (Becton Dickinson) and cultured for 10 days. Clones were then genotyped by PCR and sequencing.

## Sμ and JH4i somatic hypermutation analysis

For Sμ mutation analysis, genomic DNA was extracted from CH12 cells cultured for 3 days with TGFβ (1 ng/ml; R&D Systems Europe), IL-4 (5 ng/ml; Peprotech) and an anti-CD40 antibody (200 ng/ml; eBioscience). For JH4i mutation analysis, genomic DNA was extracted from B cells isolated from the Peyer's patches and sorted by flow cytometry using anti-B220-PE-Cy7 (eBiosciences), anti-Fas-PE (BD Pharmingen) and anti-GL7-pacific blue (BioLegend) antibodies. Sμ and JH4i sequences were amplified by PCR (one cycle at 98 °C for 30 s; 35 cycles at 98 °C for 10 s, 70 °C for 10 s and 72 °C for 30 s; and one cycle at 72 °C for 5 min) using the Q5 polymerase (New England BioLabs) and cloned into pUC57 using the MEGAWHOP<sup>31</sup> method. Inserts were sequenced (Sanger sequencing) using the M13 forward universal primer. Sequences were aligned with Lasergene (DNASTAR) and analysed with the SHMTool<sup>32</sup> server.

## SHM analysis of AID off-targets

Genomic regions within the *Bcl6*, *Cd83* and *Pim1* genes were amplified by PCR using previously described conditions and primers<sup>19</sup> from genomic DNA obtained from germinal centre B cells isolated from the Peyer's patches of unimmunized mice. Next-generation sequencing was performed according to the user guide Ion Xpress Plus gDNA Fragment Library Preparation (Life Technologies) from PCR products (100 ng each), and libraries were sequenced on an Ion torrent S5 system. Raw sequencing reads obtained in fastq format were analysed using the DeMinEr pipeline<sup>19</sup>. Samples from *Aicda*<sup>-/-</sup> B cells were used to determine the baseline mutation background and filter out mutations due to sequencing errors. Python scripts used in this study are available in a public repository under GNU General Public License v3.0 (<https://github.com/fboyerCRIBL/deminer>) and can also be provided upon demand from the corresponding author.

## Cloning

All gRNA/Cas9-HF1 plasmids were generated through golden gate cloning<sup>33</sup>. cDNAs were cloned into the pMX-PIE retrovirus through SLICE<sup>34</sup>.

## Mice

*Fam72a*<sup>-/-</sup> mice (C57BL6) were generated by the knockout mouse program (KOMP)<sup>11</sup>. Mice were bred under pathogen-free (SPF) conditions with standard dark/light cycles, ambient temperature and humidity. *Fam72a*<sup>-/-</sup> mice gave offspring at Mendelian ratios and showed no obvious deleterious phenotype. In all experiments 8–12-week-old age-matched littermates (males and females) were used. Animal work was performed under protocols approved by the IGBMC's ethics committee (ComEth/APAFIS#23104-2019112915476749) and by the French Ministry of Higher Education, Research and Innovation (D6721837).

## Igh FISH

Metaphases were prepared using standard procedures<sup>35</sup>. DNA FISH on metaphase spreads was performed as previously described<sup>35</sup> using BAC probes RP24-134G24 (5' *Igh* C) and RP24-386J17 (3' *Igh* V) and XCyting mouse chromosome 12 (Orange) paint from MetaSystems. Metaphases were imaged using a ZEISS AxioImager.Z2 microscope and the Metafer automated capture system (MetaSystems), and counted manually.

## Western blot analysis

Protein extracts were prepared using standard techniques. Proteins were separated by SDS-PAGE using gradient gels (4–12%; Invitrogen), transferred to PVDF membranes (Immobilon; Millipore) and analysed using anti-AID<sup>36</sup>, anti-β-actin (Sigma), anti-UNG1/2 (a gift from B. Kavli), anti-Nbs1 (a gift from A. Nussenzweig), anti-histone H3 (Abcam) and anti-Flag (Sigma) antibodies. Cell fractionation experiments were performed as previously described<sup>37</sup>.

## B cell development

Total splenic or bone marrow cells were labelled with suitable antibodies and analysed on an FACS-Fortessa flow cytometer (Becton Dickinson).

## Antibodies

Flag (1/1,000; M2, Sigma), IgA-biotin (1/500; Southern Biotech), UNG (a gift from B. Kavli; 2 μg/ml), AID (1/10,000; IGBMC), β-actin (1/200,000; Sigma), IgG1-biotin (1/500; BD Pharmingen), IgG2a-biotin (1/250; BioLegend), IgG2b-biotin (1/500; BioLegend), IgG3-biotin (1/250; BD Pharmingen), IgA-PE (1/500; Southern Biotech), GL7-pacific blue (1/200; BioLegend), B220-PE-Cy7 (1/100; eBiosciences), Fas (CD95)-PE (1/100; BD Pharmingen), streptavidin-PE (1/500; Jackson ImmunoResearch), histone H3 (1/10,000; Abcam), Nbs1 (a gift from A. Nussenzweig; 1/2,000), anti-CD40 (200 ng/ml; eBioscience) and anti-chicken IgM-PE (1/500; Southern Biotech).

## Statistics and reproducibility

For Fig. 3c, *P* values are pCH12 versus pCH12 + FAM72A = 0.45, pCH12 versus *Fam72a*<sup>-/-</sup> =  $2.2 \times 10^{-8}$ , pCH12 versus *Fam72a*<sup>-/-</sup> + pMX =  $1.9 \times 10^{-7}$ , pCH12 versus *Fam72a*<sup>-/-</sup> + FAM72A = 0.15, pCH12 versus *Fam72a*<sup>-/-</sup> + FAM72A(W125R) =  $8.85 \times 10^{-8}$ , pCH12 versus *Fam72a*<sup>-/-</sup> + FAM72A(W125A) =  $4.81 \times 10^{-9}$ , pCH12 versus *Fam72a*<sup>-/-</sup> + Ugi =  $7.31 \times 10^{-9}$ , *Fam72a*<sup>-/-</sup> versus *Fam72a*<sup>-/-</sup> + pMX = 0.77, *Fam72a*<sup>-/-</sup> versus *Fam72a*<sup>-/-</sup> + FAM72A =  $1.59 \times 10^{-5}$ , *Fam72a*<sup>-/-</sup> versus *Fam72a*<sup>-/-</sup> + FAM72A(W125R) = 0.73, *Fam72a*<sup>-/-</sup> versus *Fam72a*<sup>-/-</sup> + FAM72A(W125A) = 0.79, *Fam72a*<sup>-/-</sup> versus *Fam72a*<sup>-/-</sup> + Ugi =  $2.43 \times 10^{-6}$ .

For Extended Data Fig. 5c, *P* values are pCH12 versus *Fam72a*<sup>-/-</sup> =  $3 \times 10^{-4}$ ; pCH12 versus *Fam72a*<sup>-/-</sup> + pMX-VPR =  $4.3 \times 10^{-1}$ ; pCH12 versus *Fam72a*<sup>-/-</sup> + FAM72A =  $5.6 \times 10^{-1}$ ; pCH12 versus *Fam72a*<sup>-/-</sup> + FAM72A + pMX-VPR =  $4.9 \times 10^{-2}$ ; pCH12 versus *Fam72a*<sup>-/-</sup> + FAM72A(W125A) =  $1.8 \times 10^{-3}$ ; pCH12 versus *Fam72a*<sup>-/-</sup> + FAM72A(W125A) + pMX-VPR =  $6.4 \times 10^{-1}$ ; *Fam72a*<sup>-/-</sup> versus *Fam72a*<sup>-/-</sup> + pMX-VPR =  $1.1 \times 10^{-2}$ ; *Fam72a*<sup>-/-</sup> + FAM72A versus *Fam72a*<sup>-/-</sup> + FAM72A + pMX-VPR =  $2 \times 10^{-1}$ ; *Fam72a*<sup>-/-</sup> + FAM72A(W125A) versus *Fam72a*<sup>-/-</sup> + FAM72A(W125A) + pMX-VPR =  $4.3 \times 10^{-3}$ .

For Extended Data Fig. 5d, *P* values are pCH12 – IAA versus pCH12 + 500 μM IAA =  $3.8 \times 10^{-1}$ ; pCH12 – IAA versus *Fam72a*<sup>-/-</sup> – IAA =  $7.7 \times 10^{-6}$ ; pCH12 – IAA versus *Fam72a*<sup>-/-</sup> + 500 μM IAA =  $7.8 \times 10^{-6}$ ; pCH12 – IAA versus *Ung1/2*<sup>-/-</sup> – IAA =  $1.9 \times 10^{-6}$ ; pCH12 – IAA versus *Ung1/2*<sup>-/-</sup> + 500 μM IAA =  $1.9 \times 10^{-6}$ ; pCH12 – IAA versus *Ung1/2*<sup>-/-</sup> + UNG2<sup>Degron</sup> + TIR1 – IAA =  $3.9 \times 10^{-4}$ ; pCH12 – IAA versus *Ung1/2*<sup>-/-</sup> + UNG2<sup>Degron</sup> + TIR1 + 5 μM IAA =  $3.2 \times 10^{-2}$ ; pCH12 – IAA versus *Ung1/2*<sup>-/-</sup> + UNG2<sup>Degron</sup> + TIR1 + 500 μM IAA =  $6.8 \times 10^{-1}$ ; *Fam72a*<sup>-/-</sup> – IAA versus *Fam72a*<sup>-/-</sup> + 500 μM IAA =  $3.9 \times 10^{-2}$ ; *Ung1/2*<sup>-/-</sup> – IAA versus *Ung1/2*<sup>-/-</sup> + 500 μM IAA =  $8.9 \times 10^{-5}$ ; *Ung1/2*<sup>-/-</sup> + UNG2<sup>Degron</sup> + TIR1 – IAA versus *Ung1/2*<sup>-/-</sup> + UNG2<sup>Degron</sup> + TIR1 + 5 μM IAA =  $1.4 \times 10^{-5}$ ; *Ung1/2*<sup>-/-</sup> + UNG2<sup>Degron</sup> + TIR1 – IAA versus *Ung1/2*<sup>-/-</sup> + UNG2<sup>Degron</sup>



# Article

+TIRI + 500  $\mu$ M IAA =  $6.8 \times 10^{-5}$ ; *Ung1/2<sup>-/-</sup>* + UNG2<sup>Degron</sup> + TIRI + 5  $\mu$ M IAA versus *Ung1/2<sup>-/-</sup>* + UNG2<sup>Degron</sup> + TIRI + 500  $\mu$ M IAA =  $4.3 \times 10^{-3}$ .

For Fig. 3d, e, data presented are representative of three and two independent experiments, respectively.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The CRISPR–Cas9 screen high-throughput sequencing data are available at the Gene Expression Omnibus (GSE184145). Source data for Figs. 1–4 can be found in Supplementary Information. Source data are provided with this paper.

- Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
- Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
- Miyazaki, K. MEGAWHOP cloning: a method of creating random mutagenesis libraries via megaprimer PCR of whole plasmids. *Methods Enzymol.* **498**, 399–406 (2011).
- Maccarthy, T., Roa, S., Scharff, M. D. & Bergman, A. SHMTool: a webserver for comparative analysis of somatic hypermutation datasets. *DNA Repair* **8**, 137–141 (2009).
- Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE* **4**, e5553 (2009).
- Motohashi, K. A simple and efficient seamless DNA cloning method using SLiCE from *Escherichia coli* laboratory strains and its application to SLiP site-directed mutagenesis. *BMC Biotechnol.* **15**, 47 (2015).
- Lescale, C. et al. RAG2 and XLF/Cernunnos interplay reveals a novel role for the RAG complex in DNA repair. *Nat. Commun.* **7**, 10529 (2016).
- Jeevan-Raj, B. P. et al. Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *J. Exp. Med.* **208**, 1649–1660 (2011).
- Holden, P. & Horton, W. A. Crude subcellular fractionation of cultured mammalian cell lines. *BMC Res. Notes* **2**, 243 (2009).

**Acknowledgements** We thank A. Frey for help with gRNA library subcloning and the CRISPR–Cas9 screen; B. Kavli for the anti-UNG1/2 antibody; L. Brino for help with gRNA library

subcloning and CRISPR–Cas9 screen analysis; B. Jost, C. Thibault-Carpentier and D. Plassard for gRNA high-throughput sequencing on the GenomEast platform of IGBMC; D. Dembele, S. Sarnataro and N. Molina for help with Python scripts and gRNA extraction and counting; B. Heller for cell culture; C. Ebel and M. Philipps for cell sorting in the FlowCytometry Facility of IGBMC; M. Li for the GL-7 antibody, M. Selloum, D. Ali-Hadji and I. Gonçalves Da Cruz for animal care; W. Yu for generating *Xrcc4<sup>-/-</sup>* CH12 cells; and C. Goujon for the HIV1-VPR cDNA. J.M. and M.T. were supported by the Ministère de l'Enseignement Supérieur et de la Recherche, France, and the Fondation ARC. M.R. was supported by La Ligue Contre le Cancer. This study was supported by grants from the Fondation Recherche Médicale (Equipe FRM EQU201903007818 to B.R.-S.-M.), Institut National du Cancer (INCa\_13852 to L.D. and B.R.-S.-M.), Fondation ARC (ARCPJA32020060002061 to B.R.-S.-M.), the Ligue Nationale contre le Cancer (Equipe labellisée to L.D.) and by the grant ANR-10-LABX-0030-INRT, a French state fund managed by the Agence Nationale de la Recherche under the program Investissements d'Avenir labelled ANR-10-IDEX-0002-02.

**Author contributions** B.R.-S.-M. conceived and designed the study. M.R. generated CH12<sup>Cas9</sup> cells, subcloned the gRNA library and performed the CRISPR–Cas9 screen with I.R. Screen data were analysed by B.R.-S.-M. and M.R. All CH12 knockout cell lines were generated and characterized by M.R. and J.M. with the help of A.A. The majority of experiments in CH12 cells and in *Fam72a<sup>-/-</sup>* mice were performed by M.R. and J.M. *Igh* FISH experiments were performed and analysed by C.L. All constructs were cloned by M.R. and J.M., with help from V.H. V.H. performed the nuclear/chromatin fractionations and provided assistance for western blot analysis. M.R. and J.M. performed the SHM experiments in CH12 cells with the help of V.H. M.R. and J.M. performed SHM analysis in germinal centre B cells, with the help of A.A. O.M. generated the AID off-target high-throughput sequencing libraries, which were sequenced and analysed by M.T. and E.P. K.C. performed and analysed SHM in DT40 cells. A.-S.T.-C. performed mRNA sequencing and analysed the data. B.L. generated VPR lentiviral particles. F.J. performed immunohistochemistry on spleen sections. E.P., K.T., M.C., S.G.C. and E.S. oversaw experiments. M.R., J.M., E.S. and B.R.-S.-M. wrote the manuscript. M.R., J.M., C.L., E.S., L.D. and B.R.-S.-M. edited the manuscript. The overall research was directed by L.D. and B.R.-S.-M.

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

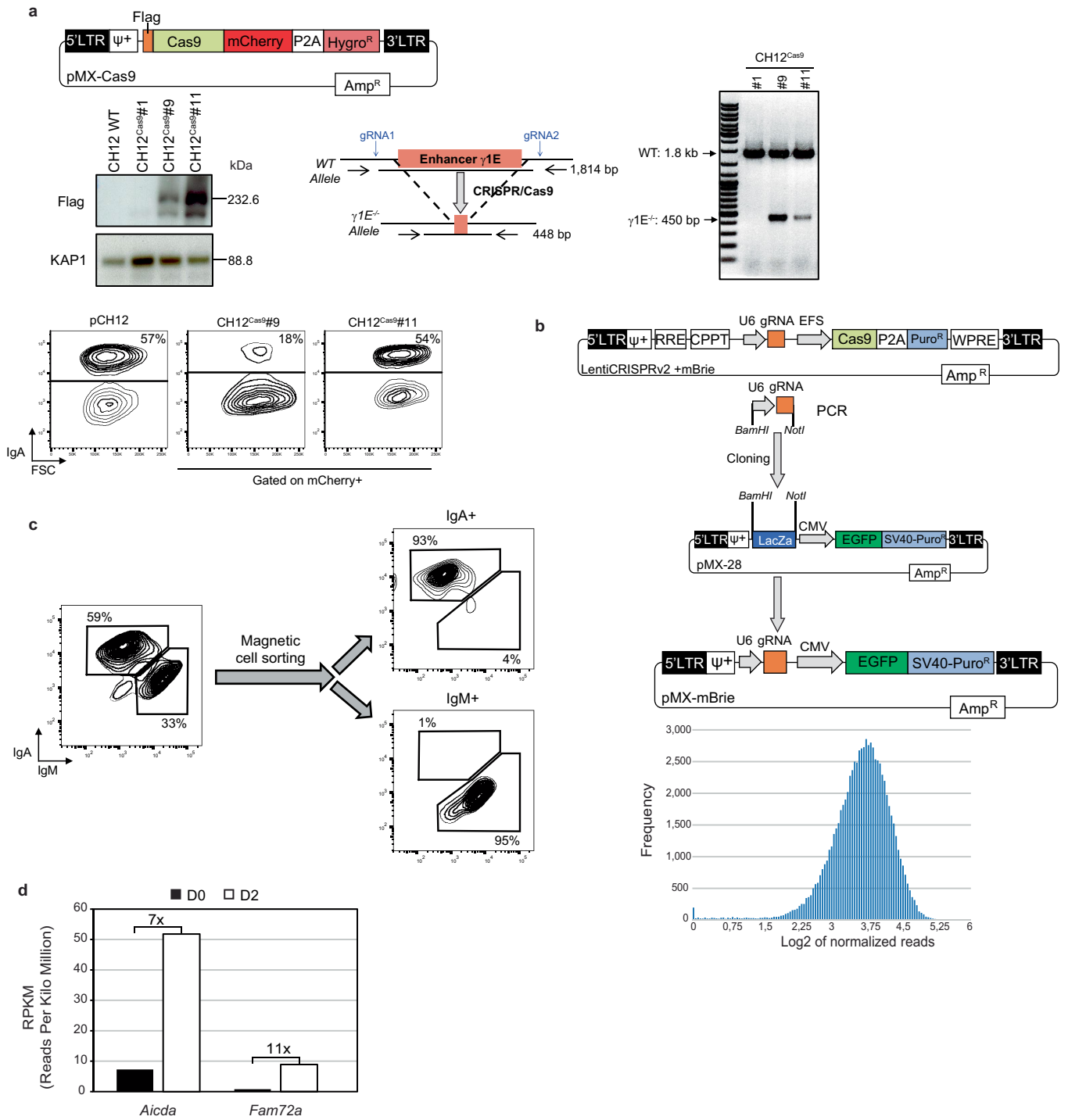
## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-04093-y>.

