



**HAL**  
open science

# Molecular mechanisms controlling immunoglobulin class switch recombination

Ebe Schiavo

► **To cite this version:**

Ebe Schiavo. Molecular mechanisms controlling immunoglobulin class switch recombination. Genomics [q-bio.GN]. Université de Strasbourg, 2013. English. NNT : 2013STRAJ084 . tel-01147174

**HAL Id: tel-01147174**

**<https://theses.hal.science/tel-01147174>**

Submitted on 29 Apr 2015

**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.

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE  
INSTITUT DE GENETIQUE ET DE BIOLOGIE MOLECULAIRE ET CELLULAIRE

Thèse présentée par :

**Ebe SCHIAVO**

soutenue le : **30 Septembre 2013**

pour obtenir le grade de

**Docteur de l'Université de Strasbourg**

Discipline : Sciences du vivant

Spécialité : Aspects moléculaires et cellulaires de la Biologie

**Molecular mechanisms controlling  
immunoglobulin class switch recombination**

<b>Dr. Bernardo REINA-SAN-MARTIN</b>	Directeur de thèse
<b>Pr. Qiang PAN-HAMMARSTRÖM</b>	Rapporteur externe
<b>Dr. Bertrand NADEL</b>	Rapporteur externe
<b>Dr. Françoise DANTZER</b>	Examineur interne
<b>Dr. Silvestro CONTICELLO</b>	Examineur

*“...though my soul may set in darkness,  
it will rise in perfect light;  
I have loved the stars too truly  
to be fearful of the night.”*  
**Sarah Williams**

*“Amo il tempo avvenuto,  
buono se è stato buono,  
cattivo perché non può tornare.”*  
**Erri De Luca**

# ACKNOWLEDGEMENTS

*The following pages represent a kind of “sum” of these last four years, but this thesis, as well as my work and everything it came along, it wouldn't have been possible without all the people I wish to acknowledge in these pages.*

*First of all, I have to say thank you to Bernardo: for accepting me in your lab in the far October 2009, for your trust and your support, for not allowing me to give up when things were turning from bad to worst; you have been my mentor and my guide and I have learnt so many things during all this time (although you still have to show me how to get 10<sup>10</sup> competence for bacteria!), and I will try to keep the enthusiasm you transmitted me as well as the positive attitude, which makes the impossible to become possible.*

*To Pan, Françoise, Bertrand and Silvo, for accepting to evaluate the work I am presenting in this dissertation.*

*To Hélène, for being part of my mid-thesis committee – as well as Françoise – and for the advices you provided me with.*

*To Susan and Philippe, for making me feel as part of a “huge lab”, for the lab meetings, the journal clubs and the barbecues!*

*To the Reina lab, my “second family”: to Anne-Sophie, with whom I shared the good and the bad time, in particular in this last year made of endless days, crazy sorts, super-high-throughput experiments: we went through all this together, by supporting each other and discovering that two brains are better than one! To Léa, who brought a ray of light in the lab: for bringing me to see Radiohead again, for our “elective affinity” and your infinite sensibility, for all the time we spent together close to or far from the bench. To Vincent, the only man able to tolerate the Reina girls!!! For struggling together with the Necker cells which didn't want to grow, for your “Sicilian accent”, the chocolate you gave me during coffee time to get some energy for the afternoon, and all the galettes de roi you baked (sometimes not really spontaneously!). To Mélanie, the new “cloning queen” of the lab, for her kindness, her Nutella muffins and for our “French lunches” in the CROUS. To Lila, for the good time we had together when she was around, and also to the “unforgettable seniors”!!! Thanks to Beena, for allowing me to discover the wonderful person you are, because since you left anything changed, and also for your help during these last weeks of “writing marathon”; and to Sara, “my wife”, who initiated me to cloning and also to the SPB: for the dances in the lab, the night chats, the “zen sessions” and Christmas decorations in the lab!*

To Isa, even though these words won't be enough to say everything I would like to say. You are not only the "mum" of the lab, but also a sister and a friend to me, one of the most sensible people I've met in my life as well as an amazing scientist; you know that if you are reading these words is also because of you.

To the Chan/Kastner lab, for the good time we had together and the reagents I borrowed! To Attila, for bugging me while waiting for the coffee and also for his advices (and for always appreciating my cakes!), Apos for sharing those problems we finally managed to overcome and for letting me skating on ice again; to the dear Deepika, for everything we shared (and share), for your support and your friendship; to Beate, for her kindness and her advices on B cells development, to Peggy, Rose, Kate, Marie, Isma, Marie-Pierre, Arnaud, for the "pots", the birthdays and the good atmosphere you created.