**Correspondence and requests for materials** should be addressed to Bernardo Reina-San-Martin.

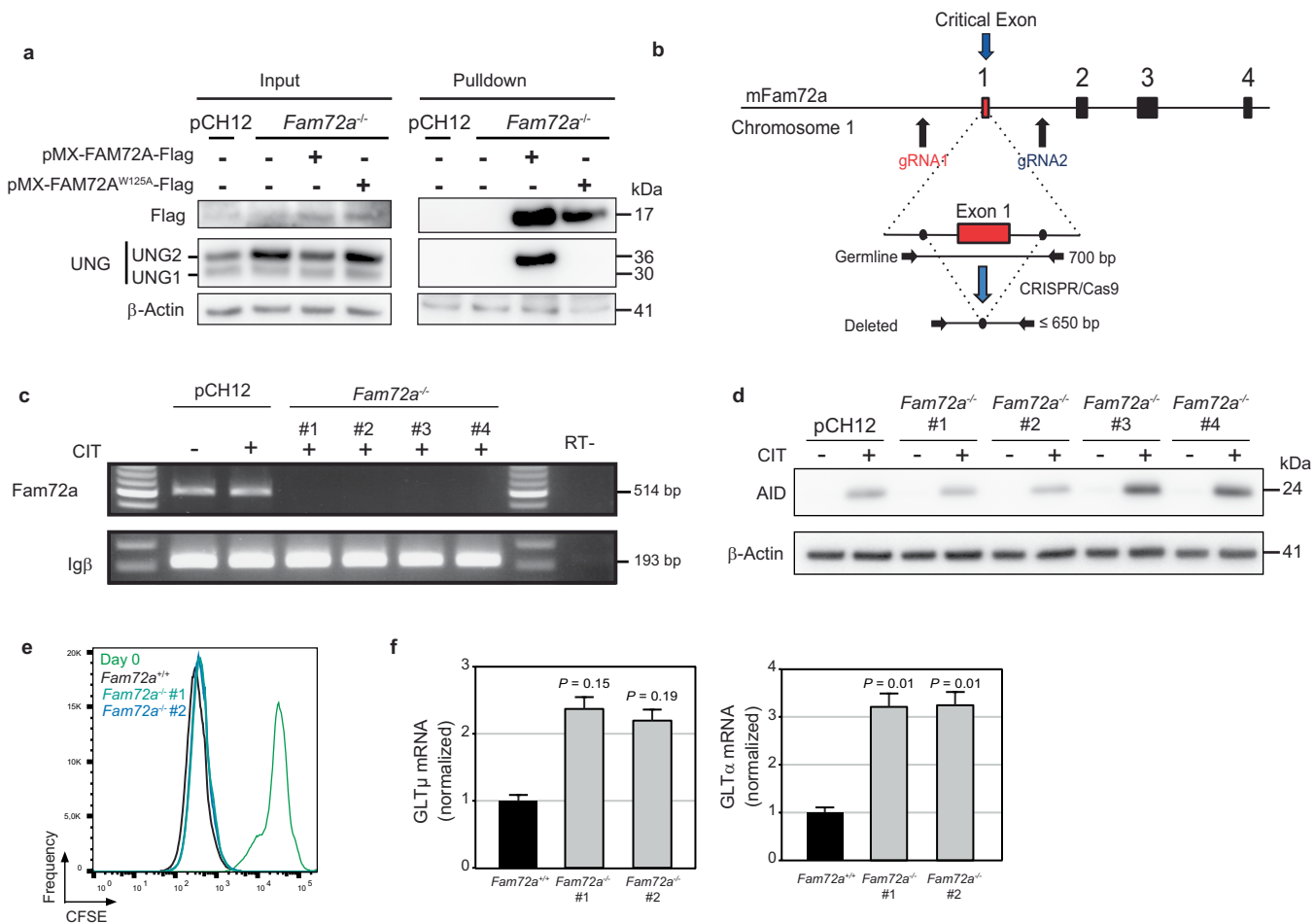
**Peer review information** Nature thanks Uttiya Basu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



**Extended Data Fig. 1 | Setting up the CRISPR/Cas9 genome-wide knock-out screen. a**, Generation and validation of Cas9-expressing (CH12<sup>Cas9</sup>) cells. Clones were verified by Western blot for Cas9 expression, by PCR for Cas9-induced deletion of the enhancer  $\gamma 1$  ( $\gamma 1E$ ) and by flow cytometry for CSR. Clone #11 was chosen for the screen. Blots and data of CSR assay are representative of 3 experiments. **b**, The U6-gRNA sequence was amplified by PCR from the mBrie library and subcloned into pMX-28 using BamHI and NotI

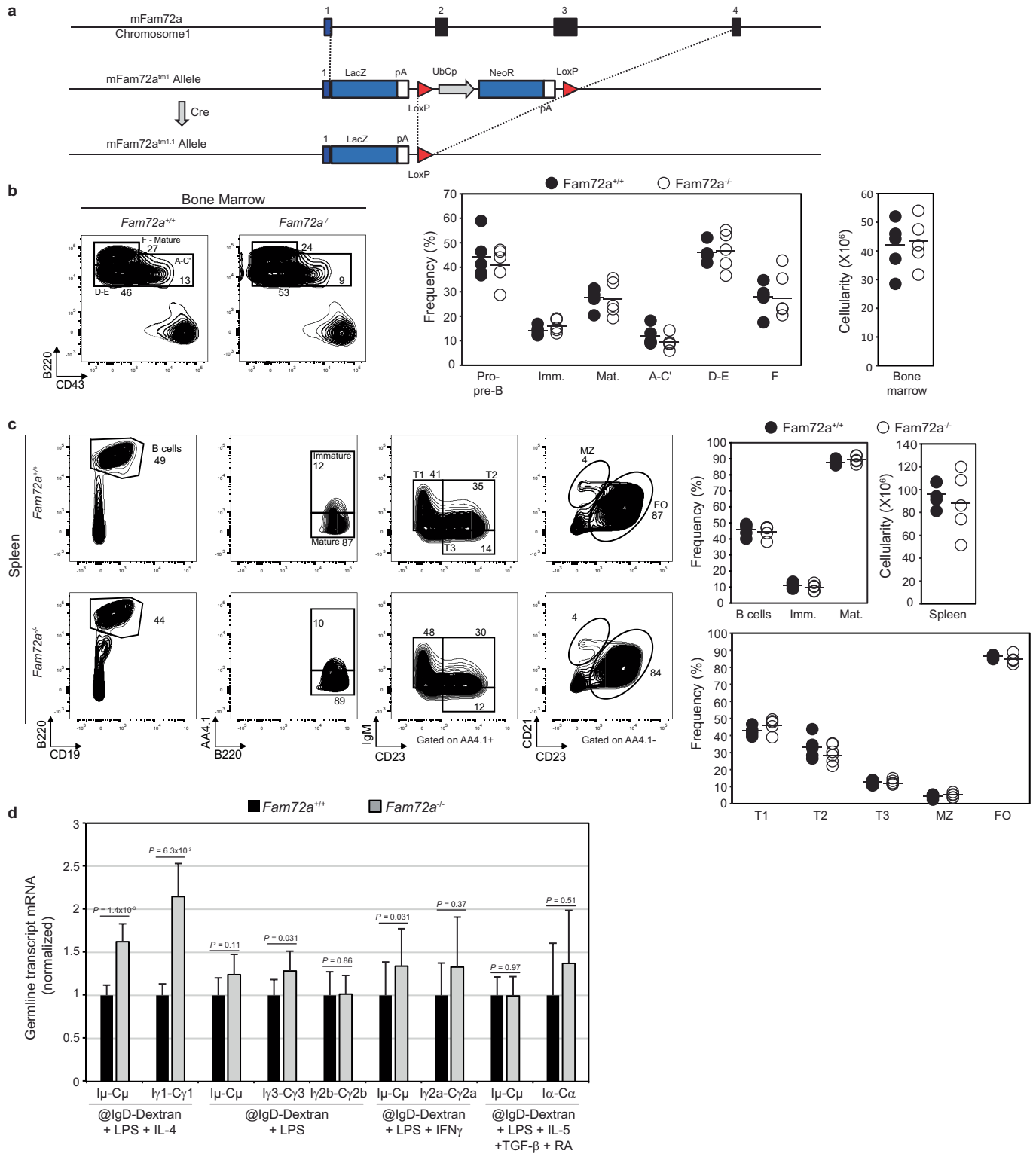
restriction enzymes to generate the pMX-mBrie gRNA library. gRNA representation was analyzed by high-throughput sequencing. **c**, Purity of IgM<sup>+</sup> and IgA<sup>+</sup> sorted populations was verified by flow cytometry. **d**, Number of reads at the *Aicda* and *Fam72a* genes in wild-type splenic B cells before and after 48h in culture with LPS and IL-4, as determined by mRNA-Seq. For gel source data, see Supplementary Fig. 1.



**Extended Data Fig. 2 | FAM72A interacts with UNG2 and *Fam72a*<sup>-/-</sup> CH12 cells have no defects in proliferation, AID expression or switch region transcription.**

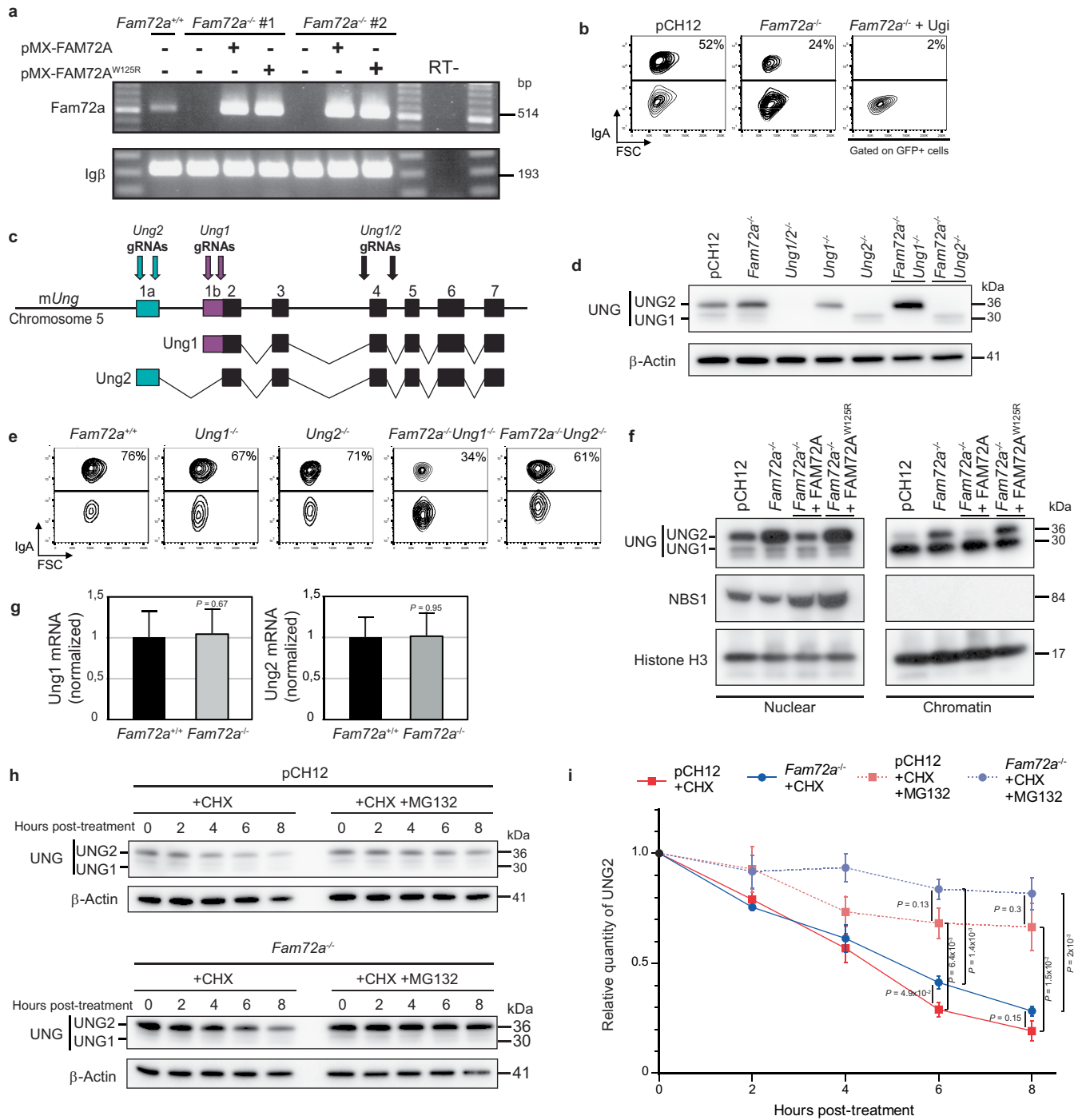
**a**, Western blot analysis for Flag, UNG (UNG1 and UNG2) and β-Actin in wildtype and *Fam72a*<sup>-/-</sup> CH12 cells transduced with a retrovirus expressing FAM72A-Flag or FAM72A<sup>W125A</sup>-Flag before (Input) or after immunoprecipitation with a Flag-specific antibody (Pulldown). **b**, Schematic representation of the murine *Fam72a* locus and location of the gRNAs used to generate *Fam72a*<sup>-/-</sup> clones using CRISPR/Cas9-HF1. **c**, RT-PCR analysis for *Fam72a* and *Igβ* in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> CH12 cells cultured or not for 3 days with TGFβ, anti-CD40 antibody and IL-4 (CIT). Data are representative of three experiments. **d**, Western blot analysis for AID and β-actin in wild-type (pCH12

and *Fam72a*<sup>-/-</sup> CH12 cells cultured or not for 3 days with TGFβ, anti-CD40 antibody and IL-4 (CIT). **e**, CFSE dye-dilution analysis by flow cytometry in *Fam72a*<sup>+/+</sup> and two independent *Fam72a*<sup>-/-</sup> CH12 cell clones cultured for 3 days with TGFβ, IL-4 and anti-CD40 antibody. Data are representative of three experiments. **f**, RT-qPCR analysis for GLTμ and GLTα in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> CH12 cells cultured for 3 days with TGFβ, IL-4 and anti-CD40 antibody. Triplicates were normalized to the abundance of *Igβ* and are expressed relative to *Fam72a*<sup>+/+</sup>, set as 1. Statistical significance was determined by a two-tailed Student's *t*-test. Data are presented as mean ± s.d. and are representative of 3 experiments. For gel source data, see Supplementary Fig. 1.



**Extended Data Fig. 3 | Robust B cell development and switch region transcription in *Fam72a*<sup>-/-</sup> mice.** **a**, Schematic representation of the murine *Fam72a* locus and strategy for the generation of *Fam72a*<sup>-/-</sup> mouse model. **b, c**, Flow cytometry and cellular analysis of B cell populations in the bone marrow (**b**) or the spleen (**c**) from *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> mice, using the indicated cell surface markers. The data are representative of 5 mice per genotype. Horizontal line: mean values. **d**, Real-time qPCR analysis for germline

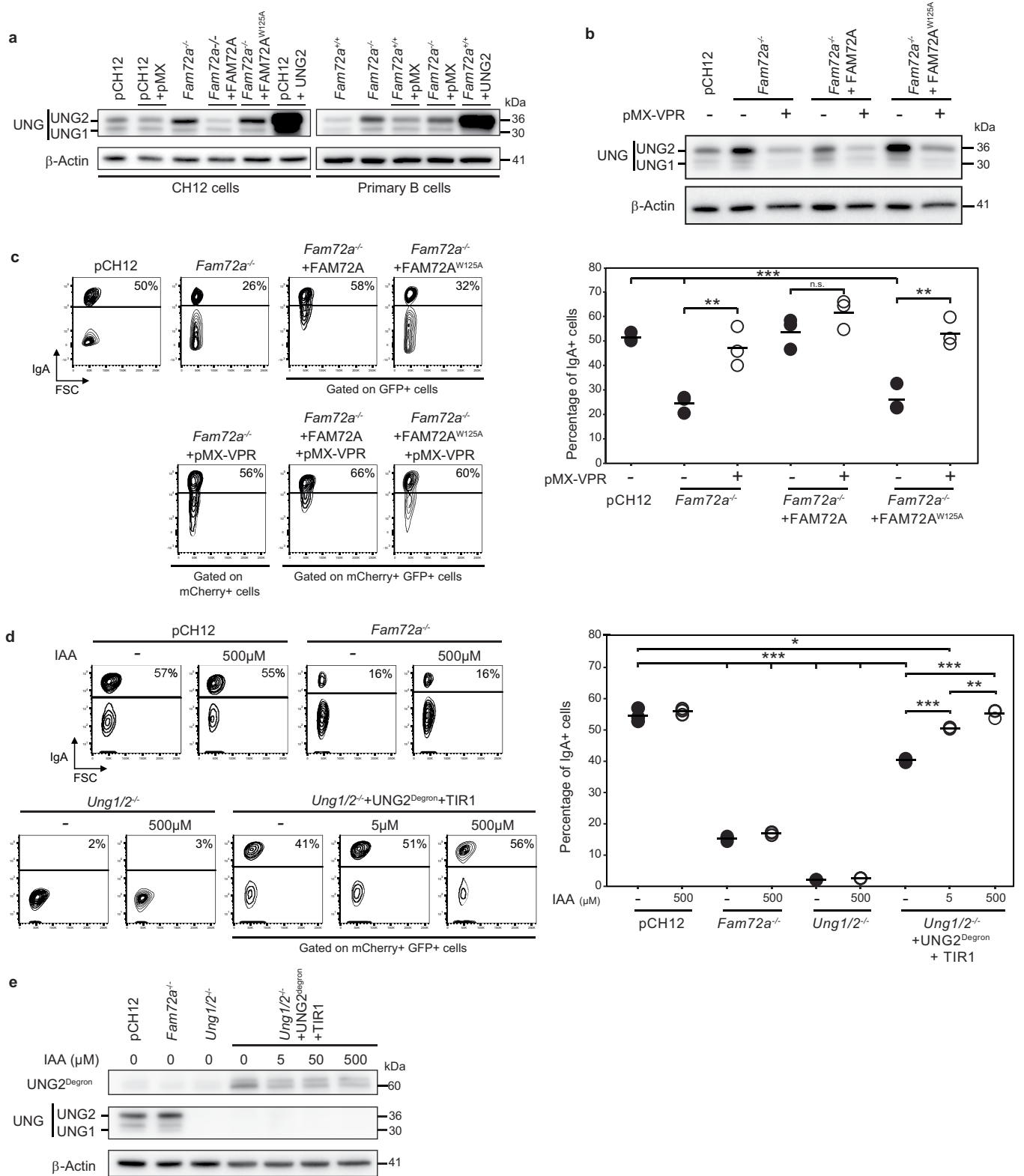
transcripts at donor (GLT $\mu$ ) and acceptor switch regions (GLT $\gamma$ 3, GLT $\gamma$ 1, GLT $\gamma$ 2b, GLT $\gamma$ 2a and GLT $\alpha$ ) in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> splenic B cells cultured for 96h as in **c**. Expression is normalized to Ig $\beta$  and is presented relative to expression in *Fam72a*<sup>+/+</sup> B cells, set as 1. Mean of three mice per genotype + SEM were calculated following the rules for error propagation while calculating a ratio. Statistical analysis was performed using two-tailed Student's *t*-test.



Extended Data Fig. 4 | See next page for caption.

**Extended Data Fig. 4 | FAM72A specifically controls UNG2 protein levels and its accession to chromatin via a proteasome-dependent mechanism. a,** RT-PCR analysis for *Fam72a* and *Igβ* in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> CH12 cells transduced (or not) with a retrovirus expressing FAM72A, FAM72A<sup>W125R</sup> or FAM72A<sup>W125A</sup>. Data are representative of three experiments. **b,** Flow cytometry analysis of IgA expression in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> CH12 cells expressing (or not) an UNG inhibitor (Ugi) and cultured for 3 days with TGFβ, anti-CD40 antibody and IL-4. Representative plots are shown. The percentage of IgA-expressing cells is indicated. **c,** Schematic representation of the murine *Ung* locus and location of the gRNAs targeting *Ung2* (exon 1a; blue), *Ung1* (exon 1b; purple) or *Ung1/2* (exon 4; black) used to generate *Ung1*<sup>-/-</sup>, *Ung2*<sup>-/-</sup>, *Ung1*<sup>-/-</sup> *Ung2*<sup>-/-</sup>, *Fam72a*<sup>-/-</sup> *Ung1*<sup>-/-</sup> and *Fam72a*<sup>-/-</sup> *Ung2*<sup>-/-</sup> CH12 cell clones using CRISPR/Cas9-HF1. Note, the gRNAs used to generate *Ung1/2*<sup>-/-</sup> CH12 cells target the same exon, which was deleted in *Ung1/2*-deficient mice. **d,** Western blot analysis for UNG (UNG1 and UNG2) and β-actin in wild-type, *Fam72a*<sup>-/-</sup>, *Ung1/2*<sup>-/-</sup>, *Ung1*<sup>-/-</sup>, *Ung2*<sup>-/-</sup>, *Fam72a*<sup>-/-</sup> *Ung1*<sup>-/-</sup> and *Fam72a*<sup>-/-</sup> *Ung2*<sup>-/-</sup> CH12 cells. Blots are representative of 3 experiments. **e,** Flow cytometry analysis of IgA expression in additional

independent wildtype, *Ung1*<sup>-/-</sup>, *Ung2*<sup>-/-</sup>, *Fam72a*<sup>-/-</sup> *Ung1*<sup>-/-</sup> and *Fam72a*<sup>-/-</sup> *Ung2*<sup>-/-</sup> CH12 cells cultured for 72 h with TGFβ, IL-4 and anti-CD40 antibody. **f,** Western blot analysis for UNG (UNG1 and UNG2), NBS1 and Histone H3 on nuclear and chromatin fractions prepared from CH12 cells (pCH12) and *Fam72a*<sup>-/-</sup> CH12 cells expressing FAM72A or FAM72A<sup>W125R</sup>. Representative blots of 2 experiments. **g,** RT-qPCR analysis for *Ung1*, *Ung2* and *Igβ* in *Fam72a*<sup>+/+</sup> and *Fam72a*<sup>-/-</sup> splenic B cells cultured for 4 days with LPS, IL-4 and anti-IgD-Dextran. Triplicates were normalized to the abundance of *Igβ* and are expressed relative to *Fam72a*<sup>+/+</sup> B cells, set as 1. Data are presented as mean of five mice per genotype ± s.d. Statistical analysis was performed using two-tailed Student's *t*-test. **h,** Western blot analysis for UNG (UNG1 and UNG2) and β-Actin in wild-type and *Fam72a*<sup>-/-</sup> CH12 cells cultured in the presence of cycloheximide (CHX) and MG132. **i,** Quantification of the protein level of UNG2, relative to time point zero from 3 independent experiments. Data are presented as mean ± s.d. Statistical analysis was performed using two-tailed Student's *t* test. For gel source data, see Supplementary Fig. 1.

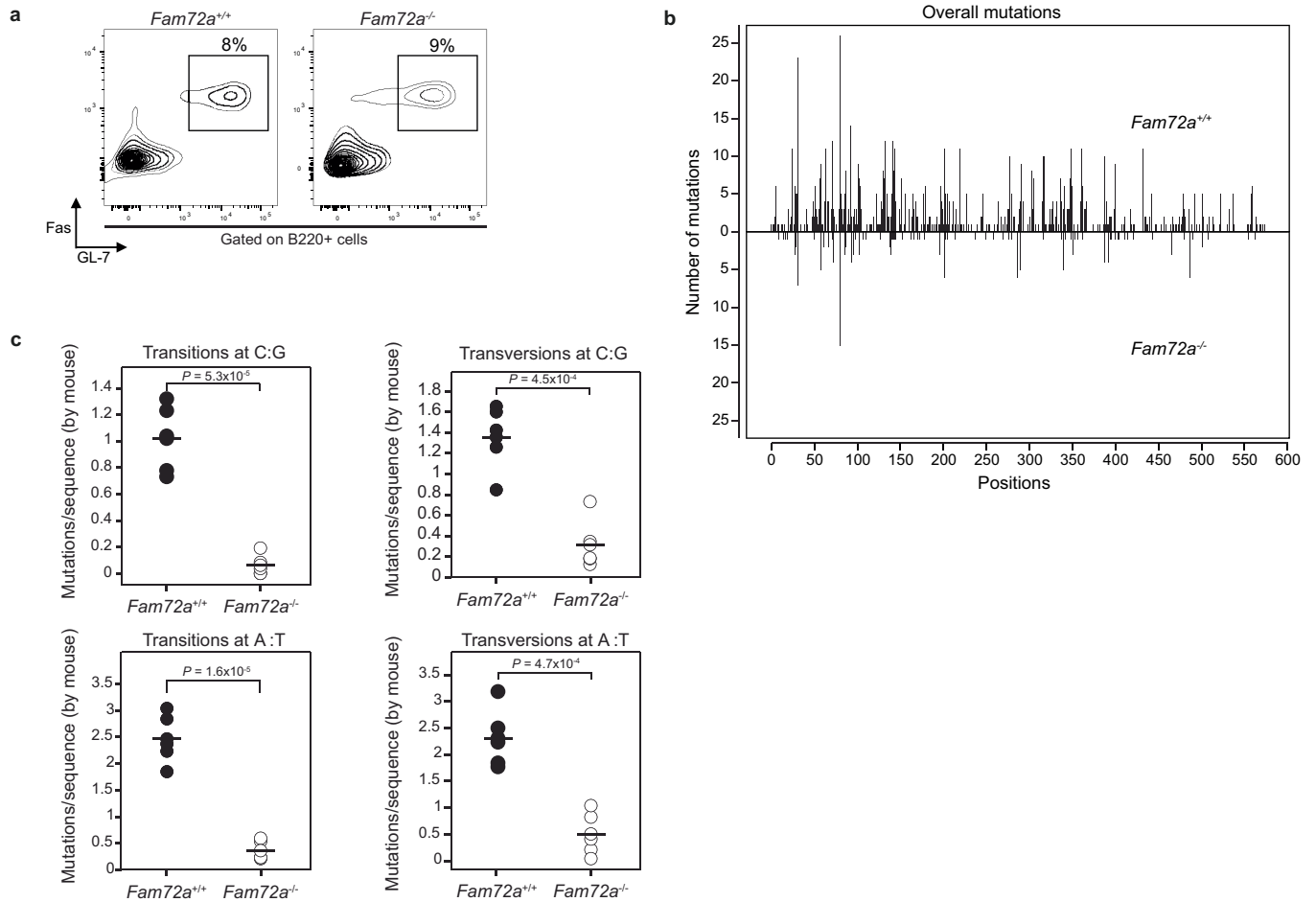


Extended Data Fig. 5 | See next page for caption.

**Extended Data Fig. 5 | UNG2 protein levels correlate with the efficiency of CSR.** **a**, Western blot analysis for UNG (UNG1 and UNG2) and  $\beta$ -Actin protein levels in wild-type CH12 cells and primary B cells transduced with a retrovirus expressing FAM72A, FAM72A<sup>W125A</sup> or UNG2 and cultured with TGF $\beta$ , IL-4 and anti-CD40 antibody or LPS + IL-4 + anti-IgD-Dextran, respectively. **b**, Western blot analysis for UNG (UNG1 and UNG2) and  $\beta$ -actin in *Fam72a*<sup>-/-</sup> CH12 cells transduced with a retrovirus expressing HIV1-VPR (pMX-VPR), FAM72A and/or FAM72A<sup>W125A</sup>. **c**, Flow cytometry analysis of IgA expression in *Fam72a*<sup>-/-</sup> CH12 cells transduced with a retrovirus expressing HIV1-VPR (pMX-VPR), FAM72A and/or FAM72A<sup>W125A</sup>. Representative plots of three experiments are shown and quantified on the right. P values were determined using two-tailed Student's *t*-test; see Statistics and Reproducibility section in Methods. n.s.: not significant;

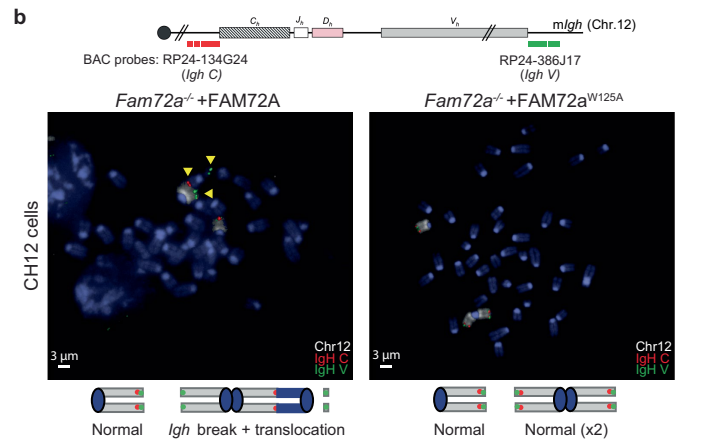
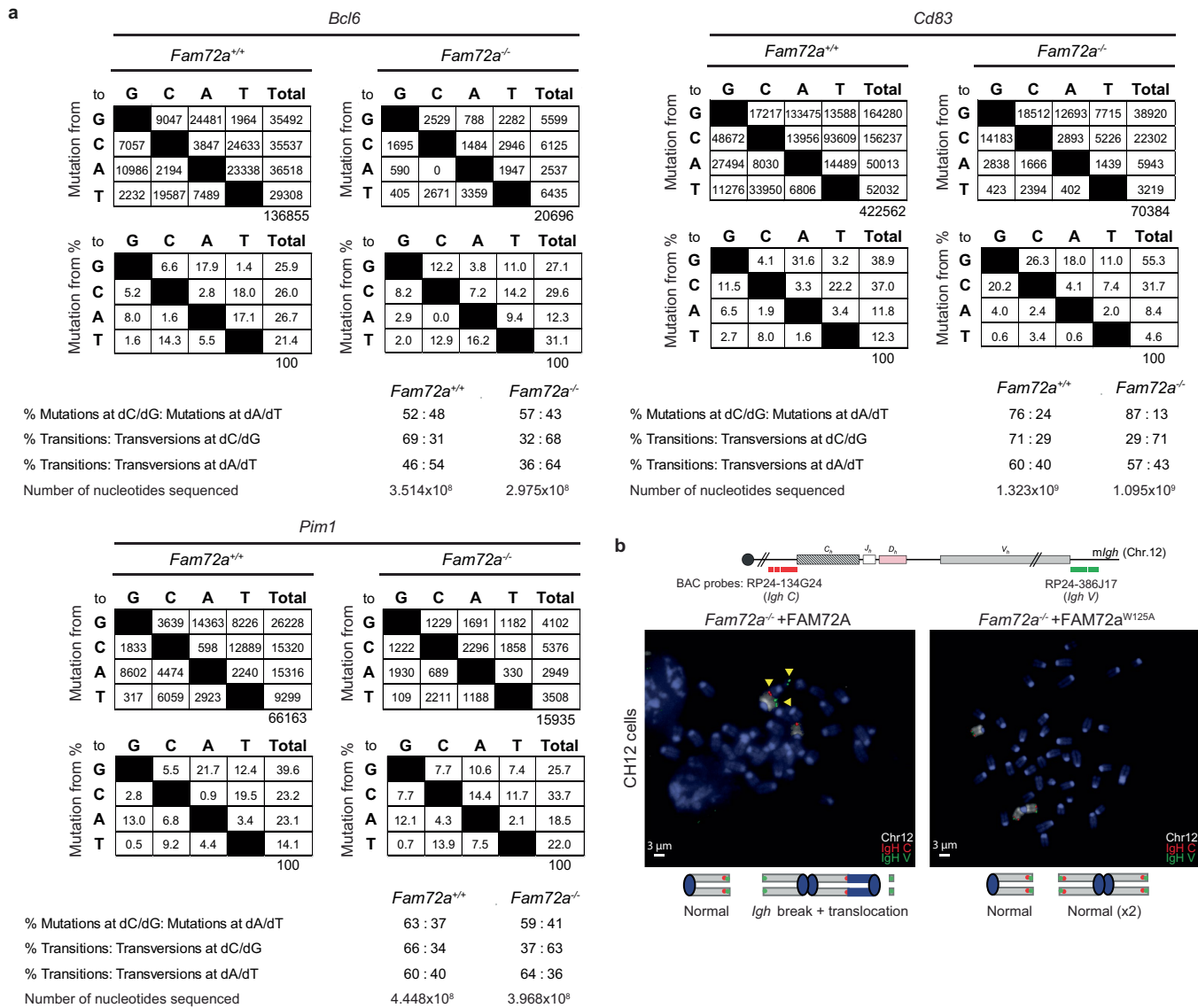
\*\*<0.01; \*\*\*<0.001. Horizontal line: mean values. **d**, Flow cytometry analysis of IgA expression in wild-type, *Fam72a*<sup>-/-</sup> and *Ung1/2*<sup>-/-</sup> CH12 cells transduced (or not) with a retrovirus expressing UNG2<sup>Degron</sup> and TIR1 and cultured for 72 h with TGF $\beta$ , IL-4 and anti-CD40 antibody in the presence or absence of auxin (IAA). Quantification is shown on the right. P-values were determined using two-tailed Student's *t*-test; see Statistics and Reproducibility section in Methods. \*<0.05; \*\*<0.005; \*\*\*<0.0005. Horizontal line: mean values. Data are from three experiments. **e**, Western blot analysis for UNG (UNG1 and UNG2) and  $\beta$ -Actin in wild-type, *Fam72a*<sup>-/-</sup>, and *Ung1/2*<sup>-/-</sup>, transduced (or not) with a retrovirus expressing UNG2<sup>Degron</sup> and Tir1 and cultured in the presence or absence of auxin (IAA). Blots **a**, **b** and **e** are representative of 3 independent experiments. For gel source data, see Supplementary Fig. 1.





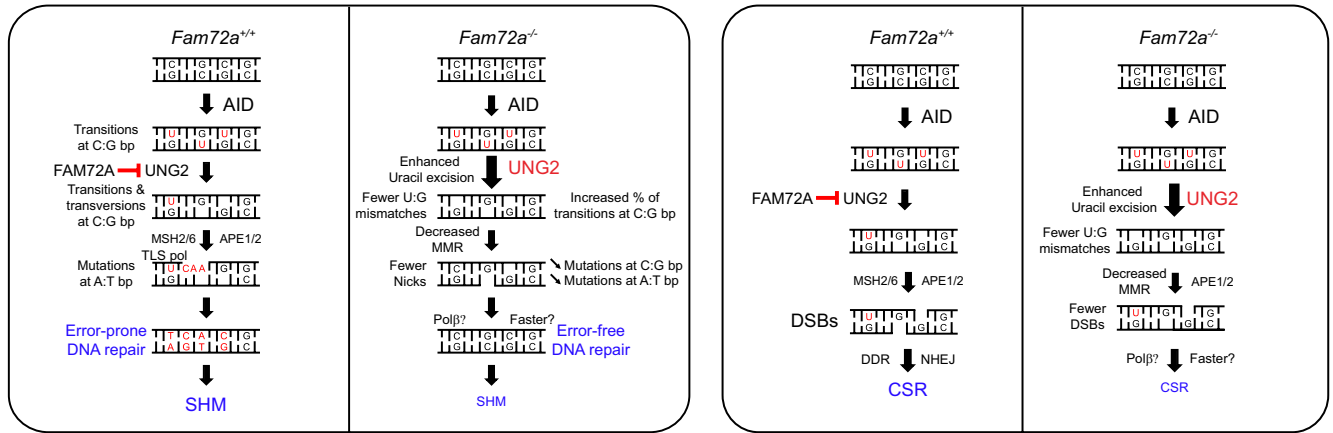
**Extended Data Fig. 6 | Somatic hypermutation at the J<sub>H</sub>4 intron (J<sub>H</sub>4i). a,** Flow cytometry analysis of germinal center B cells (B220+ Fas+ GL-7+) isolated from the Peyer's patches of unimmunized *Fam72a<sup>+/+</sup>* and *Fam72a<sup>-/-</sup>* mice. Plots are gated on B220+ cells. **b,** Distribution of mutations at the J<sub>H</sub>4 intron (J<sub>H</sub>4i) in *Fam72a<sup>+/+</sup>* (top) and *Fam72a<sup>-/-</sup>* (bottom) sequences obtained from germinal center B cells (B220+ Fas+ GL-7+) isolated from the Peyer's patches (PPs) of

unimmunized *Fam72a<sup>+/+</sup>* and *Fam72a<sup>-/-</sup>* mice. **c,** Mean frequencies (at JH4i) of transition and transversion mutations at G/C and A/T base pairs, per sequence for each individual mouse analyzed. Data come from 5 mice of each genotype. p-value was determined using two-tailed Student's *t*-test. Horizontal line: mean values.



**Extended Data Fig. 7 | Somatic hypermutation at AID off-targets and *Igh* FISH strategy. a**, Tables depicting the mutation profiles observed for *Bcl6*, *Cd83* and *Pim1* in germinal center B cells (B220+ Fas+ GL-7+) isolated from the Peyer's patches (PPs) of unimmunized *Fam72a<sup>+/+</sup>* and *Fam72a<sup>-/-</sup>* mice. The respective number of nucleotides sequenced for each gene is also indicated. Data are from five mice per genotype. **b**, Schematic representation of the *Igh*

locus with positions of the BACs used for generation of DNA FISH probes. Lower panel: DNA-FISH on representative metaphases from day 2 stimulated *Fam72a<sup>-/-</sup>* cells complemented with FAM72A or FAM72A<sup>W125A</sup> using the *Igh* C BAC probe (red) combined with *Igh* V BAC probe (green) and chromosome 12 paint (white). Yellow arrowheads point to broken or translocated chromosome 12.



**Extended Data Fig. 8 | Working model for SHM and CSR in the presence or absence of FAM72A.** Immunoglobulin genes are diversified through Somatic Hypermutation (SHM) and Class Switch Recombination (CSR). They are both initiated by the deamination of cytosines in DNA induced by Activation Induced Cytidine Deaminase (AID). The resulting uracils are processed, mainly by the nuclear isoform of Uracil DNA Glycosylase (UNG2), and by proteins of the Base Excision Repair (BER) and Mismatch Repair (MMR) pathways to introduce mutations or double-stranded DNA breaks (DSBs) during SHM (left panel) or

CSR (right panel). FAM72A interacts with UNG2 to control its physiological level by triggering its degradation and to enforce error-prone DNA repair. Consequently, deficiency in *Fam72a* leads to the specific upregulation of UNG2 and its accumulation on chromatin. This would enhance uracil excision, resulting in a reduction in the efficiency of SHM and CSR and enforced error-free DNA repair. Therefore, FAM72A controls the balance between error-prone and error-free DNA repair during antibody diversification.

### **Annexe 3. Continuous MYD88 activation is associated with expansion and then transformation of IgM differentiating plasma cells**

Catherine Ouk, Lilian Roland, Nathalie Gachard, Stephanie Poulain, Christelle Oblet, David Rizzo, Alexis Saintamand, Quentin Lemasson, Claire Carrion, Morgane Thomas, Karl Balabanian, Marion Espeli, Marie Parrens, Isabelle Soubeyran, Melanie Boulin, Nathalie Faumont, Jean Feuillard and Christelle Vincent-Fabert

Front Immunol, 2021 May 4;12:641692. doi: 10.3389/fimmu.2021.641692. eCollection 2021.





# Continuous MYD88 Activation Is Associated With Expansion and Then Transformation of IgM Differentiating Plasma Cells

## OPEN ACCESS

### Edited by:

Pablo Engel,  
University of Barcelona, Spain

### Reviewed by:

Hiromi Kubagawa,  
German Rheumatism Research  
Center (DRFZ), Germany  
Rachel Maurie Gerstein,  
University of Massachusetts Medical  
School, United States

### \*Correspondence:

Christelle Vincent-Fabert  
christelle.vincent-fabert@unilim.fr  
Jean Feuillard  
jean.feuilleard@unilim.fr

<sup>†</sup>These authors have contributed  
equally to this work

<sup>‡</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
B Cell Biology,  
a section of the journal  
Frontiers in Immunology

**Received:** 14 December 2020

**Accepted:** 14 April 2021

**Published:** 04 May 2021

### Citation:

Ouk C, Roland L, Gachard N,  
Poulain S, Oblet C, Rizzo D,  
Saintamand A, Lemasson Q,  
Carrion C, Thomas M, Balabanian K,  
Espéli M, Parrens M, Soubeyran I,  
Boulin M, Faumont N, Feuillard J and  
Vincent-Fabert C (2021) Continuous  
MYD88 Activation Is Associated With  
Expansion and Then Transformation of  
IgM Differentiating Plasma Cells.  
Front. Immunol. 12:641692.  
doi: 10.3389/fimmu.2021.641692

Catherine Ouk<sup>1†</sup>, Lilian Roland<sup>1†</sup>, Nathalie Gachard<sup>1‡</sup>, Stéphanie Poulain<sup>2‡</sup>, Christelle Oblet<sup>1</sup>, David Rizzo<sup>1</sup>, Alexis Saintamand<sup>1</sup>, Quentin Lemasson<sup>1</sup>, Claire Carrion<sup>1</sup>, Morgane Thomas<sup>1</sup>, Karl Balabanian<sup>3</sup>, Marion Espéli<sup>3</sup>, Marie Parrens<sup>4</sup>, Isabelle Soubeyran<sup>5</sup>, Mélanie Boulin<sup>1</sup>, Nathalie Faumont<sup>1</sup>, Jean Feuillard<sup>1\*</sup> and Christelle Vincent-Fabert<sup>1\*</sup>

<sup>1</sup> UMR CNRS 7276/INSERM U1262 CRIBL, University of Limoges, and Hematology Laboratory of Dupuytren Hospital University Center (CHU) of Limoges, Limoges, France, <sup>2</sup> UMR CANTHER « CANCER Heterogeneity, Plasticity and Resistance to THERapies » INSERM 1277-CNRS 9020 UMRS 12, University of Lille, Hematology Laboratory, Biology and Pathology Center, CHU de Lille, Lille, France, <sup>3</sup> Institut de Recherche Saint-Louis, EMILy, INSERM U1160, University of Paris, Paris, France, <sup>4</sup> Pathology Department, Hospital University Center of Bordeaux, Bordeaux, France, <sup>5</sup> Laboratory of Pathology, Institut Bergonié, Bordeaux, France

Activating mutations of *MYD88* (*MYD88*<sup>L265P</sup> being the far most frequent) are found in most cases of Waldenström macroglobulinemia (WM) as well as in various aggressive B-cell lymphoma entities with features of plasma cell (PC) differentiation, such as activated B-cell type diffuse large B-cell lymphoma (DLBCL). To understand how MYD88 activation exerts its transformation potential, we developed a new mouse model in which the MYD88<sup>L252P</sup> protein, the murine ortholog of human MYD88<sup>L265P</sup>, is continuously expressed in CD19 positive B-cells together with the Yellow Fluorescent Protein (Myd88<sup>L252P</sup> mice). In bone marrow, IgM B and plasma cells were expanded with a CD138 expression continuum from IgM<sup>high</sup> CD138<sup>low</sup> to IgM<sup>low</sup> CD138<sup>high</sup> cells and the progressive loss of the B220 marker. Serum protein electrophoresis (SPE) longitudinal analysis of 40 Myd88<sup>L252P</sup> mice (16 to 56 weeks old) demonstrated that ageing was first associated with serum polyclonal hyper gammaglobulinemia (hyper Ig) and followed by a monoclonal immunoglobulin (Ig) peak related to a progressive increase in IgM serum levels. All Myd88<sup>L252P</sup> mice exhibited spleen enlargement which was directly correlated with the SPE profile and was maximal for monoclonal Ig peaks. Myd88<sup>L252P</sup> mice exhibited very early increased IgM PC differentiation. Most likely due to an early increase in the Ki67 proliferation index, IgM lymphoplasmacytic (LP) and plasma cells continuously expanded with age being first associated with hyper Ig and then with monoclonal Ig peak. This peak was consistently associated with a spleen LP-like B-cell lymphoma. Clonal expression of both membrane and secreted  $\mu$  chain isoforms was demonstrated at the mRNA level by high throughput sequencing. The Myd88<sup>L252P</sup> tumor transcriptomic signature identified both proliferation and canonical NF- $\kappa$ B p65/RelA activation. Comparison with *MYD88*<sup>L265P</sup> WM showed that Myd88<sup>L252P</sup> tumors also

shared the typical lymphoplasmacytic transcriptomic signature of WM bone marrow purified tumor B-cells. Altogether these results demonstrate for the first time that continuous MYD88 activation is specifically associated with clonal transformation of differentiating IgM B-cells. Since MYD88<sup>L252P</sup> targets the IgM PC differentiation continuum, it provides an interesting preclinical model for development of new therapeutic approaches to both WM and aggressive MYD88 associated DLBCLs.