To the Soutoglou lab: Evi, Tibor, Anne-Sophie, Zita, Audrey, Charlène, Céline, Alkmini and Anastazja: for the time we shared while spending whole days in tissue culture room, for the cakes, the advices, the NCS and H<sup>2</sup>O! And for Nespresso orders (thanks Tibor!!!).

To Shankar, who once told me something such as "if you share the problems, you divide them by two; if you share your happiness, you multiply it by two". This is what you did (and do): share feelings, thought, experience. Thanks so much for being my "living Maniatis", for the night chats in the lab about Fellini movies or Italian football players, for being the friend I can rely on.

To Patricia, for her enthusiasm and her participation to thesis movies, and for her smiles! And to Claudine, for her patience with our impossible sorts, for explaining me 100 times that, at the LSR, I need to "replicate without data" an existing folder to keep the setting, for her amazing cakes and biscuits and for our chats in Italian!

To Marco, for our discussions about DGE, and to Bernard, Serge, Stephanie, Céline, Doulaye and Laurent, for all their help with the DGE and data analysis, and for the interesting discussions we had.

To Christophe and Marie-Laure, for letting me discover the wonderful world of protein co-expression, and for all the attempts we made to make AID soluble!

To Betty and the "angels" of the culture facility: for your kindness, for the last minute orders of medium which saved many experiments of mine and for caring about the Necker cells stock.

To all the people of the common IG BMC services: for making this institute a very nice place to work in; to Bénédicte and Armelle, who cared about the administrative "details" and to Eveline and Maité for their kindness.

To all the present and former members of the IGBMC Students and Post-Docs board: for their enthusiasm, for the time we shared, and for the activities; for showing me that is still possible to make science and have fun! To Ben, Jérôme, Thomas (Tommaso!), David and Adrien, for reaching the end of this path together.

To David, because when he sent me the link for the PhD program call while we were in Cambridge I couldn't imagine that it would have such an impact on my future. To Mr. Grey, because his "sunglassed smile" turned from bad to good a few days of mine. To the volunteers of the SPA, for our nice chats, and especially to the cats and the dogs, hoping they will find a home.

To the "Italian family in Strasbourg": Claudia, Serena, Maria Vittoria, Manuela and Floriana, for being my "safety net" in those tough moments, for the dinners, parties, trips, movies, for the grappa and Cannonau, pan di stelle and aperol spritz. To Manu, my "older sister", and Floriana, because when we met for the interview here in 2009 we couldn't imagine that it was the beginning of a wonderful friendship (and Sicily is waiting for us!!!). To Monica, Pietro and the sweet Edoardo; to Rocco for her enthusiasm and support, to Goffredo, for his "technical advices", and to Angelo, for being able to understand my thoughts without the need of words.

To my friends (and cousins!): Sonia, Emi, Antonia, Raf, Alejandro, Gianlorenzo, Laura, Francesca, Vera, Ale: for their support and their friendship.

To my family: my cousins, aunts and uncles, because no matters where I am or what I do, as I know that you will be there for me, and me for you as well.

To my grandmothers, Ebe and Italia, for their support and their love (and prayers! Thanks!!!) and to my grandfathers, Vincenzo and Domenico, who would have been happy to know that I am writing these pages.

To my beloved sister Irene, for being the "good part of me", for her advices, her support and for her love; to my mum and my dad, for believing in me and reminding me that "everything is going to be ok", for the freedom they always gave me although the distance is hard to accept sometimes. And for knowing that, in the end, I always come back home. To Bizet, Pisolo and Asia, for proving that tolerance between cats and dogs is possible!

# TABLE OF CONTENTS

<b>LIST OF FIGURES .....</b>	<b>3</b>
<b>LIST OF TABLES .....</b>	<b>5</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>6</b>
<b>INTRODUCTION .....</b>	<b>9</b>
<b>I. B cell receptor diversification.....</b>	<b>10</b>
1. The B cell receptor .....	10
2. Antigen-independent Ig diversification: V(D)J recombination .....	11
3. Antigen-dependent Ig diversification .....	16
3.1. Gene conversion.....	16
3.2. Somatic hypermutation .....	17
3.3. Class switch recombination .....	17
<b>II. Activation-induced cytidine deaminase (AID).....</b>	<b>19</b>
1. The RNA editing model.....	20
2. The DNA deamination model.....	20
3. Role of AID in somatic hypermutation.....	21
4. Role of AID in class switch recombination .....	23
4.1. Transcription at the IgH locus .....	23
4.2. Sequence specificity and IgH locus regulatory elements .....	25
4.3. Formation of double stranded DNA breaks at the IgH locus .....	26
4.4. Processing of double stranded DNA breaks: DNA damage response and repair .....	27
4.4.1. "Sensing" the lesions: the DNA damage response (DDR) .....	27
4.4.2. The DSBs repair through non-homologous end joining pathway .....	28
4.4.3. The alternative non-homologous end joining pathway .....	29
5. Role of AID outside the immune system.....	30
6. AID and pathogenesis: B cell lymphomas .....	31
7. AID regulation: how B cells limit AID-mediated DNA damage .....	33
7.1. Transcriptional regulation .....	33
7.2. Post-translational modifications .....	34
7.3. Subcellular localization .....	34
7.4. AID and its cofactors.....	35
<b>III. Class switch recombination-immunodeficiencies.....</b>	<b>39</b>
1. CSR-ID due to a CD40 signaling defect .....	39
1.1. CD40 ligand deficiency .....	40
1.2. Loss of CD40 .....	40
1.3. Impaired activation of NF- $\kappa$ B pathway .....	40
2. CSR-ID due to an intrinsic B cell defect.....	41
2.1. AID deficiency .....	41
2.1.1. AID C-terminal mutations .....	41
2.1.2. AID N-terminal mutations .....	42
2.2. Loss of UNG .....	42
2.3. Deficiency of PMS2 .....	43
2.4. CSR-ID due to a known DNA repair defect .....	43
2.5. CSR-ID due to an unknown defect and associated to normal SHM .....	43
2.5.1. Defect upstream of DSBs: AID CSR-specific cofactor hypothesis .....	44
2.5.2. Defect downstream of DSBs: DNA repair factor hypothesis .....	44

<b>IV. Spt5 and RNA polymerase II: the breakthrough</b> .....	<b>45</b>
1. Spt5: the missing link between AID and transcription .....	46
<b>V. Transcription and chromatin-regulating factors</b> .....	<b>47</b>
1. Spt6: more than a chaperone .....	47
2. The PAF complex: the “transcription platform” .....	48
2.1. Role of the PAF complex in histone modifications and transcription .....	49
<b>VI. The Structural maintenance of chromosomes (Smc) complexes in genome regulation</b> .....	<b>51</b>
1. The cohesin complex .....	52
1.1. Cohesins and Ig loci reorganization.....	54
1.2. Cohesin deficiency and pathological consequences .....	54
2. The Smc5/6 complex .....	55
<b>VII. Working hypothesis</b> .....	<b>58</b>
<b>RESULTS</b> .....	<b>59</b>
<b>I. Overview of thesis work</b> .....	<b>60</b>
<b>II. Identification of CSR-specific factors</b> .....	<b>63</b>
1. Transcriptome profiling of human B cell lines derived from healthy donors, CSR-ID and AID deficient-patients .....	63
2. Comparative proteomic analysis between human and mouse B cell lines to identify CSR-specific AID cofactors.....	74
<b>III. Spt6: the “missing factor” in CSR-ID patients?</b> .....	<b>75</b>
<b>IV. Role of the PAF complex in CSR</b> .....	<b>80</b>
Publication - A role for the RNA pol II-associated PAF complex in AID-induced immune diversification .....	81
<b>V. Smc5: a new potential regulator of CSR</b> .....	<b>98</b>
<b>VI. Smc5/6 complex: is Smc6 required for CSR?</b> .....	<b>103</b>
<b>VII. Role of the cohesin complex in CSR</b> .....	<b>106</b>
Manuscript - The cohesin complex regulates class switch recombination.....	108
<b>DISCUSSION</b> .....	<b>133</b>
1. Patient analysis and transcriptome profiling: the dark side of the approach.....	134
2. Validated or not validated: that is the question .....	135
3. Missing factor: is it really downregulated? .....	136
4. The importance of being within the nucleus.....	137
5. Does AID size and domains matter? .....	138
6. The short story of Spt6.....	139
7. The PAF complex, AID and transcription-associated factors: a complicated relationship.....	141
8. AID targeting: what does it mean? .....	142
9. The Smc5/6 complex: a functional “divorce”? .....	143
10. The cohesin complex in CSR regulation: long-range interactions, repair or both? .....	145
<b>Working model for CSR</b> .....	<b>147</b>
<b>General conclusions</b> .....	<b>149</b>
<b>LITERATURE CITED</b> .....	<b>150</b>
<b>ANNEX I - MATERIALS AND METHODS</b> .....	<b>I</b>
<b>ANNEX II - RÉSUMÉ DE THÈSE</b> .....	<b>VII</b>