**Keywords:** MYD88 L265P mutation, lymphoplasmacytic lymphoma/Waldenström's macroglobulinemia, IgM secretion, monoclonal Ig peak, B-cell lymphoma, plasma cell

## INTRODUCTION

Waldenström's macroglobulinemia (WM) is an incurable indolent B-cell lymphoma of the elderly accounting for less than 5% of B-cell lymphomas with, as unique characteristics, a serum IgM peak and primary medullary localization of lymphoplasmacytic cells that exhibit continuous differentiation from mature B lymphocytes to IgM secretory plasma cells (1). Secondary lymphoid organ infiltration and/or a leukemic phase is found in 20% cases. Other manifestations include neuropathy, cryoglobulinemia, skin rash, cold-agglutinin hemolytic anemia, and amyloidosis (2). The discovery of the activating mutation of MYD88 (MYD88<sup>L265P</sup> being the far most frequent) in more than 90% of WM cases contributed to the concept that this entity is genetically distinct from other B-cell lymphomas (3, 4). Being present in 50% of IgM monoclonal gammopathies of undetermined significance (MGUS), MYD88 mutations are most likely a primary event in WM (5). Considered as secondary genetic events, activating mutations of CXCR4 (CXCR4<sup>S338X</sup> or CXCR4<sup>WHIM</sup>), a receptor implicated in migration and bone marrow (BM) homing of leucocytes, are found in 30% of WM cases (6). Additional mutations of CD79b, ARID1A or TP53 have been reported (7).

Despite these advances, WM pathophysiology is incompletely understood. Its treatment remains challenging and the exact role of MYD88 mutations in the emergence of lymphoplasmacytic B-cell clones is not known (7, 8). Indeed, MYD88 mutations are also found in 30% of activated B-cell type diffuse large B-cell lymphomas (ABC-DLBCL), more than half of primary cutaneous DLBCLs, leg type, and many DLBCLs at immune-privileged sites but not in plasma cell myelomas, even IgM types (9). It should be noted that IgM expression is a surrogate marker of ABC-DLBCLs (10). Moreover, all these aggressive B-cell tumors associated with MYD88, which often exhibit morphological features of plasma cell (PC) differentiation, are all associated with expression of the PC differentiation marker IRF4. MYD88 protein is the canonical adapter for inflammatory signaling pathways to downstream members of the Toll-like receptor (TLR) and interleukin-1 receptor (IL-1R) families. Forming the myddosome complex, MYD88 binds IL-1R or TLR family members to IRAK kinases family. IRAK activation leads to activation of the NF kappa B (NF-κB) transcription factor and interferon 3 and 7 regulatory factors (IRF3 and 7). MYD88<sup>L265P</sup> constitutively increases formation of the myddosome complex with downstream NF-κB activation (3, 11, 12).

Experimentally, MYD88<sup>L265P</sup> is essential for survival of ABC-DLBCL and WM cell lines (3, 11). A recent publication suggests the involvement of HOIP and LUBAC-dependent NF-κB activation in the transformation potential of MYD88 activation in a mouse model (13). The current published mouse models with continuous MYD88 activation in the B-cell lineage develop aggressive clonal B-cell lymphomas that resemble human ABC-DLBCLs (13–15). Although discussed by Jo et al. in the HOIP/LUBAC activation context, no IgM peak was reported in these models. Therefore, the question of a direct role for MYD88 in the development of a lymphoplasmacytic lymphoma with monoclonal IgM secretion is still open. Recently, K Schmidt et al. reported a mouse model in which MYD88 activation was responsible for an indolent lymphoproliferative disorder resembling to IgM monoclonal gammopathy of unknown significance (IgM MGUS), the asymptomatic preclinical stage of WM (16).

Here, we present a new mouse model in which the MYD88<sup>L252P</sup> protein, the murine ortholog of human MYD88<sup>L265P</sup>, is continuously expressed in the B-cell lineage together with Yellow Fluorescent Protein (YFP) (Myd88<sup>L252P</sup> mice). We show that these mice first developed early expansion of CD93<sup>neg</sup> IgM PCs with an increase in both IgM secretion and bone marrow relocalization of IgM B-cells. Moreover, these mice also had increased percentages of IgM<sup>high</sup> CD138<sup>low</sup> and IgM<sup>low</sup> CD138<sup>high</sup> cells with a CD138 expression continuum between both cell types. Then, these mice developed an oligoclonal or clonal IgM lymphoplasmacytic-like B-cell lymphoma together with a serum IgM monoclonal peak. These tumors had marked transcriptomic similarities to WM but they were located in the spleen and exhibited significant increased proliferation. Despite differences between Myd88<sup>L252P</sup> LP-like B-cell tumors and WM, our results demonstrate that the MYD88 transformation potential is strongly associated with a shift in B-cell peripheral differentiation toward plasma cells with IgM secretion. These results help explain why MYD88 activation is found in most WM and in various aggressive B-cell lymphomas with IgM PC differentiation engagement such as ABC DLBCLs.

## MATERIALS AND METHODS

### Generation of the Transgenic Mouse Line Myd88<sup>L252P</sup>

The transgenes (cDNA) Myd88<sup>WT</sup>-IRES-Yfp and Myd88<sup>L252P</sup>-IRES-Yfp were synthesized (Genecust, Dudelange, Luxembourg)

and inserted into the pcDNA3.1 vector. Sequences of these transgenes are given in **Supplementary Materials and Methods**. The *Myd88<sup>L252P</sup>-IRES-Yfp* insert was cloned into the pROSA26-1 vector (17) containing a LoxP-flanked region, consisting of a stop cassette and the *Neomycin* gene (18). The transgene *Myd88<sup>L252P</sup>-IRES-Yfp* was inserted downstream from the stop cassette. JM8 embryonic stem (ES) cells were transfected with the targeting vector according to a previously described protocol (19). Targeted ES cells were screened for homologous recombination by PCR. Genomic DNA (gDNA) extraction was performed using an in house protocol and consisted of gDNA precipitation with absolute ethanol. Twenty nanograms gDNA were used for each PCR (primer sequences in **Supplementary Materials and Methods**) using LongAmp<sup>®</sup> Taq DNA Polymerase (New England Biolabs, Ipswich, MA) according to the manufacturer's recommendations. Recombined ES cells were injected into C57BL/6J blastocysts which were transferred into foster mothers to obtain chimeric mice (Myd88<sup>L252P-flSTOP</sup> mice).

## Mice

Cd19<sup>Cre</sup> mice (20) and mice carrying the *Myd88<sup>L252P</sup>-IRES-Yfp* allele were crossed to induce the expression of the transgene in B cells (Myd88<sup>L252P</sup> mice). Offspring were routinely screened by PCR using specific primers for insertion of the transgene (**Supplementary Materials and Methods**). Animals were housed at 21–23°C with a 12-hour light/dark cycle. All procedures were conducted under an approved protocol according to European guidelines for animal experimentation (French national authorization number: 8708503 and French ethics committee registration number APAFIS#14581-2018041009469362 v3).

## Cell Transfection and NF-κB Dependent Dual-Luciferase Reporter Assay

A20 cells (5.10<sup>6</sup>) were co-transfected with 5 μg of either empty pcDNA3.1, pcDNA3.1-*Myd88<sup>WT</sup>* or pcDNA3.1-*Myd88<sup>L252P</sup>* vectors, plus 100 ng pRL-TK *Renilla luciferase* expression vector (Promega Corporation, Madison, WI) and 5 μg of either the 3X-κB-L vector with three copies of the major histocompatibility complex (MHC) class I κB element or its mutated counterpart, the 3X-mutκB-L vector (21) using Amaxa L013 program (AMAXA Biosystems, Cologne, Germany). Twenty four hours after transfection, cells were lysed and luciferase activities were measured using the Dual-Luciferase Reporter Assay System and the Turner Designs TD-20/20 Luminometer (Promega Corporation, Madison, WI).

## Sera Analyses

Serum was obtained from blood collected retro-orbitally. Specific ELISA and serum electrophoresis assay were performed as previously described (19, 22).

## Flow Cytometry and *In Vivo* Proliferation Assays

In order to collect BM cells of Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> mice, femurs from both hind legs were rinsed with PBS and sternum

was gently crushed and cells filtered on a nylon meshwork that was rinsed with PBS. Spleen cells from Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> were filtered through a sterile nylon membrane. Blood samples were collected retro-orbitally. Red cells were lysed by RBC Lysis Buffer (Biolegend, San Diego, CA). Cell suspensions were resuspended at 4°C in a labeling buffer (PBS, 1% BSA, 2mM EDTA) and labeled with fluorescent conjugated monoclonal antibodies listed in **Supplementary Materials and Methods**. Labeled cells were analyzed using a BD Fortessa SORP flow cytometer (BD Bioscience France, Le Pont de Claix, France). Results were analyzed using Kaluza Flow Cytometry software 1.2 (Beckman Coulter, Brea, CA).

## Immunohistochemistry

Paraffin embedded tissue sections (5μm) were deparaffinized as follows: slides were immersed successively in xylene twice for 3 min, 3 times for 3 min in 100% ethanol, once for 3 min in 95% ethanol and 3 times in PBS for 5 min. Then, slides were immersed in citrate buffer pH7 and heated 4 times for 5 min 40 sec in a microwave at 800W. Image acquisition was performed with the Nanozoomer 2.0RS Hamamatsu and NDP.scan software (812 Joko-cho, Higashi-ku, Hamamatsu City, 431-3196, Japan). Quantification of Ki67 nuclear labelling was performed with the imageJS and the Ki67 module: <http://imagejs.org/?http://module.imagejs.googlecode.com/git/mathbiol.chromomarkers.js&http://module.imagejs.googlecode.com/git/ki67> (23).

## Gene Expression Profiling

A series of seven mice (three Cd19<sup>Cre</sup> and four Myd88<sup>L252P</sup>) was studied in parallel with bone marrow purified tumor B-cells from a series of 11 patients with *MYD88<sup>L265P</sup>* WM (series 1) as well as lymph nodes from a series of 58 patients: 19 *MYD88<sup>WT</sup>* chronic lymphocytic leukemia, 15 *MYD88<sup>L265P</sup>* WM, 12 *MYD88<sup>wt</sup>* Nodal marginal zone lymphoma, 5 *MYD88<sup>wt</sup>* WM, 4 follicular lymphoma and 3 patients with benign follicular hyperplasia (series 2, **Supplementary Tables 1 and 2**). Approval of this protocol was obtained from the local IRB of the CHRU of Lille (CSTMT043). *MYD88* and *CXCR4* mutational status was determined as previously described (6). Total mRNA was extracted from whole infiltrated tissues and purified B-cells as reported (24). For humans and mice, RNA amplification and hybridization onto microarrays were performed on an Affymetrix Human Genome U133 Plus 2.0 Array and on an Affymetrix Gene Atlas system<sup>®</sup> with the MoGene-2\_1-st-v1 Affymetrix chip (Affymetrix, Santa Clara, CA) respectively according to a previously described protocol (25) (GEO accession number GSE138273). Bioinformatic analyses are detailed in **Supplementary Materials and Methods**.

## Repertoire Analysis

RNA was extracted from total spleen, and 1μg was used for sequencing. Transcripts were amplified by 5'RACE PCR using reverse primers hybridizing within either the membrane or secreted exon of the μ or γ genes. ProtoScript<sup>®</sup> II Reverse Transcriptase (New England Biolabs, Ipswich, MA) was used for reverse transcription and amplicons were obtained using



Phusion® High Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions. Primers used are listed in **Supplementary Materials and Methods**. Illumina sequencing adapter sequences were added by primer extension, and resulting amplicons were sequenced on an Illumina MiSeq sequencing system using MiSeq kit Reagent V2 500 cycles. Paired reads were merged using FLASH (26). Repertoire analysis was done using the IMGT/HighV-QUEST online tool ([http://www.imgt.org/IMGT\\_vquest/vquest](http://www.imgt.org/IMGT_vquest/vquest)). The resulting output was parsed by in-house R script.

## RESULTS

### Generation of a Mouse Model With Insertion of the Mouse Mutation Myd88<sup>L252P</sup> Into the Rosa26 Locus

To study the effect of continuous MYD88 activation on B-cell differentiation, we created a transgene containing the mutant murine cDNA sequence of *Myd88* (*Myd88<sup>L252P</sup>*) which is orthologous to the human mutant sequence *MYD88<sup>L265P</sup>*, in frame with the Yellow Fluorescent Protein (*Yfp*) sequence and separated by an Internal Ribosome Entry Site (IRES) sequence (*Myd88<sup>L252P</sup>-IRES-Yfp*) (**Supplementary Figure 1A**). To validate this transgene, we checked that it induced expression of both MYD88<sup>L252P</sup> and YFP proteins in the murine A20 B-cell line (**Supplementary Figures 1B, C**) and that it was responsible for constitutive NF-κB activation (**Supplementary Figure 1D**). The *Myd88<sup>L252P</sup>-IRES-Yfp* insert was cloned into the pROSA26.1 vector (17) (**Supplementary Figure 1E**). In this construct, the insert was subcloned downstream from a *Neomycin-STOP* cassette flanked by LoxP sites. Chimeric mice were intercrossed to obtain stable germinal transmission of the *Myd88<sup>L252P</sup>-IRES-Yfp* transgene (*Myd88<sup>L252P</sup>-flSTOP* mice). *Myd88<sup>L252P</sup>-flSTOP* and *Cd19<sup>Cre</sup>* mice were crossed. Mice with both transgenes (*Myd88<sup>L252P</sup>* mice) were then studied, with their age matched *Cd19<sup>Cre</sup>* littermates as controls (*Cd19<sup>Cre</sup>* LMC). Specific B-cell expression of the transgene was found in more than 90% of blood and spleen B cells compared to virtually no expression in the T-cell compartment (**Supplementary Figure 1F**). As expected and as evidence of NF-κB activation, *Myd88<sup>L252P</sup>* splenocytes over-expressed the NF-κB target gene *Tnfrsf10b* at the mRNA level (**Supplementary Figure 1G**).

### Serum Protein Electrophoresis Profiles Segregate Myd88<sup>L252P</sup> Mice According to Age

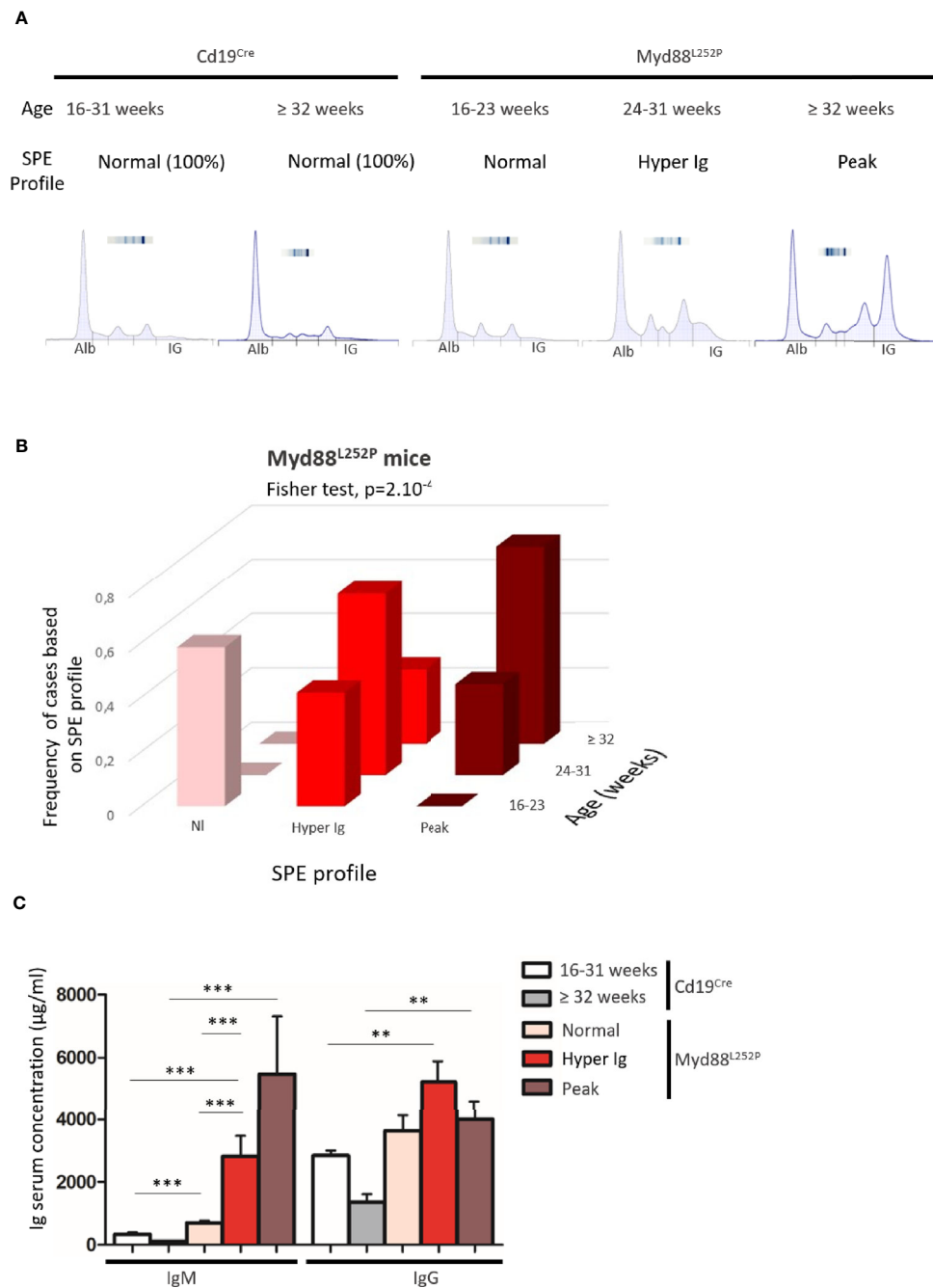
As a first exploratory step, serum protein electrophoresis (SPE) was systematically performed on a series of 40 *Myd88<sup>L252P</sup>* mice and 26 age matched *Cd19<sup>Cre</sup>* LMCs. As shown in **Figure 1A** three SPE profiles were seen: normal, polyclonal hyper gammaglobulinemia (hyper Ig) and a monoclonal Ig peak. All *Cd19<sup>Cre</sup>* LMCs exhibited normal SPE regardless of their age. In other words, SPEs with hyper Ig or Ig peaks were found only in *Myd88<sup>L252P</sup>* mice (**Figures 1A, B**). **Figure 1B** shows the relationship between the age of *Myd88<sup>L252P</sup>* mice and the SPE

profile. Young *Myd88<sup>L252P</sup>* mice (16-23 weeks) had a normal or hyper Ig SPE profile. In contrast, most mice older than 32 weeks had an Ig peak. In between these two groups, 24 to 31 week old mice (middle age) had a hyper Ig or an Ig peak in 65% and 35% cases respectively (Fisher test,  $p=2.10^{-4}$ ). ELISA quantification of serum Ig revealed that young *Myd88<sup>L252P</sup>* mice with a normal SPE exhibited a moderate IgM and IgG hyper Ig when compared to their *Cd19<sup>Cre</sup>* LMC (**Figure 1C** and **Supplementary Figure 2**). Serum IgG levels of middle aged and old *Myd88<sup>L252P</sup>* mice were variable when compared to their young counterparts. This was in contrast to serum IgM levels that were significantly increased in middle aged mice and even more so in old mice and correlated with the SPE profile and age (**Figure 1C** and **Supplementary Figure 2**).

These results first indicate that continuous MYD88 activation in B cells was associated with a global increase in Ig secretion. Second, age related occurrence of polyclonal hyper Ig and then monoclonal Ig peaks correlated with the increase in serum IgM levels. This suggests that, after a polyclonal expansion phase, aging of *Myd88<sup>L252P</sup>* mice was associated with clonal restriction of IgM-secreting B-cells, very likely reflecting the MYD88<sup>L252P</sup> B-cell transformation power in these cells. Therefore, these first results point to a strong relationship between MYD88<sup>L252P</sup> and IgM-secreting B-cells.

### Myd88<sup>L252P</sup> Bone Marrow IgM Plasma Cell Content Was Increased and Displayed a CD138 Expression Continuum

As shown in **Figure 2A**, *Myd88<sup>L252P</sup>* bone marrow global B-cell content was comparable to that of *Cd19<sup>Cre</sup>* LMCs in terms of percentages in 16 week old mice with normal SPE. Transgene expression was mainly found in *Myd88<sup>L252P</sup>* *CD19<sup>high</sup>* B-cells. Indeed, with an on/off effect, percentages of YFP positive cells (*i.e.* of LoxP rearranged cells) was directly correlated with CD19 mean fluorescence intensity (**Figure 2B**). Comparison of *CD19<sup>Cre</sup>* LMC and *Myd88<sup>L252P</sup>* mice did not reveal any significant bone marrow B-cell increase with age. However older *Myd88<sup>L252P</sup>* mice had increased levels of IgM<sup>pos</sup> *CD19<sup>high</sup>* B-cells (**Figures 3A, B**, left panel and **Supplementary Figure 3** for the gating strategy). Strikingly, a CD138 expression continuum was clearly evident in a triple parametric B220/CD138/IgM histogram gated on mature B-cells and/or PCs in *Myd88<sup>L252P</sup>* mice only (**Figure 3A**, lower panel see in red the decrease in B220 and the increase in CD138 expression). This CD138 expression continuum, that we recently showed to be characteristics of MYD88<sup>L265P</sup> WM bone marrow tumor B-cells (Gayet et al, Cytometry B 2021), started from IgM<sup>high</sup> CD138<sup>low</sup> and ended at IgM<sup>low</sup> CD138<sup>high</sup> cells (**Figure 3A**, lower panel). This CD138 expression continuum was absent in *Cd19<sup>Cre</sup>* LMCs. Consequently, *Myd88<sup>L252P</sup>* mice showed increased percentages of both IgM<sup>high</sup> CD138<sup>low</sup> B-cells (most likely precursors of IgM PCs) and total bone marrow PCs (**Figure 3B**, right panel). Moreover, the proportion of bone marrow IgM PCs was significantly increased in young *Myd88<sup>L252P</sup>* mice and even more in older *Myd88<sup>L252P</sup>* mice (**Figure 3C**). Indeed, *Myd88<sup>L252P</sup>* *CD19<sup>pos</sup>/YFP<sup>pos</sup>* B cells tended to accumulate in the

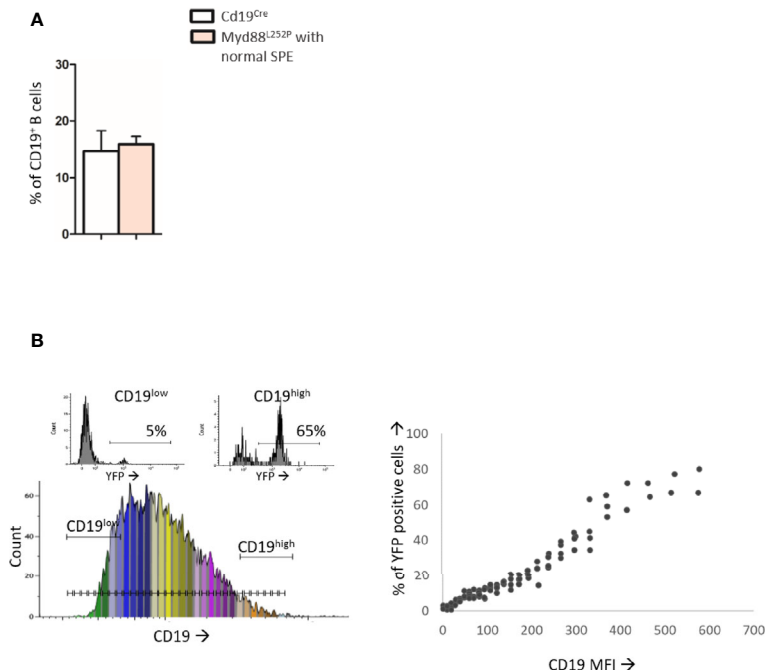


**FIGURE 1** | Myd88<sup>L252P</sup> transgenic mice exhibited serum IgM hypergammaglobulinemia and then monoclonal IgM peaks when ageing. **(A)** Examples of serum protein electrophoresis of Cd19<sup>Cre</sup> mice (respectively 16 and 36 weeks-old) and three Myd88<sup>L252P</sup> mice (16, 24 and 36 week-old) with normal, polyclonal hypergammaglobulinemia (hyper Ig) and monoclonal Ig peaks respectively. **(B)** Frequencies of cases according to SPE profile and age for Myd88<sup>L252P</sup> mice (n = 40). **(C)** IgM and IgG serum levels in Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> mice. For Cd19<sup>Cre</sup> (n = 15), Myd88<sup>L252P</sup> (n = 36). Results are expressed as the mean ± SEM. Mann Whitney test p-value < 0.01 and < 0.001 are symbolized by \*\* and \*\*\* respectively.

IgM<sup>high</sup> B-cell compartment when compared to its Myd88<sup>L252P</sup> CD19<sup>pos</sup>/YFP<sup>neg</sup> counterpart (**Supplementary Figure 4**).

Altogether, these results indicate that transgene expression started in a minority of CD19<sup>weak</sup> B-cell precursors and was

mainly expressed at the latest stages of B-cell lymphopoiesis when CD19 expression was high. Evidence for bone marrow increase in both IgM<sup>high</sup> CD138<sup>low</sup> and PCs with the characteristic CD138 expression continuum may either suggest



**FIGURE 2** | Analysis of B-cell differentiation in bone marrow from 16 week-old Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> mice. **(A)**: Percentage of CD19<sup>pos</sup> and/or B220<sup>pos</sup> B-cells in bone marrow from 16 week-old Cd19<sup>Cre</sup> (n = 9) and Myd88<sup>L252P</sup> mice (n = 9). Results are expressed as the mean ± SEM. **(B)**: Flow cytometry analysis of the transgene expression according to CD19 expression levels (n=9). Left panel: CD19 monoparametric histogram sliced according to CD19 MFI intervals. For each CD19 MFI interval, the percentage of bone marrow B220<sup>pos</sup> YFP positive cells was noted. Two examples of YFP monoparametric histograms are presented in the upper part, one for CD19<sup>low</sup> B220<sup>pos</sup> B-cells and one for CD19<sup>high</sup> B220<sup>pos</sup> B-cells with their respective percentages of YFP<sup>pos</sup> cells. Right panel: percentage of YFP<sup>pos</sup> B220<sup>pos</sup> B-cells (Y axis) according to CD19 MFI (X axis).

that the bone marrow of Myd88<sup>L252P</sup> mice had the ability to home IgM PC precursors and PCs and/or a shift in peripheral B-cell differentiation toward IgM PC in Myd88<sup>L252P</sup> mice.

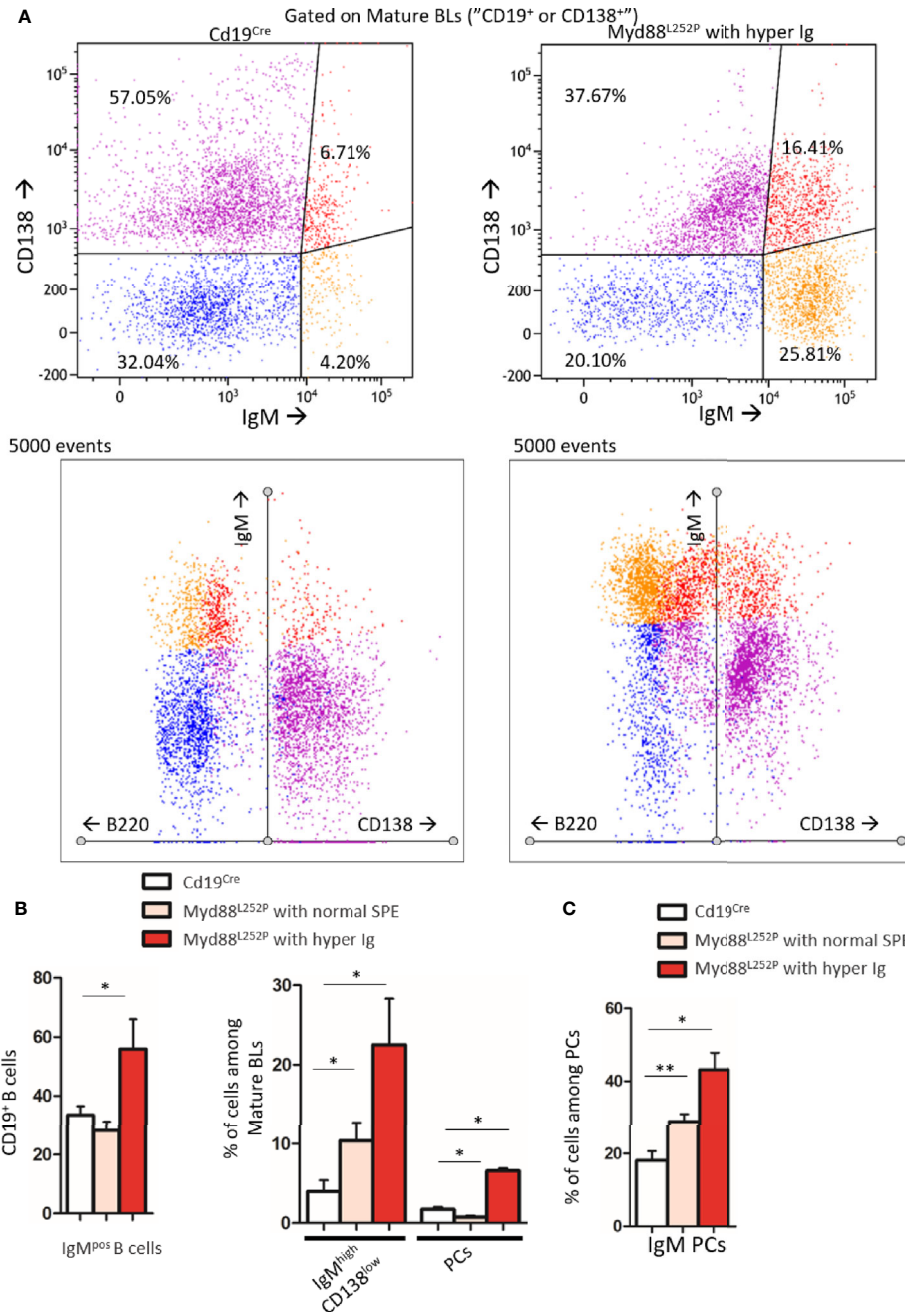
## MYD88<sup>L252P</sup> First Induced Peripheral Early Lymphoplasmacytic and Plasma Cells Expansion and Then B-Cell Transformation Into a Lymphoplasmacytic-Like Lymphoma

Extended white blood cell differential was not altered and mice did not exhibit any palpable/visible peripheral lymphadenopathy (data not shown) regardless of the SPE profile. Young Myd88<sup>L252P</sup> mice with normal SPE tended to have spleen enlargement when compared to their age related Cd19<sup>Cre</sup> LMCs (**Figure 4**). Spleen enlargement was dramatically increased in Myd88<sup>L252P</sup> mice with hyper Ig and even more so in those with an IgM peak, a feature that was most likely related to B-cell transformation. Indeed, the B/T cell ratio was markedly increased in these latter mice (**Supplementary Figure 5**).

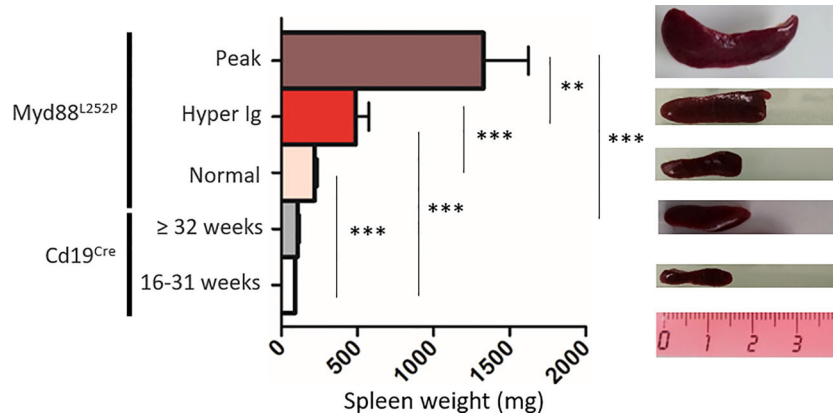
While spleen histology of young Myd88<sup>L252P</sup> mice was comparable to that of their Cd19<sup>Cre</sup> LMCs, with an apparent normal spleen architecture, Myd88<sup>L252P</sup> mice with hyper Ig or an Ig peak exhibited enlarged and congruent lymphoid nodules (**Figures 5A–E**). At high magnification, a marked lymphoplasmacytic aspect

consisting of a mixture of small to large lymphocytes with numerous lymphoplasmacytic cells (LP cells) and PCs was noted in all Myd88<sup>L252P</sup> mice whatever their SPE profile (**Figures 5F–J**). This spleen aspect was particularly striking for mice with Ig peaks, and was characterized by massive and diffuse infiltration of lymphoplasmacytic cells that evoked human B-cell lymphomas with features of PC differentiation, further called “LP-like lymphoma” or “LP-like tumors”. (see also the cytological imprint in **Supplementary Figure 6**). Presence of LP cells and PCs in Myd88<sup>L252P</sup> spleen was invariably confirmed by immunohistochemistry after intracellular Ig labeling regardless of SPE status, with numerous LP cells and terminally differentiated PCs (cells with intermediate or strong intracytoplasmic Ig labeling respectively). Noteworthy, cell densities were markedly increased in Myd88<sup>L252P</sup> mice with hyper Ig or with an Ig peak (**Figures 5K–O**).

Based on B-cell expression of CD21 and CD23, frequencies of CD21<sup>pos</sup> CD23<sup>high</sup> follicular B-cells were not significantly altered in Myd88<sup>L252P</sup> mice regardless of their SPE status (**Figure 6A** and **Supplementary Figure 7**). In contrast, a decrease of CD21<sup>high</sup> CD23<sup>pos</sup> marginal zone B-cells was observed. This cell content nearly disappeared in Myd88<sup>L252P</sup> mice with an Ig peak. Only total PCs were increased in these mice (**Figure 6B**). However, among total spleen PCs, percentages of IgM PCs were increased in a similar manner in all Myd88<sup>L252P</sup> mice no matter what their SPE status was (**Figure 6C** and **Supplementary**



**FIGURE 3 |** Increase in the IgM PC compartment in bone marrow from Myd88<sup>L252P</sup> transgenic mice. **(A)** Example of bi and triple parametric flow cytometry histograms gated on mature B or plasma cells for expression of IgM, B220 and CD138 for Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> mice (left and right panels respectively). Upper panels: IgM<sup>low</sup> or neg CD138<sup>neg</sup>, IgM<sup>high</sup> CD138<sup>low</sup>, IgM<sup>high</sup> CD138<sup>high</sup> and IgM<sup>low</sup> CD138<sup>high</sup> cells are colored in blue, orange, red and purple respectively using a hinged quadstat of the Kaluza software and an IgM/CD138 2-dimensional plot. The hinged quadstat was set-up for one Cd19<sup>Cre</sup> mouse. Lower panels: triple parametric histograms using the radar function of the Kaluza software. Note the CD138 expression continuum on Myd88<sup>L252P</sup> bone marrow B-cells that correlated with a progressive decrease in B220 expression. This CD138 expression continuum was virtually absent in Cd19<sup>Cre</sup> mice. **(B)** Percentages of total IgM<sup>pos</sup> B cells and IgM<sup>high</sup> CD138<sup>low</sup> pre PCs and B220<sup>low</sup> CD138<sup>high</sup> PCs in bone marrow from 16-24 week-old Cd19<sup>Cre</sup> (n = 6) and Myd88<sup>L252P</sup> mice with normal SPE or hyper Ig (n = 6 and n = 3 respectively). Results are shown as the mean ± SEM. Mann Whitney test p-value < 0.05 is symbolized by \*. **(C)** Percentages of IgM<sup>pos</sup> CD138<sup>high</sup> PCs among total PCs in bone marrow from 16-31 week-old Cd19<sup>Cre</sup> (n = 6) and Myd88<sup>L252P</sup> mice with normal SPE or hyper Ig (n = 6 and n = 3 respectively). Results are shown as the mean ± SEM. Mann Whitney test p-value < 0.05 and < 0.01 are symbolized by \* and \*\* respectively.



**FIGURE 4** | Myd88<sup>L252P</sup> mice exhibited a progressively increasing splenomegaly consistently related to the SPE profile. Spleen size of Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> age-paired mice. Myd88<sup>L252P</sup> mice were sacrificed together with at least one Cd19<sup>Cre</sup> mouse of the same age. Left panel: distribution of spleen weights; right panel: examples of spleen macroscopy (Cd19<sup>Cre</sup> n = 9 for 16-31 weeks-old and n = 16 for ≥ 32 weeks; Myd88<sup>L252P</sup> with normal SPE: n = 9; with hyper Ig: n = 13; with Ig peak: n = 19). Results are given as the mean ± SEM. Mann Whitney test p-value < 0.01 and p-value < 0.001 are symbolized by \*\* and \*\*\* respectively.

**Figure 7**). Moreover IgM PCs were predominantly CD93<sup>neg</sup> suggesting that they belonged to the proliferating PC compartment (27) (**Figure 6C** and **Supplementary Figure 7**).

Therefore, morphological and immunophenotypic results indicated that continuous MYD88 activation was associated with continuous peripheral IgM PC differentiation very early on and that these LP and PC subsets continuously expanded with age first being associated with hyper Ig and then with an Ig monoclonal peak and a LP-like lymphoma aspect in the spleen.

### Proliferation Index of Myd88<sup>L252P</sup> LP-Like Tumors Was Moderately Increased in Myd88<sup>L252P</sup> Tumors With an LP-Aspect

To better study these Myd88<sup>L252P</sup> LP-like tumors, we compared their Ki67 proliferation index to that of Cd19<sup>Cre</sup> LMCs as well as to L.CD40, L.CD40/Δc-MYC mice as controls. L.CD40 mice are a model of marginal zone spleen B cell indolent lymphomas without plasma cell differentiation but with NF-κB activation (28). L.CD40/Δc-MYC mice are a model of ABC-DLBCLs with both c-Myc and NF-κB activation in B-cells (24). Very few Ki67 positive cells were seen outside germinal centers in spleen sections from Cd19<sup>Cre</sup> mice (**Figure 7A** panel A). The Ki67 index was weak in L.CD40 mice (**Figure 7A** panel B). By contrast, the vast majority of cells from L.CD40/Δc-MYC tumors were Ki67 positive (**Figure 7A** panel C). Being moderately increased in young mice with a normal SPE, the Ki67 proliferative index was further enhanced in mice with hyper Ig and even higher in those with an Ig peak, but never reached the proliferation index levels of L.CD40/Δc-MYC tumors (**Figures 7A**, panels D–7 and **7B**). *In vivo* incorporation of BrdU was tested for four mice with an Ig peak and one with polyclonal hyper Ig. Results confirmed that the proliferation index was consistently increased (**Supplementary Figure 8**) albeit always less than 30% with this technique.

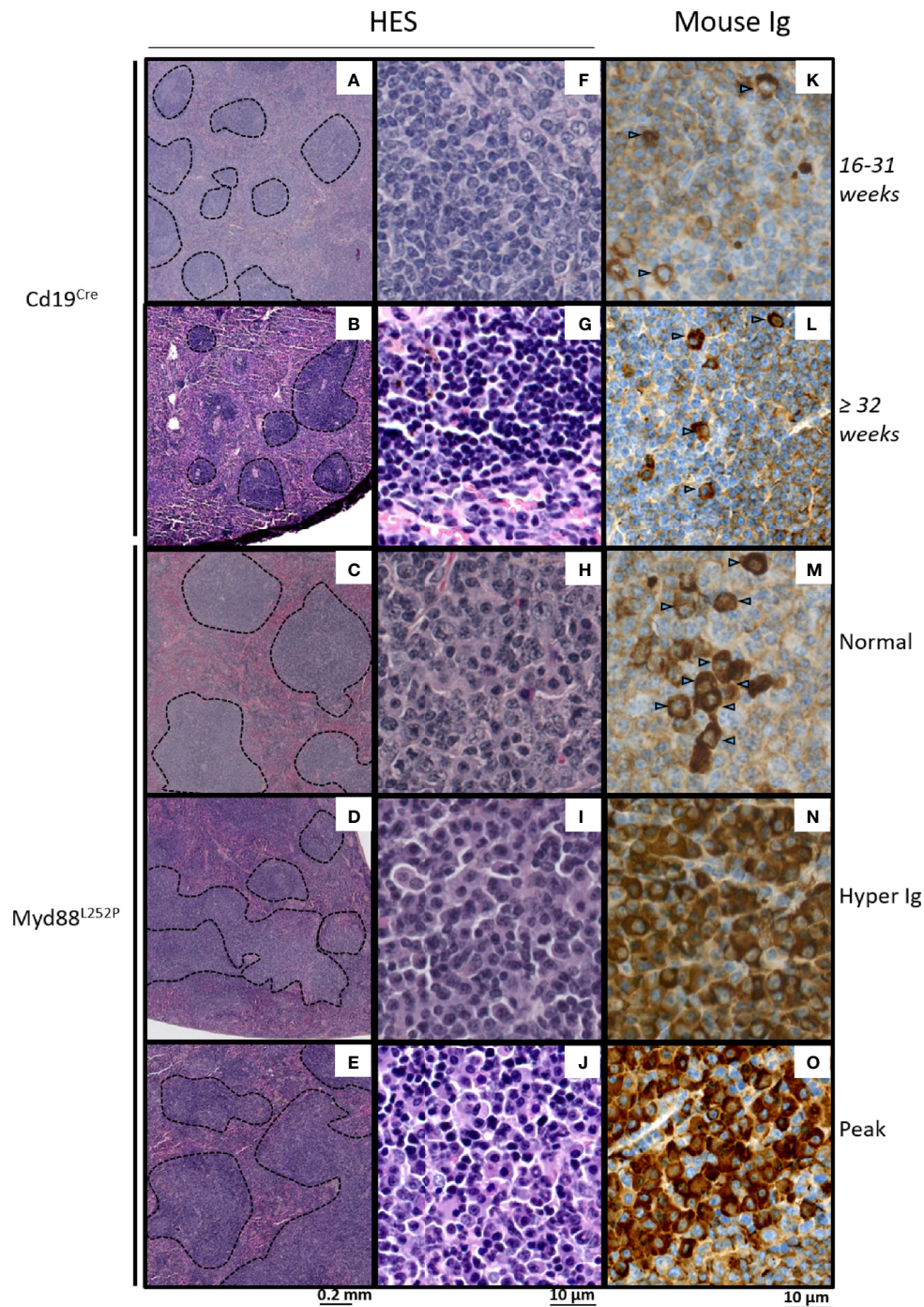
Altogether, these results show that MYD88<sup>L252P</sup> expression in B-cells was responsible for progressive peripheral B-cell

expansion related to an early increase in B-cell proliferation. Mice with an Ig peak clearly exhibited a lymphoproliferative disease with a marked increase in proliferation index but with features of an LP-like lymphoma such as the presence of numerous LP cells and PCs.

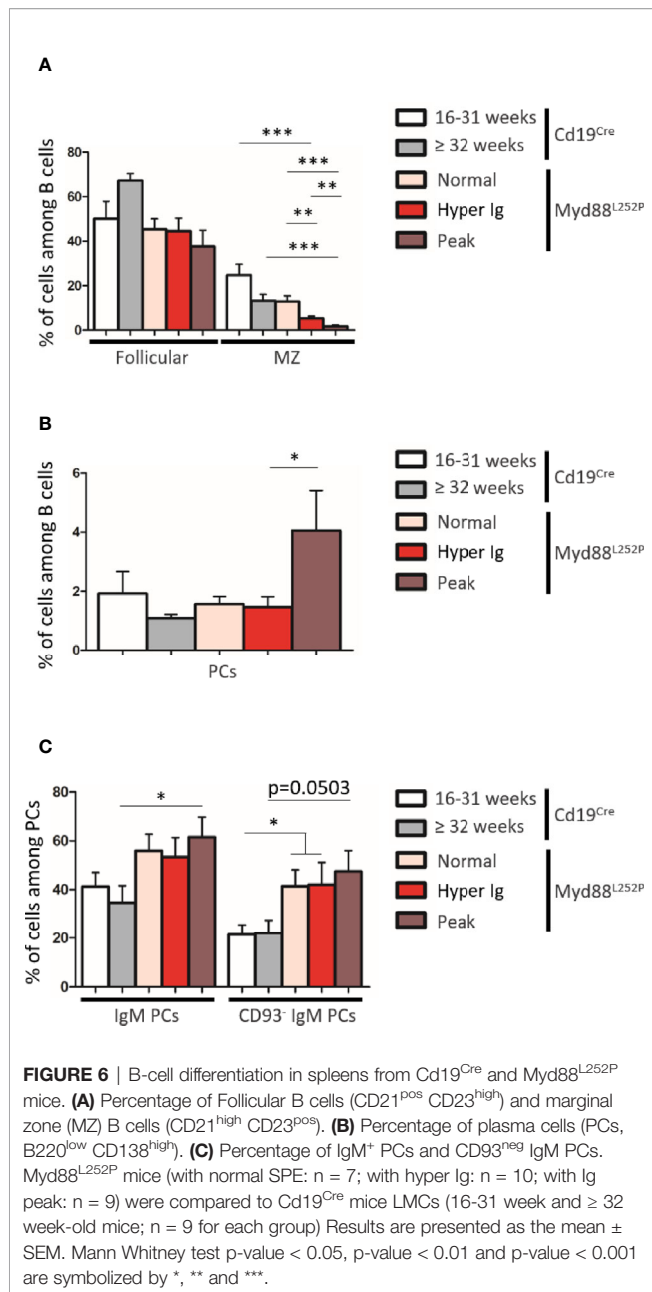
### Myd88<sup>L252P</sup> Mice With Ig Peaks Developed IgM but Not Igg Monoclonal or Oligoclonal B-Cell Lymphomas With Expression of Both Membrane and Secretory Heavy Chain mRNA

Since Myd88<sup>L252P</sup> mice had a global hyper Ig even if predominantly IgM, it was important to assess μ or γ chain clonality of tumor surface and secreted immunoglobulins at the molecular level. Six LP-like cases with monoclonal Ig peaks and five Cd19<sup>Cre</sup> mice were studied. mRNA reverse transcription followed by RACE PCR with primers specific for either the membrane or secreted forms of mouse μ or γ heavy chains was performed, followed by high throughput sequencing (HTS) of the VDJ-Cμ or VDJ-Cγ region (**Figure 8A**). **Figures 8B, C** show the relative frequency of the five most abundant mRNA clones for the μ or γ heavy chains respectively. RACE PCR and HTS results indicate that Myd88<sup>L252P</sup> mice developed clonal or oligoclonal B-cell expansion with expression of both secreted and membrane forms of the μ heavy chain that had the same VDJ-Cμ clonal rearrangement (**Figure 8B**), without any bias in terms of V segment usage (not shown). The same RACE PCR technique with primers specific for either the membrane or secreted form of the mouse γ heavy chain did not identify any significant B-cell clonal expansion in Myd88<sup>L252P</sup> tumors (**Figure 8C**).

These results indicate that, despite initial IgM and IgG hyper Ig in young Myd88<sup>L252P</sup> mice, MYD88<sup>L252P</sup> specifically promoted IgM B-cell lymphomagenesis with clonal expression of both membrane and secreted μ chain isoforms identical VDJ gene rearrangements. These results are in full agreement with the



**FIGURE 5** | Morphological and immunophenotypic plasma cell engagement of Myd88<sup>L252P</sup> tumors. Hematein eosin morphological aspect (**A–J**) and intracytoplasmic Ig labeling (**K–O**) of Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> spleens: one 16–31 week-old Cd19<sup>Cre</sup> (**A, F, K**), one ≥ 32 week-old Cd19<sup>Cre</sup> (**B, G, L**) and three Myd88<sup>L252P</sup> (**C, H, M**) for Normal group, (**D, I, N**) for Hyper Ig group and **E, J, O** for Peak group) mice are shown. Panels (**A–E**) at low magnification, Myd88<sup>L252P</sup> tumors often exhibited a nodular pattern (**C–E**). Panels (**F–J**): at high magnification, most Myd88<sup>L252P</sup> tumors had small B-cell aspects with marked lymphoplasmacytic engagement (panels **H–J**). Panels (**K–O**) Myd88<sup>L252P</sup> tumors with a lymphoplasmacytic aspect exhibited marked plasma cell differentiation as revealed by the presence of intracytoplasmic Ig in numerous cells (arrows) with various labeling intensities (**M–O**).



lymphoplasmacytic aspect Myd88<sup>L252P</sup> tumors and the persistent IgM plasma cell differentiation continuum.

## Transcriptomic Signature of Myd88<sup>L252P</sup> LP-like Tumors Revealed NF- $\kappa$ B RelA Activation, Proliferation and Plasma Cell Differentiation and Overlapped With Waldenström's Macroglobulinemia Gene Expression Profile (GEP)

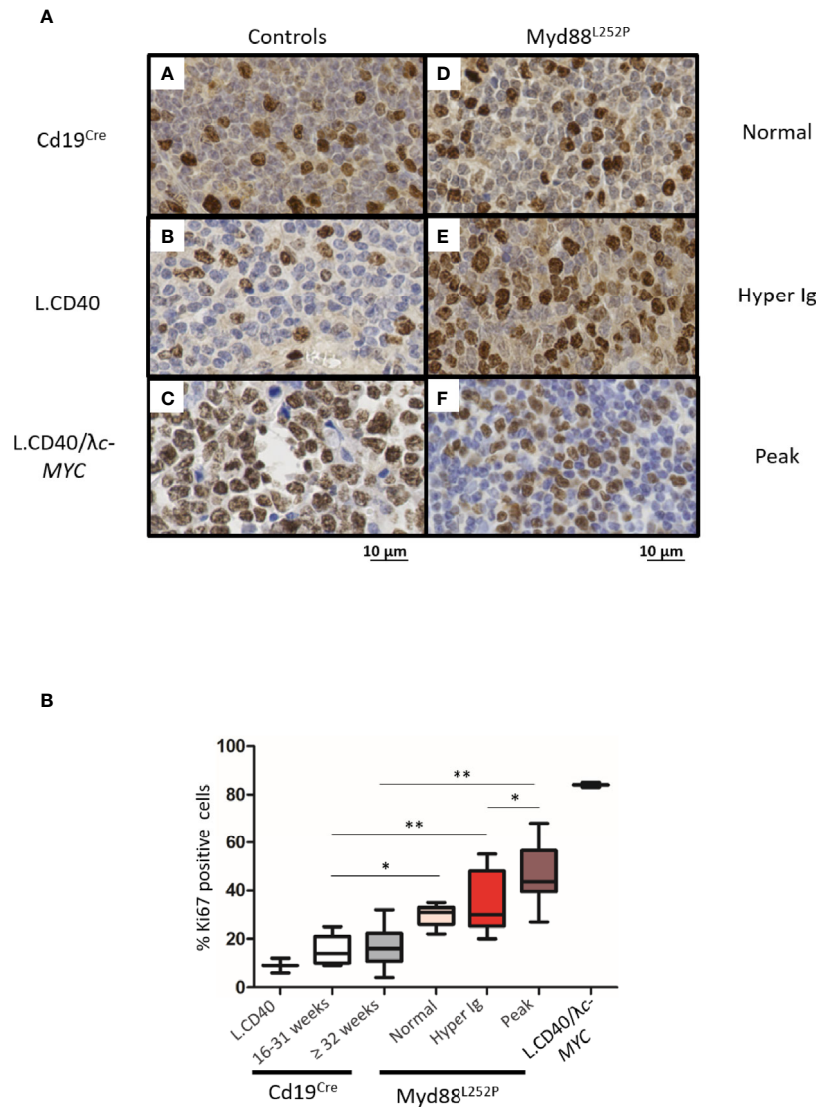
To explore the transcriptomic signature of Myd88<sup>L252P</sup> LP-like tumors in conditions similar to those of most studies on human B-cell lymphomas and to look for common features with WM,

we selected a short series of massively invaded Myd88<sup>L252P</sup> spleen tumors with monoclonal Ig peaks. Comparisons were first done with their Cd19<sup>Cre</sup> LMCs, and then with WM patients.

With a fold change of 2, a set of 1515 differentially expressed genes were found in Myd88<sup>L252P</sup> spleen tumors when compared to spleens from Cd19<sup>Cre</sup> LMCs (**Supplementary Table 3**). To analyze this set of genes, we combined both K-mean and hierarchical clustering and principal component analysis, as already published (29). Following this methodology, deregulated genes in Myd88<sup>L252P</sup> spleen tumors could be segmented into 14 classes of co-regulated genes (**Figure 9** with methodological details in **Supplementary Materials and Methods, Supplementary Figure 9** and **Supplementary Table 4**). Consistent with increased B/T ratios in Myd88<sup>L252P</sup> LP-like tumors, expression of genes belonging to the T-cell lineage, as well as T-cell signaling and activation signatures was down-regulated in Myd88<sup>L252P</sup> spleen tumors (**Figure 9**, see clusters I, K and L). Of note, RelB signature was associated with that of T-cells and was decreased in Myd88<sup>L252P</sup> spleen tumors. In contrast, expression of genes belonging to the proliferation, RelA NF- $\kappa$ B activation pathway and plasma cell differentiation signatures were up-regulated (**Figure 9**, cluster J mainly as well as cluster B, G and M for proliferation).

To identify MYD88<sup>L252P</sup> deregulated genes in common with those of WM patients having the MYD88<sup>L265P</sup> mutation, transcriptomes of Myd88<sup>L252P</sup> tumors were compared to those of purified WM bone marrow B-cells from a series of 11 MYD88<sup>L265P</sup> WM patients. A subset of 319 coherently dysregulated genes in both Myd88<sup>L252P</sup> LP-like tumors and WM tumor B-cells (163 up and 156 down, **Supplementary Tables 5, 6**) was extracted from the 1515 differentially expressed genes in Myd88<sup>L252P</sup> LP-like tumors. Unsupervised hierarchical clustering showed a 95% coherency between the branches of up and down regulated genes in both WM tumor B-cells and LP-like Myd88<sup>L252P</sup> spleen tumors (**Figure 10A**, left and middle panels and **Supplementary Tables 5, 6**). To check whether this Myd88<sup>L252P</sup>/WM signature could discriminate WM from other indolent NHLs, an independent series of 58 patients, including 15 MYD88<sup>L265P</sup> WM, five MYD88<sup>wt</sup> WM, 12 MYD88<sup>wt</sup> NMZLs and 19 MYD88<sup>wt</sup> CLL was analyzed (**Supplementary Tables 2, 7**). All MYD88<sup>L265P</sup> WM patients clustered together after unsupervised hierarchical clustering (**Figure 10A**, right panel and **Supplementary Tables 7, 8**). Also belonging to the MYD88<sup>L265P</sup> WM cluster were 3/5 (60%) MYD88<sup>wt</sup> WM and 2/12 (17%) MYD88<sup>wt</sup> NMZLs. We used the linear predicting score described by Wright et al. (31) to estimate the informativeness of the Myd88<sup>L252P</sup>/WM signature for WM diagnosis. As shown in **Figure 10B**, as set of 174 genes (84 up and 90 down, **Figure 10B**) was found to predict WM with over 90% probability (**Supplementary Tables 9, 10**).

The plasma cell signature was the main component of the Myd88<sup>L252P</sup>/WM GEP. Genes such as *EDEM1* and 2, *IRF4* or *XBPI* were over-expressed while others such as *PAX5* or *BCL6* were markedly down-regulated (**Figure 10**). Consistently, functions revealed by Gene Set Enrichment analysis mainly corresponded to endoplasmic reticulum and Golgi apparatus



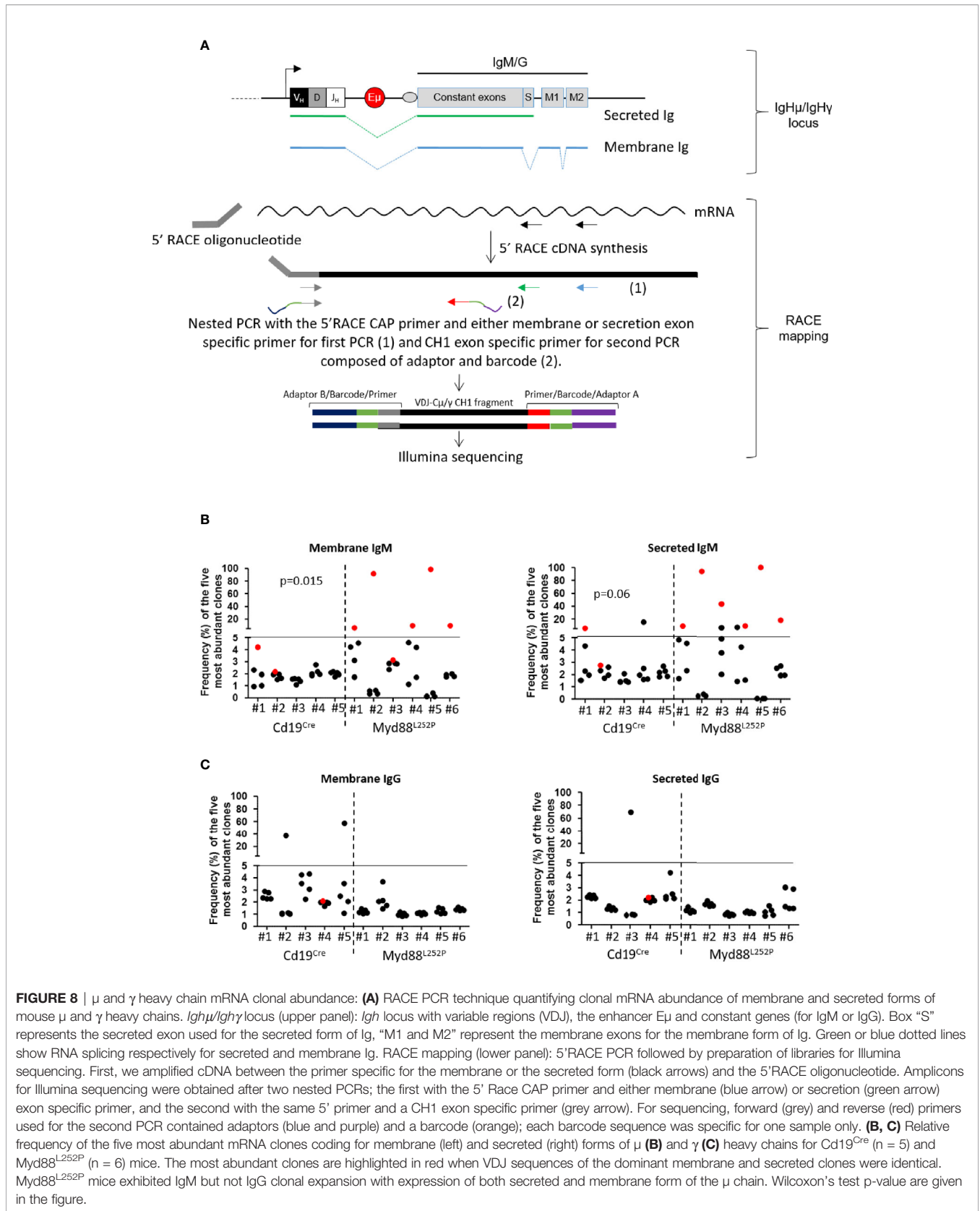
**FIGURE 7** | Intermediate increase in Myd88<sup>L252P</sup> tumor proliferation rate. **(A)** Examples of Ki67 labeling in spleen sections from three controls (panels A-C) and three Myd88<sup>L252P</sup> mice (panels D-F, n=28). Controls were Cd19<sup>Cre</sup> LMCs (panel A, n = 5 for 16-31 week-old mice and n = 8 for ≥ 32 week-old mice), L.CD40 mice (panel B) with indolent splenic lymphomas of marginal zone B-cells (n=2), as well as L.CD40/λc-Myc mice (panel C, n=2) with a ABC-DLBCL lymphomas (24, 28). The three Myd88<sup>L252P</sup> tumors (panels D-F) are one representative example of each group defined according the SPE profile (normal n = 5, hyper Ig n = 8 and Ig peak n = 15). Here, L.CD40 mice were used as a model of indolent B-cell lymphoma with a low proliferation index while L.CD40/λc-Myc mice were a model of aggressive B-cell lymphoma with a high proliferation index. **(B)** Quantification of Ki67 labeling. Box plots represent the median and quartile of percentages of Ki67 positive cells. Mann Whitney test p-value < 0.05, p-value < 0.01 are symbolized by \* and \*\* respectively.

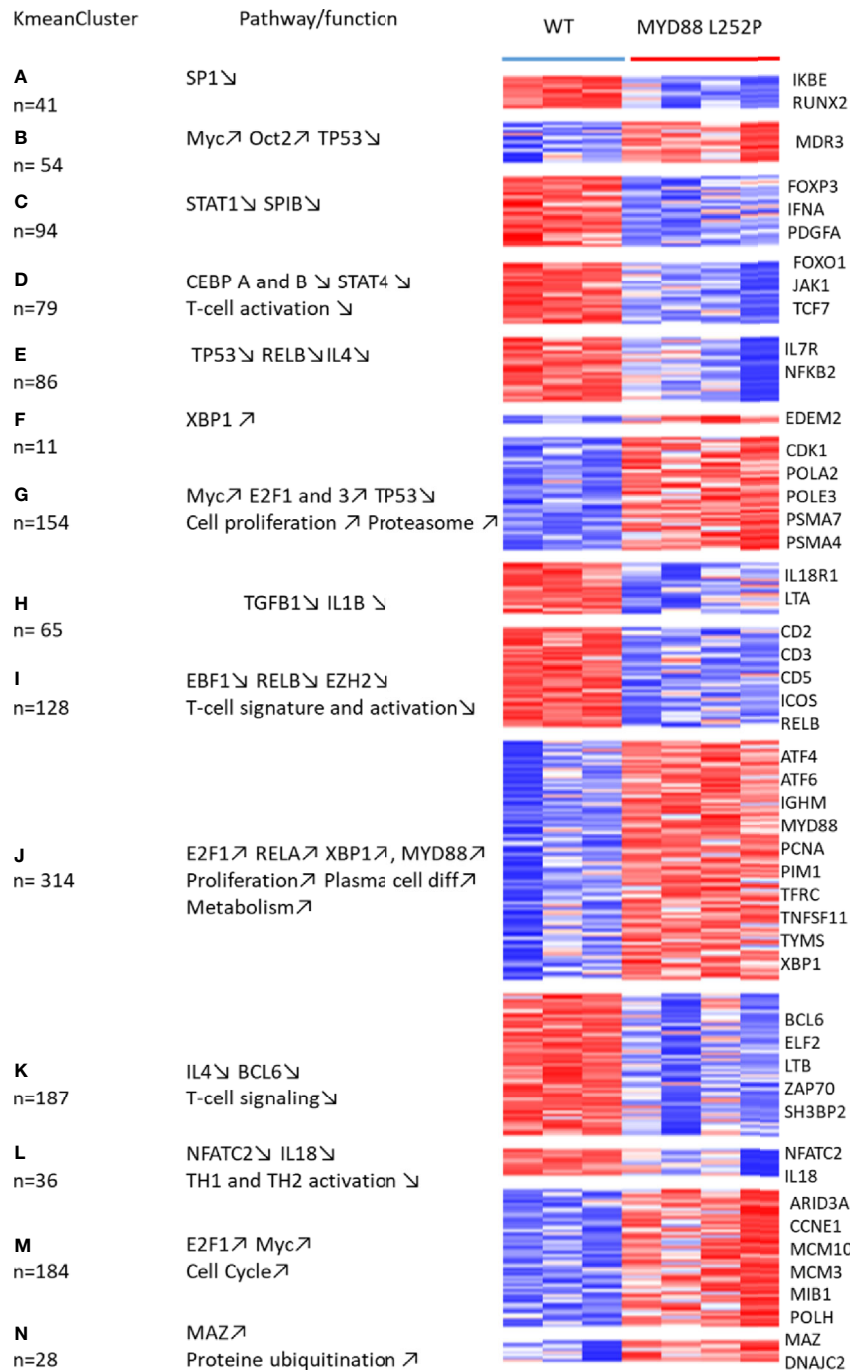
(not shown). In accordance with results published by Hunter et al. (30), expression of genes such as *CXCR4*, *DUSP22*, *PIM1* and 2 or *TRAM1* was increased while expression of *SNED1* was decreased. Few genes of the Myd88<sup>L252P</sup>/WM signature overlapped with those of ABC/GC DLBCLs published by Wright et al. (31). These overlapping genes, corresponding only to those that are overexpressed in ABC DLBCLs, were *IRF4*, *IGHM*, *CXCR4*, *P2RX5*, *PIM1* and *PIM2*. In other words, the Myd88<sup>L252P</sup>/WM signature did not significantly overlap with that of GCB DLBCLs. Among other deregulated genes were cyclin kinase inhibitors *CDKN1B* (*p27<sup>kip1</sup>*) and *CDKN2C* (*p18/*

*INK4AC*), mutations of the former being found in hairy cell leukemias (32). *RASSF3* and *KRAS* were also up-regulated. *RASSF3* belongs to the Ras association domain family (RASSF).

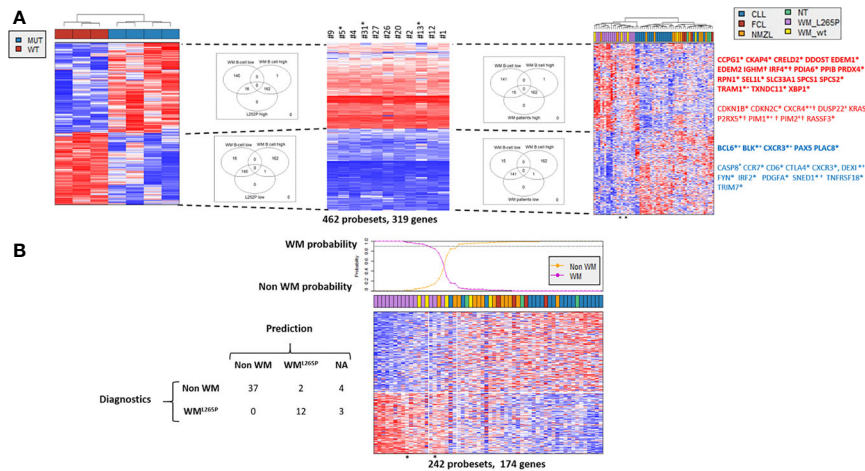
Altogether, the Myd88<sup>L252P</sup> tumor signature highlights proliferation as well as canonical NF-κB p65/RelA activation (but not RelB), which is in agreement with the known fact that MYD88 activates the classical NF-κB pathway. The Myd88<sup>L252P</sup> tumor signature also strikingly confirms that lymphoplasmacytic differentiation is at the heart of MYD88 related B-cell transformation in mice, a feature shared with WM tumors with the MYD88<sup>L265P</sup> mutation.







**FIGURE 9** | Whole transcriptome analysis of Cd19<sup>Cre</sup> and Myd88<sup>L252P</sup> mice: total mRNA was extracted from whole spleen tissues. Gene expression profiles were obtained using the MoGene-2\_1-st-v1 Affymetrix chip. mRNA transcripts (3236) were selected to be differentially expressed using the Limma R package. Expressed genes that were too heterogeneous were eliminated, resulting in a final selection of 1515 genes. These genes were segregated into 40 Kmean clusters. The closest Kmean clusters were merged 2 by 2 according to their proximity by principal component analysis of the mean vectors. This was repeated until maximization of the absolute value of Chi2 (29). This resulted in 14 aggregated clusters. Functional annotation of the aggregated Kmean clusters was performed using the Ingenuity Pathway Analysis (IPA) Software. Annotated heatmap of the 1515 genes are segregated into the 14 aggregated clusters. Left: the aggregated Kmean clusters with the corresponding number of genes; middle: main pathways and or function identified with the IPA software; right: some relevant genes.



**FIGURE 10** | Comparison of gene expression profile (GEP) of *Myd88*<sup>L252P</sup> mice and patients with WM or other indolent B-cell NHL. Affymetrix differential gene expression profiles (GEP) between four *Myd88*<sup>L252P</sup> mice (MUT, n=4) and three *Cd19*<sup>Cre</sup> mice (WT, n=3) were compared to the Affymetrix GEP of purified bone marrow tumor B-cells from 11 WM patients with the *MYD88*<sup>L265P</sup> mutation, resulting in selection of 462 probesets (319 genes). This selection was used on the Affymetrix transcriptome of an independent series of lymph node biopsies from 58 patients: 19 *MYD88*<sup>wt</sup> chronic lymphocytic leukemias (CLL), 15 *MYD88*<sup>L265P</sup> WM (*WM*<sub>L265P</sub>), 12 *MYD88*<sup>wt</sup> nodal marginal zone lymphomas (NMZL), 5 *MYD88*<sup>wt</sup> WM with IgM peaks (*WM*<sub>WT</sub>), 4 follicular lymphomas (FCL) and 3 patients with follicular hyperplasia (NT). **(A)** Hierarchical clustering and heatmap of the 462 selected probesets for mice (left), purified bone marrow *MYD88*<sup>L265P</sup> WM B-cells (middle) and lymph nodes (right). Down and up-regulated genes are in blue and red respectively. Branches of down and up regulated genes in *Myd88*<sup>L252P</sup> mice, *MYD88*<sup>L265P</sup> WM bone marrow B-cells and *MYD88*<sup>L265P</sup> WM lymph nodes are delineated by dashed lines. Venn diagrams of the intersections between the branches are shown, highlighting the consistency between branches across the different clustering. Some genes of interest are noted on the right. In bold are those of the plasma cell signature; \*: genes in the predictor (see **Figure 7B**); +: genes reported by Hunter et al. in WM (30); †: genes of the ABC/GC DLBCL signature (31). **(B)**: Informativeness of *MYD88*<sup>L265P</sup> WM diagnosis using the 462 selected probesets defined in **Figure 8A**. The 462 probesets defined from *MYD88*<sup>L252P</sup> mice and bone marrow tumor B-cells from *MYD88*<sup>L265P</sup> WM patients were used to predict *MYD88*<sup>L265P</sup> WM diagnosis (*WM*<sup>L265P</sup> versus non WM) from other lymphomas within the series of 58 lymph node biopsies. Probabilities that each sample belongs to *WM*<sup>L265P</sup> versus non WM group are indicated. The *WM*<sup>L265P</sup> versus non WM or not attributed (NA) assignment is shown on the left.

## DISCUSSION

Different mouse preclinical models with continuous MYD88 activation in B-cells have been published; All but one demonstrate the B-cell transformation potential of MYD88 but without presenting evidence for a correlation between IgM B-cell LP and PC differentiation (13–15). The first report was published by Knittel et al. (14). The authors generated a mouse model that allows B-cell conditional expression of the *Myd88*<sup>L252P</sup> allele from the endogenous *Myd88* locus. In this model, expression of *Myd88*<sup>L252P</sup> would be regulated in a manner similar to that of the wild type allele. At least three regulatory levels of MYD88 activity have been reported. The MYD88 regulatory region harbors various transcription binding sites such as those for NF- $\kappa$ B, IRF1, SP1 or STAT factors and it was shown that this gene is regulated by IL-6 (33), which suggests a role for its expression in either plasma cell differentiation or inflammation. An alternative splice variant of *MYD88*, *MYD88s*, lacks exon 2 and is unable to activate NF- $\kappa$ B. This variant is also able to form a heterodimer with the full length MYD88 protein, resulting in decreased formation of the myddosome complex (34). *MYD88s* is increased during sepsis and is thought to ensure robust termination of MYD88 dependent inflammation (35, 36). As a third regulation mechanism, hypomethylation and upregulation of MYD88 are important for NF- $\kappa$ B activation (37). MYD88 promoter demethylation is important in glioblastoma and

is associated with increased MYD88 protein expression in lung cancers (38, 39). Mice from Knittel's model occasionally developed DLBCLs when they aged. Therefore, the Knittel mouse model raises the question of the relationship between B-cell lymphomagenesis and regulation of the endogenous *Myd88* locus throughout B-cell life. In addition, this model also showed hyper Ig, even if polyclonal, which may suggest that expansion of Ig secreting B-cells could precede the lymphoma development. In that view, K Schmidt et al. recently published a mouse model with a transgene closed to the one of the Knittel's model. But, after exploring the effect of MYD88 activation in three different Cre context, *Cd19*-Cre, *Cy1*-Cre and *Cd19*-Cre<sup>ERT2</sup>, as well as analyzing the effect of NP-immunization, the authors concluded that MYD88 continuous activation promotes survival of long term IgM expressing B-cells with clonal restriction a monoclonal serum IgM paraprotein resulting in an IgM MGUS-like disorder (16).

To our knowledge, two papers, from Wang et al. and Sewastianik et al. reported the effect of the human MYD88<sup>L265P</sup> protein in murine B-cells (15, 40). In the model of Wang et al, primary murine B-cells were retrovirally infected *ex vivo* before reinjection. In this model, an initial boost of B-cell proliferation was seen followed by B-cell apoptosis in a Bim-dependent manner. Importantly, no increase in immunoglobulin secretion was reported. Beside the fact that retroviral infection of B-cells may have its own interfering effects, this model raises the question of whether the human MYD88<sup>L265P</sup> protein may have exactly the

same activation properties on a mouse B-cell background as in human B-lymphocytes. In the model of Sewastianik et al, the human loxP-flanked-stop-MYD88<sup>L265P</sup> transgene was inserted downstream from the mouse *Collagen type I alpha 1 chain* (*Col1A1*) gene and MYD88<sup>L265P</sup> expression was induced by crossing with *Aid<sup>Cre</sup>* mice (15). In addition to focal skin rashes, some WM features were noted such as expansion of lymphoplasmacytic cells and increased IgM serum levels. However, only DLBCL clonal transformation was seen. Because AID is mainly expressed in germinal center B-cells and because the promoter of the *Col1A1* gene is highly active in fibroblasts and osteoblasts (41), this model raises the question whether temporality and/or expression pattern could be important in MYD88 driven B-cell lymphomagenesis.

The *Rosa26* locus has been solidly proven to be valuable for expression of numerous oncogenes in the B-cell lineage (see reference (42) for discussion as well as the literature of the K Rajewsky's group). By inserting our *Myd88<sup>L252P</sup>-IRES-Yfp* transgene in this locus, we forced continuous expression of the mutated MYD88 protein in a heterozygous-like context while respecting the native MYD88 activation pathways of mouse B-cells. Moreover, we were able to monitor our transgene expression by flow cytometry due to YFP. Thereby, we created a conditional *Myd88<sup>L252P</sup>* mouse model closed to the one published by Jo et al. (13). However, these authors mainly focused their work on the synergy between MYD88 and the catalytic subunit HOIP which increases LUBAC ligase activity that in turns promotes NF- $\kappa$ B canonical activation; only four so-called CD19-cre-MYD88LP have been studied at the tumor stage.

Here, a longitudinal analysis of a series of 40 *Myd88<sup>L252P</sup>* mice compared to their age matched Cd19<sup>Cre</sup> LMCs demonstrated that IgM plasma cell expansion is at the heart of MYD88 dependent B-cell transformation. Indeed, by examining clonal restriction of IgM secreting B-cells, we first showed that ageing of *Myd88<sup>L252P</sup>* mice was associated with polyclonal hyper Ig followed by monoclonal Ig peak due to increased serum IgM. Second, we provide evidence indicating that bone marrow relocation of IgM B-cells, IgM<sup>high</sup> CD138<sup>low</sup> cells and IgM PCs was increased in *Myd88<sup>L252P</sup>* mice with a CD138 expression continuum, that is a characteristic of WM tumor B-cells. Third, analysis of spleen morphology and spleen B-cell subsets by flow cytometry indicated that continuous MYD88 activation was associated very early with peripheral LP cell and CD93<sup>neg</sup> IgM PC expansion and that these cell subsets were markedly increased at the time of the Ig monoclonal peak. Fourth, appearance of a monoclonal Ig peak was constantly associated with a B-cell lymphoma with marked features of lymphoplasmacytic differentiation, so-called *Myd88<sup>L252P</sup>* LP-like lymphoma. Fifth, at the molecular level, *Myd88<sup>L252P</sup>* specifically promoted IgM B-cell lymphomagenesis with mRNA clonal expression of both membrane and secreted  $\mu$  (but not  $\gamma$ ) chain isoforms. Finally, the *Myd88<sup>L252P</sup>* tumor gene expression profile not only highlights the canonical NF- $\kappa$ B p65/RelA activation pathway and proliferation, but also strikingly shares an *Xbp1* centered lymphoplasmacytic B-cell differentiation signature with

MYD88<sup>L265P</sup> WM. This signature differentiates MYD88<sup>L265P</sup> WM from other indolent B-cell tumors including marginal zone lymphomas. Our results contradict the conclusions of Sewastianik et al. (15), and, being in line with those of K Schmidt et al. (16), firmly demonstrate the specific transforming effect of MYD88 activation in IgM PC differentiating B-cells.

To establish the gene expression signature from bulk *Myd88<sup>L252P</sup>* spleen tumors rather than from purified B-cells is matter of discussion. The very obvious disadvantage of working on bulk tumors is certainly that all mRNA species from all cell types present in the tissue are mixed together. Even if massively invaded, both stromal and other residual immune cells (T-cells, macrophages, dendritic cells...) persist constantly in the tumor. In these conditions, specifically assigning a given mRNA expression pattern to tumor cells is always hazardous. However, because tumors were immediately snap-frozen, all mRNA species are supposed to be well preserved without any significant experimental bias. On the other hand, working on purified cells may also induce artefacts since the abundance of different mRNA species may vary during the time of purification which can also stress the cells. Above all, in the specific case of *Myd88<sup>L252P</sup>* spleen tumors, which exhibit an LP aspect with continuous PC differentiation, the key question would have been to choose the right negative selection marker. Indeed, tumor cells are phenotypically heterogeneous with variable expression of B220, surface Ig or CD138 for example. Should we have selected B220<sup>high</sup> versus B220<sup>low</sup> B-cells or CD138<sup>low</sup> versus CD138<sup>high</sup> cells? Rather than make wrong or partial choices concerning which tumor cells to purify, we chose to work on bulk *Myd88<sup>L252P</sup>* spleen tumors with well-preserved total mRNA and to compare this bulk signature to that of WM, including purified WM bone marrow tumor B-cells.

One characteristic feature of *Myd88<sup>L252P</sup>* B-cells was the strong reduction of the marginal zone B-cell compartment. While also characterized by continuous activation of NF- $\kappa$ B, the L.CD40 mouse model published by Hömig-Hözel et al. (28), in which B-cells are submitted to continuous CD40 activation signaling, showed expansion of marginal zone B-lymphocytes. This points out the differences between TLRs and CD40 in terms of NF- $\kappa$ B activation. Indeed, CD40 is able to activate both the classical and alternative pathways, i.e. to induce the nuclear translocation of RelA and RelB NF- $\kappa$ B containing complexes while TLRs only activate the classical pathway. The effect of both pathways as well as the strength of NF- $\kappa$ B activation on B-cell fate has been extensively discussed by Pillai et al. (43). In this review, the authors also indicate that BTK activation blocks the Notch signaling pathway that is essential for marginal zone B-cell differentiation. It turns out that Hunter et al. have shown that *Myd88* is able to activate BTK in a BCR independent manner (30), which consequently could repress B-cell maturation toward the marginal zone B-cell lineage.

Gene expression profiles of LP-like *Myd88<sup>L252P</sup>* tumors distinctly suggest the involvement of RelA rather than RelB. RelA, but not RelB, is clearly associated with EBV-dependent B-cell immortalization and with EBV-associated DLBCL tumors, which exhibit a phenotype close to that of ABC-DLBCLs (25).

RelA is also essential for development of GC-derived PCs (44) and immunohistochemistry detected nuclear RelA in WM B-cells (45). Indeed, LP-like Myd88<sup>L252P</sup> B-cell tumors shared strong overlaps with human WM in terms of gene expression profile. Even if a few genes were in common with the ABC-DBCL signature such as *IGHM*, *CXCR4* or *PIM1* and *PIM2*, the Myd88<sup>L252P</sup>/WM signature points to dysregulation of plasma cell differentiation as the keystone of MYD88 transforming physiopathology. It also suggests that *KRAS* activation could be important. *RASSF3* and *KRAS* itself were up-regulated in both Myd88<sup>L252P</sup> B-cell tumors and human WM. Consistently, and in agreement with results from the group of Treon (30), we also found *RASSF6* overexpression in WM patients (NG and JF unpublished results). *CXCR4*, whose expression is increased in WM, may activate the RAS pathway through RasGAP-associated proteins (46). The most frequent mutations involve *KRAS* and *NRAS* genes in multiple myeloma (47). Even if such mutations have not been reported in WM (6), our results highlight the putative role of the Ras activating pathway in WM, which may lead to the design of novel therapies.

Despite similarities between WM and Myd88<sup>L252P</sup> LP-like tumors such as serum Ig monoclonal peak, increase in IgM prePCs, PC bone marrow relocation and marked lymphoplasmacytic differentiation of tumor cells some major differences exist. The first difference, is the predominant site of tumor involvement. Even if splenomegaly is found in 20% of bona fide WM patients, it is largely admitted that bone marrow is the primary tumor site. In contrast, Myd88<sup>L252P</sup> LP-like mouse tumors mainly developed in the spleen. Physiologically, IgM PCs tend to reside mainly in the spleen whereas switched IgG PCs migrate to the bone marrow (48). This raises the question of why WM IgM tumor B-cells migrate to bone marrow. In that view, it is largely suspected that *CXCR4* mutation could play a role in bone marrow homing. Further studies could also evaluate the transforming potential of Myd88 by adoptive transfer of LP-like tumor B-cells. Whether the transferred tumor will retain the lymphoplasmacytic aspect would also be interesting.

Another significant difference is the presence of large cells and increased proliferation in Myd88<sup>L252P</sup> LP-like mouse tumors. The increased proliferation index was an early event since it was also found in young Myd88<sup>L252P</sup> mice with normal SPE. As we previously discussed (24), only three mouse models for indolent lymphomas of the spleen have been published, one mimicking TRAF3 inactivation, the second with constitutive expression of BCL10 and the last one with continuous CD40 signaling (the L.CD40 model that we used as a control in **Figure 7**) (28, 49, 50). These three models are characterized by increased RelB activation. We previously demonstrated that immune surveillance may influence morphology and proliferation in the L.CD40 model. In this model, immunosuppression led to transformation of small indolent B-cell L.CD40 tumors into large B-cells with increased proliferation. Reactivating the anti-tumor response using anti-PD-L1 immunotherapy led to tumor regression (51, 52). These results on this mouse model suggest that the immunologically silent “indolent phenotype” of a B-cell tumor could be related to the immune pressure exerted on tumor B-cells. Whether and how activation of the alternative and canonical NF-κB pathways

differently disturb immune surveillance remains to be determined, and comparison of both L.CD40 and Myd88<sup>L252P</sup> mouse models could provide answers. However, as in the L.CD40 tumor model and in spleen marginal zone lymphomas (52, 53), the PD1/PD-L1 axis is most likely playing a role in the immune escape of aggressive tumor B-cells with MYD88 activation. Using Eμ-*MYC* transgenic hematopoietic stem cells (HSC) stably transduced with naturally occurring NF-κB mutants to generate various primary mouse lymphomas, Reimann et al. recently showed that MYD88 tumors express high levels of PD-L1 and that anti-PD-1 therapies induce T-cell dependent senescence of tumor cells (54). PD-L1 surface expression is weak or absent on WM tumor B-cells. However soluble PD-L1 serum levels are increased in WM patients and PD-L1 is upregulated by IL6 (55).

In summary, our longitudinal study of Myd88<sup>L252P</sup> mice demonstrated that continuous MYD88 activation is able to promote early expansion of IgM LP cells and PCs with, first, serum polyclonal hyper Ig and then a monoclonal Ig peak. Ig peaks were constantly associated with B-cell lymphomas sharing characteristics with WM. Two major differences with WM were the spleen localization of Myd88<sup>L252P</sup> tumors and increased proliferation. Here, we showed for the first time that IgM lymphoplasmacytic B-cell differentiation is at the heart of Myd88<sup>L252P</sup> transforming potential. Thus, we also provide an interesting preclinical model for development of new therapeutic approaches or to study immune surveillance for example not only in WM but also in others B-cell lymphomas with features of plasma cell differentiation. Indeed, a better understanding of the underlying molecular mechanisms is necessary in order to develop new therapies for these incurable B-cell cancers.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/geo/>, GSE138273.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the local IRB of the CHRU of Lille (CSTMT043). The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by French national authorization number: 8708503 and French ethics committee registration number APAFIS#14581-2018041009469362 v3.

## AUTHOR CONTRIBUTIONS

COu and LR contributed equally to this work. COu and LR performed and analyzed experiments. AS helped to perform the

repertoire analysis. COB performed and analyzed ELISA. MD and NG performed the transcriptomic experiments. JF, SP, AS and LR performed the bioinformatics analyses. QL and CC performed flow cytometry analysis of bone marrow B-cell subsets. NF participated in the design of the project. KB and ME participated in the development of this study. CV-F created the mouse model, contributed to the experiments and analyzed the results. JF and CV-F directed the study and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

The group of JF is supported by grants from the Ligue Nationale Contre le Cancer (Equipe labellisée Ligue), the Comité Orientation Recherche Cancer (CORC), the France Lymphome Espoir association, the Nouvelle Aquitaine Region and the Haute-Vienne and Corrèze committees of the Ligue Nationale Contre le Cancer. CV-F was supported by the France Lymphome Espoir association of patients. SP is supported by the Septentrion committee of Ligue contre le Cancer. ME is supported by an ANR @RAction grant (ANR-14-ACHN-0008), an ANR JCJC

## REFERENCES

- Preud'homme JL, Seligmann M. Immunoglobulins on the Surface of Lymphoid Cells in Waldenström's Macroglobulinemia. *J Clin Invest* (1972) 51(3):701–5. doi: 10.1172/JCI106858
- Ghobrial IM. Are You Sure This is Waldenström Macroglobulinemia? *Hematol Am Soc Hematol Educ Prog* (2012) 2012:586–94. doi: 10.1182/asheducation.V2012.1.586.3798562
- Treon SP, Xu L, Yang G, Zhou Y, Liu X, Cao Y, et al. Myd88 L265P Somatic Mutation in Waldenström's Macroglobulinemia. *N Engl J Med* (2012) 367(9):826–33. doi: 10.1056/NEJMoa1200710
- Gachard N, Parrens M, Soubeyran I, Petit B, Marfak A, Rizzo D, et al.IGHV Gene Features and MYD88 L265P Mutation Separate the Three Marginal Zone Lymphoma Entities and Waldenström Macroglobulinemia/Lymphoplasmacytic Lymphomas. *Leukemia* (2013) 27(1):183–9. doi: 10.1038/leu.2012.257
- Jiménez C, Sebastián E, Chillón MC, Giraldo P, Mariano Hernández J, Escalante F, et al. Myd88 L265P is a Marker Highly Characteristic of, But Not Restricted to, Waldenström's Macroglobulinemia. *Leukemia* (2013) 27(8):1722–8. doi: 10.1038/leu.2013.62
- Poulain S, Roumier C, Venet-Caillault A, Figeac M, Herbaux C, Marot G, et al. Genomic Landscape of CXCR4 Mutations in Waldenström Macroglobulinemia. *Clin Cancer Res* (2016) 22(6):1480–8. doi: 10.1158/1078-0432.CCR-15-0646
- Baron M, Simon L, Poulain S, Leblond V. How Recent Advances in Biology of Waldenström's Macroglobulinemia may Affect Therapy Strategy. *Curr Oncol Rep* (2019) 21(3):27. doi: 10.1007/s11912-019-0768-4
- Leblond V, Kastiris E, Advani R, Ansell SM, Buske C, Castillo JJ, et al. Treatment Recommendations From the Eighth International Workshop on Waldenström's Macroglobulinemia. *Blood* (2016) 128(10):1321–8. doi: 10.1182/blood-2016-04-711234
- Onaindia A, Medeiros LJ, Patel KP. Clinical Utility of Recently Identified Diagnostic, Prognostic, and Predictive Molecular Biomarkers in Mature B-cell Neoplasms. *Mod Pathol* (2017) 30(10):1338–66. doi: 10.1038/modpathol.2017.58
- Ruminy P, Etancelin P, Couronné L, Parmentier F, Rainville V, Mareschal S, et al. The Isotype of the BCR as a Surrogate for the GCB and ABC Molecular Subtypes in Diffuse Large B-cell Lymphoma. *Leukemia* (2011) 25(4):681–8. doi: 10.1038/leu.2010.302

grant (ANR-19-CE15-0019-01), an IDEX Université de Paris grant, a Fondation Arthritis grant and a Fondation ARC grant (PJA20181208173). KB is supported by an ANR PRC grant (ANR-17-CE14-0019), an INCa grant (PRT-K 2017) and the Association Saint Louis pour la Recherche sur les Leucémies.

## ACKNOWLEDGMENTS

We thank the animal, histology and cytometry facilities of the BISCEM US 42 INSERM/UMS 2015 CNRS technological platform of the University of Limoges as well as the immunology laboratory of the Limoges University Hospital Center for their technical assistance. We thank Dr Jeanne Cook Moreau (UMR CNRS 7276/INSERM 1276) for careful English editing.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.641692/full#supplementary-material>

- Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim K-H, et al. Oncogenically Active MYD88 Mutations in Human Lymphoma. *Nature* (2011) 470(7332):115–9. doi: 10.1038/nature09671
- de Groen RAL, Schrader AMR, Kersten MJ, Pals ST, Vermaat JSP. MYD88 in the Driver's Seat of B-cell Lymphomagenesis: From Molecular Mechanisms to Clinical Implications. *Haematologica* (2019) 104(12):2337–48. doi: 10.3324/haematol.2019.227272
- Jo T, Nishikori M, Kogure Y, Arima H, Sasaki K, Sasaki Y, et al. LUBAC Accelerates B-cell Lymphomagenesis by Conferring Resistance to Genotoxic Stress on B Cells. *Blood* (2020) 136(6):684–97. doi: 10.1182/blood.2019002654
- Knittel G, Liedgens P, Korovkina D, Seeger JM, Al-Baldawi Y, Al-Maarri M, et al. B-Cell-Specific Conditional Expression of Myd88p.L252P Leads to the Development of Diffuse Large B-cell Lymphoma in Mice. *Blood* (2016) 127(22):2732–41. doi: 10.1182/blood-2015-11-684183
- Sewastianik T, Guerrero ML, Adler K, Dennis PS, Wright K, Shanmugam V, et al. Human MYD88L265P is Insufficient by Itself to Drive Neoplastic Transformation in Mature Mouse B Cells. *Blood Adv* (2019) 3(21):3360–74. doi: 10.1182/bloodadvances.2019000588
- Schmidt K, Sack U, Graf R, Winkler W, Popp O, Mertins P, et al. B-Cell-Specific Myd88 L252p Expression Causes a Premalignant Gammopathy Resembling Igm MGUS. *Front Immunol* (2020) 11:602868. doi: 10.3389/fimmu.2020.602868
- Soriano P. Generalized LacZ Expression With the ROSA26 Cre Reporter Strain. *Nat Genet* (1999) 21(1):70–1. doi: 10.1038/5007
- Lakso M, Sauer B, Mosinger B, Lee EJ, Manning RW, Yu SH, et al. Targeted Oncogene Activation by Site-Specific Recombination in Transgenic Mice. *Proc Natl Acad Sci USA* (1992) 89(14):6232–6. doi: 10.1073/pnas.89.14.6232
- Vincent-Fabert C, Fiancette R, Pinaud E, Truffinet V, Cogné N, Cogné M, et al. Genomic Deletion of the Whole IgH 3' Regulatory Region (hs3a, hsl2, hs3b, and hs4) Dramatically Affects Class Switch Recombination and Ig Secretion to All Isotypes. *Blood* (2010) 116(11):1895–8. doi: 10.1182/blood-2010-01-264689
- Rickert RC, Roes J, Rajewsky K. B Lymphocyte-Specific, Cre-mediated Mutagenesis in Mice. *Nucleic Acids Res* (1997) 25(6):1317–8. doi: 10.1093/nar/25.6.1317
- Faumont N, Chanut A, Benard A, Cogne N, Delsol G, Feuillard J, et al. Comparative Analysis of Oncogenic Properties and Nuclear Factor-KappaB Activity of Latent Membrane Protein 1 Natural Variants From Hodgkin's Lymphoma's Reed-Sternberg Cells and Normal B-Lymphocytes. *Haematologica* (2009) 94(3):355–63. doi: 10.3324/haematol.13269

22. Chauzeix J, Laforêt M-P, Deveza M, Crowther L, Marcellaud E, Derouault P, et al. Normal Serum Protein Electrophoresis and Mutated IGHV Genes Detect Very Slowly Evolving Chronic Lymphocytic Leukemia Patients. *Cancer Med* (2018) 7(6):2621–8. doi: 10.1002/cam4.1510
23. Almeida JS, Iriabho EE, Gorrepati VL, Wilkinson SR, Grüneberg A, Robbins DE, et al. Imagejs: Personalized, Participated, Pervasive, and Reproducible Image Bioinformatics in the Web Browser. *J Pathol Inform* (2012) 3:25. doi: 10.4103/2153-3539.98813
24. David A, Arnaud N, Fradet M, Lascaux H, Ouk-Martin C, Gachard N, et al. c-Myc Dysregulation is a Co-Transforming Event for Nuclear Factor- $\kappa$ B Activated B Cells. *Haematologica* (2017) 102(5):883–94. doi: 10.3324/haematol.2016.156281
25. Chanut A, Duguet F, Marfak A, David A, Petit B, Parrens M, et al. RelA and RelB Cross-Talk and Function in Epstein-Barr Virus Transformed B Cells. *Leukemia* (2014) 28(4):871–9. doi: 10.1038/leu.2013.274
26. Magoč T, Salzberg SL. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* (2011) 27(21):2957–63. doi: 10.1093/bioinformatics/btr507
27. Chevrier S, Genton C, Kallies A, Karnowski A, Otten LA, Malissen B, et al. CD93 is Required for Maintenance of Antibody Secretion and Persistence of Plasma Cells in the Bone Marrow Niche. *Proc Natl Acad Sci U S A* (2009) 106(10):3895–900. doi: 10.1073/pnas.0809736106
28. Hömig-Hölzel C, Hojer C, Rastelli J, Casola S, Strobl LJ, Müller W, et al. Constitutive CD40 Signaling in B Cells Selectively Activates the Noncanonical NF- $\kappa$ B Pathway and Promotes Lymphomagenesis. *J Exp Med* (2008) 205(6):1317–29. doi: 10.1084/jem.20080238
29. Faumont N, Durand-Panteix S, Schlee M, Grömminger S, Schuhmacher M, Hölzel M, et al. c-Myc and Rel/NF- $\kappa$ B are the Two Master Transcriptional Systems Activated in the Latency III Program of Epstein-Barr Virus-Immortalized B Cells. *J Virol* (2009) 83(10):5014–27. doi: 10.1128/JVI.02264-08
30. Hunter ZR, Xu L, Yang G, Tsakmaklis N, Vos JM, Liu X, et al. Transcriptome Sequencing Reveals a Profile That Corresponds to Genomic Variants in Waldenström Macroglobulinemia. *Blood* (2016) 128(6):827–38. doi: 10.1182/blood-2016-03-708263
31. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A Gene Expression-Based Method to Diagnose Clinically Distinct Subgroups of Diffuse Large B Cell Lymphoma. *Proc Natl Acad Sci USA* (2003) 100(17):9991–6. doi: 10.1073/pnas.1732008100
32. Durham BH, Getta B, Dietrich S, Taylor J, Won H, Bogenberger JM, et al. Genomic Analysis of Hairy Cell Leukemia Identifies Novel Recurrent Genetic Alterations. *Blood* (2017) 130(14):1644–8. doi: 10.1182/blood-2017-01-765107
33. Harroch S, Gothelf Y, Revel M, Chebath J. 5' Upstream Sequences of MyD88, an IL-6 Primary Response Gene in M1 Cells: Detection of Functional IRF-1 and Stat Factors Binding Sites. *Nucleic Acids Res* (1995) 23(17):3539–46. doi: 10.1093/nar/23.17.3539
34. Janssens S, Burns K, Tschopp J, Beyaert R. Regulation of interleukin-1- and Lipopolysaccharide-Induced NF- $\kappa$ B Activation by Alternative Splicing of Myd88. *Curr Biol* (2002) 12(6):467–71. doi: 10.1016/S0960-9822(02)00712-1
35. Adib-Conquy M, Adrie C, Fitting C, Gattolliat O, Beyaert R, Cavaillon J-M. Up-Regulation of MyD88s and SIGIRR, Molecules Inhibiting Toll-like Receptor Signaling, in Monocytes From Septic Patients. *Crit Care Med* (2006) 34(9):2377–85. doi: 10.1097/01.CCM.0000233875.93866.88
36. Lee FF-Y, Davidson K, Harris C, McClendon J, Janssen WJ, Alper S. NF- $\kappa$ B mediates lipopolysaccharide-induced alternative pre-mRNA splicing of MyD88 in mouse macrophages. *J Biol Chem* (2020) 295(18):6236–48. doi: 10.1074/jbc.RA119.011495
37. Meng R, Li D, Feng Z, Xu Q. MyD88 Hypermethylation Mediated by DNMT1 is Associated With LTA-induced Inflammatory Response in Human Odontoblast-Like Cells. *Cell Tissue Res* (2019) 376(3):413–23. doi: 10.1007/s00441-019-02993-0
38. Šutić M, Motzek A, Bubanović G, Linke M, Sabol I, Vugrek O, et al. Promoter Methylation Status of ASC/TMS1/PYCARD is Associated With Decreased Overall Survival and TNM Status in Patients With Early Stage non-Small Cell Lung Cancer (NSCLC). *Transl Lung Cancer Res* (2019) 8(6):1000–15. doi: 10.21037/tlcr.2019.12.08
39. Wang W, Zhao Z, Wu F, Wang H, Wang J, Lan Q, et al. Bioinformatic Analysis of Gene Expression and Methylation Regulation in Glioblastoma. *J Neurooncol* (2018) 136(3):495–503. doi: 10.1007/s11060-017-2688-1
40. Wang JQ, Jeelall YS, Beutler B, Horikawa K, Goodnow CC. Consequences of the Recurrent MYD88(L265P) Somatic Mutation for B Cell Tolerance. *J Exp Med* (2014) 211(3):413–26. doi: 10.1084/jem.20131424
41. Komori T. Mouse Models for the Evaluation of Osteocyte Functions. *J Bone Metab* (2014) 21(1):55–60. doi: 10.11005/jbm.2014.21.1.55
42. Schmidt-Suppran M, Rajewsky K. Vagaries of Conditional Gene Targeting. *Nat Immunol* (2007) 8(7):665–8. doi: 10.1038/ni0707-665
43. Pillai S, Cariappa A. The Follicular Versus Marginal Zone B Lymphocyte Cell Fate Decision. *Nat Rev Immunol* (2009) 9(11):767–77. doi: 10.1038/nri2656
44. Heise N, De Silva NS, Silva K, Carette A, Simonetti G, Pasparakis M, et al. Germinal Center B Cell Maintenance and Differentiation are Controlled by Distinct NF- $\kappa$ B Transcription Factor Subunits. *J Exp Med* (2014) 211(10):2103–18. doi: 10.1084/jem.20132613
45. Merzianu M, Jiang L, Lin P, Wang X, Weber DM, Vadhan-Raj S, et al. Nuclear BCL-10 Expression is Common in Lymphoplasmacytic Lymphoma/Waldenström Macroglobulinemia and Does Not Correlate With P65 NF- $\kappa$ B Activation. *Mod Pathol* (2006) 19(7):891–8. doi: 10.1038/modpathol.3800609
46. Okabe S, Fukuda S, Kim Y-J, Niki M, Pelus LM, Ohyashiki K, et al. Stromal Cell-Derived factor-1 $\alpha$ /CXCL12-induced Chemotaxis of T Cells Involves Activation of the RasGAP-associated Docking Protein p62Dok-1. *Blood* (2005) 105(2):474–80. doi: 10.1182/blood-2004-03-0843
47. Perrot A, Corre J, Avet-Loiseau H. Risk Stratification and Targets in Multiple Myeloma: From Genomics to the Bedside. *Am Soc Clin Oncol Educ Book* (2018) 38:675–80. doi: 10.1200/EDBK\_200879
48. Bohannon C, Powers R, Satyabhama L, Cui A, Tipton C, Michaeli M, et al. Corrigendum: Long-lived Antigen-Induced IgM Plasma Cells Demonstrate Somatic Mutations and Contribute to Long-Term Protection. *Nat Commun* (2016) 7:12687. doi: 10.1038/ncomms12687
49. Li Z, Wang H, Xue L, Shin D-M, Roopenian D, Xu W, et al. Emu-BCL10 Mice Exhibit Constitutive Activation of Both Canonical and Noncanonical NF- $\kappa$ B Pathways Generating Marginal Zone (MZ) B-cell Expansion as a Precursor to Splenic MZ Lymphoma. *Blood* (2009) 114(19):4158–68. doi: 10.1182/blood-2008-12-192583
50. Moore CR, Liu Y, Shao C, Covey LR, Morse HC, Xie P. Specific Deletion of TRAF3 in B Lymphocytes Leads to B-lymphoma Development in Mice. *Leukemia* (2012) 26(5):1122–7. doi: 10.1038/leu.2011.309
51. Vincent-Fabert C, Saintamand A, David A, Alizadeh M, Boyer F, Arnaud N, et al. Reproducing Indolent B-cell Lymphoma Transformation With T-cell Immunosuppression in LMP1/CD40-expressing Mice. *Cell Mol Immunol* (2019) 16(4):412–4. doi: 10.1038/s41423-018-0197-6
52. Vincent-Fabert C, Roland L, Zimber-Strobl U, Feuillard J, Faumont N. Pre-Clinical Blocking of PD-L1 Molecule, Which Expression is Down Regulated by NF- $\kappa$ B, JAK1/JAK2 and BTK Inhibitors, Induces Regression of Activated B-cell Lymphoma. *Cell Commun Signal* (2019) 17(1):89. doi: 10.1186/s12964-019-0391-x
53. Vincent-Fabert C, Soubeyran I, Velasco V, Parrens M, Jeannot R, Lereclus E, et al. Inflamed Phenotype of Splenic Marginal Zone B-cell Lymphomas With Expression of PD-L1 by Intratumoral Monocytes/Macrophages and Dendritic Cells. *Cell Mol Immunol* (2019) 16(6):621–4. doi: 10.1038/s41423-019-0228-y
54. Reimann M, Schrezenmeier JF, Richter-Pechanska P, Dolnik A, Hick TP, Schleich K, et al. Adaptive T-cell Immunity Controls Senescence-Prone MyD88- or CARD11-mutant B-Cell Lymphomas. *Blood* (2020). doi: 10.1182/blood.2020005244
55. Jalali S, Price-Troska T, Paludo J, Villasboas J, Kim H-J, Yang Z-Z, et al. Soluble PD-1 Ligands Regulate T-cell Function in Waldenström Macroglobulinemia. *Blood Adv* (2018) 2(15):1985–97. doi: 10.1182/bloodadvances.2018021113

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ouk, Roland, Gachard, Poulain, Oblet, Rizzo, Saintamand, Lemasson, Carrion, Thomas, Balabanian, Espéli, Parrens, Soubeyran, Boulain, Faumont, Feuillard and Vincent-Fabert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





## **Fonctions des protéines liant la matrice nucléaire dans l'organisation et l'intégrité du génome des cellules B.**

---

Pour produire des Immunoglobulines variées et affines, les cellules B ont recours à certains remaniements géniques, tels que la commutation de classe et l'hypermutation somatique (SHM). Ces processus, prenant place aux loci des gènes d'Ig, induisent des cassures d'ADN et des mutations. Leur altération peut conduire à l'apparition d'évènements illégitimes qui sont souvent à l'origine de l'émergence de lymphomes. Une surveillance accrue de ces mécanismes s'effectue *via* des régions dites régulatrices des gènes d'Ig dont l'activateur intronique  $E\mu$  qui est encadré de régions de liaison à la matrice nucléaire : les MARs. Notre équipe a constaté l'implication de ces régions dans le processus de SHM et dans le positionnement nucléaire des loci. Pour expliquer ces effets, nous soupçonnions l'intervention de certaines protéines qui ont la capacité de lier ces régions MAR, telles que SATB1 et ARID3A. Par l'étude de modèles murins de délétion conditionnelle de l'une et l'autre de ces protéines, nous démontrons qu'elles contribuent toutes les deux au processus d'hypermutation somatique. Nos études semblent indiquer que SATB1 pourrait limiter l'introduction des mutations alors que ARID3A favoriserait ce processus. Nous montrons également que ARID3A contribue à la différenciation plasmocytaire. Ces données confortent nos études précédentes et placent les régions MARs comme des éléments cis-régulateurs de la SHM. Nos études ouvrent de nouveaux champs d'investigation visant à étudier en détail la fonction de ces deux protéines dans le contrôle de l'hypermutation. La compréhension du mode d'action de ces nouveaux acteurs devrait aider à préciser leur implication dans l'émergence de certaines pathologies du lymphocyte B.

---

Mots-clés : Lymphocytes B, Hypermutation somatique, SATB1, ARID3A, Région MARs $E\mu$

### **MAR binding protein functions in B cells nuclear organization and integrity**

---

In order to produce diverse and affine Immunoglobulins, B cells employ certain gene alterations, such as class switching and somatic hypermutation (SHM). These processes, taking place at Ig gene loci, induce DNA breaks and mutations. Their alteration can lead to illegitimate events that are often at the origin of the emergence of lymphomas. Increased surveillance of these mechanisms is carried out via so-called regulatory regions of Ig genes, including the intronic activator  $E\mu$  which is flanked by nuclear matrix binding regions: MARs. Our team found that these regions are involved in the SHM process and in the nuclear positioning of loci. To explain these effects, we suspected the intervention of certain proteins that have the capacity to bind these MARs regions, such as SATB1 and ARID3A. By studying mouse models of conditional deletion of both proteins, we demonstrate that they both contribute to the somatic hypermutation process. Our studies suggest that SATB1 may limit the introduction of mutations while ARID3A may promote this process. We also show that ARID3A contributes to plasma cell differentiation. These data confirm our previous studies and place MARs regions as cis-regulatory elements of SHM. Our studies open new fields of investigation to study in detail the function of these two proteins in the control of hypermutation. Understanding the mode of action of these new players should help to clarify their involvement in the emergence of certain B cell pathologies.

---

Keywords: B lymphocytes, Somatic Hypermutation, SATB1, ARID3A, MARs $E\mu$  region