# LIST OF FIGURES

## INTRODUCTION

Figure 1. Structure of the B cell receptor (BCR).....	10
Figure 2. Antibody diversification mechanisms.....	11
Figure 3. Organization of the mouse Ig loci.....	12
Figure 4. The steps of V(D)J recombination.....	14
Figure 5. Model of immunoglobulin gene conversion (IGC) at the chicken Igλ locus.....	16
Figure 6. Somatic hypermutation (SHM).....	17
Figure 7. Class switch recombination (CSR).....	18
Figure 8. AID domains organization.....	19
Figure 9. The DNA deamination model.....	21
Figure 10. AID-mediated mutagenesis during SHM.....	22
Figure 11. Class switch recombination reaction.....	23
Figure 12. Formation and repair of DSBs during CSR.....	27
Figure 13. Cellular regulation of AID.....	38
Figure 14. Main roles of the PAF complex.....	49
Figure 15. Domain organization of the Smc protein and Smc complex. ....	51
Figure 16. Cohesin complex.....	52
Figure 17. Smc5/6 complex.....	57

## RESULTS

Figure 18. Workflow of identification and functional characterization of CSR-specific regulators.....	62
Figure 19. Comparison of transcriptomic data obtained from healthy donors, CSR-ID and AID <sup>-/-</sup> EBV-immortalized B cell lines.....	64
Figure 20. DGE data validation by RT-qPCR: genes upregulated and downregulated in patients and controls when compared to AID <sup>-/-</sup> .....	70
Figure 21. DGE data validation by RT-qPCR: genes upregulated and downregulated in patients when compared to controls.....	71
Figure 22. Comparative analysis of deregulated genes in CSR-ID patients.....	73
Figure 23. Strategy for proteome identification of CSR-specific AID partners on human and mouse B cell lines.....	74
Figure 24. Spt6 and Spt5 expression in human B cell lines.....	75
Figure 25. Retrovirus-mediated Spt6 knockdown on CH12 B cells.....	77
Figure 26. Lentivirus-mediated Spt6 knockdown on CH12 B cells might have an effect on CSR.....	78
Figure 27. Smc5 expression is variable in CSR-ID patients.....	98

Figure 28. Lentivirus-mediated Smc5 knockdown induces a partial CSR reduction upon 48h stimulation.....	101
Figure 29. Lentivirus-mediated Smc5 knockdown induces opposite effects on CSR upon 72h stimulation.....	102
Figure 30. Lentivirus-mediated Smc6 knockdown has an impact on CSR upon 48h stimulation.....	104
Figure 31. Lentivirus-mediated Smc6 knockdown affects CSR upon 72h stimulation.....	105

## **DISCUSSION**

Figure 32. AID N-terminal and C-terminal truncations.....	139
Figure 33. Working model for CSR.....	148

# LIST OF TABLES

## INTRODUCTION

Table 1. Class switch recombination immunodeficiencies (CSR-ID) .....	39
Table 2. Cohesin subunits and regulatory factors.....	53
Table 3. Smc5/6 complex subunits and associated non-Smc elements (Nse) .....	57

## RESULTS

Table 4. Characteristics of the human B cell lines analyzed in this study.....	60
Table 5. Deregulated genes in patients when compared to AID <sup>-/-</sup> .....	65
Table 6. Deregulated genes in controls when compared to AID <sup>-/-</sup> .....	66
Table 7. Deregulated genes in patients when compared to controls.....	67
Table 8. List of top ten Ingenuity networks.....	72

## LIST OF ABBREVIATIONS

3'RR	IgH locus 3' regulatory region
53BP1	p53 binding protein 1
A-NHEJ	Alternative non-homologous end joining pathway
<i>Aicda</i>	Activation-induced cytidine deaminase gene (mouse)
<i>AICDA</i>	Activation-induced cytidine deaminase gene (human)
AID	Activation-induced cytidine deaminase
APE	Apurinic/aprimidinic endonuclease
APLF	Aprataxin and PNK-like factor
APRIL	A proliferation-inducing ligand
ATM	Ataxia telangiectasia mutated
BATF	Basic leucine zipper transcriptional factor ATF-like
bp	Base pairs
BCR	B cell antigen receptor
BER	Base excision repair pathway
C-NHEJ	Classical non-homologous end joining pathway
C/EBP	CCAAT/enhancer-binding protein
CD40L	CD40 ligand
CDR	Complementarity-determining region
CE	Coding end
CFSE	Carboxyfluorescein succinimidyl ester
ChIP	Chromatin immunoprecipitation
CSR	Class switch recombination
CSR-ID	Class switch recombination immunodeficiency
CTNBL1	Catenin beta like 1
DDR	DNA damage response
DGE	Digital gene expression-tag profiling
DNA-PKcs	DNA-dependent protein kinase (catalytic subunit)
DSB	Double stranded DNA break
dsDNA	double stranded DNA
DSIF	5,6-dichloro-1- $\beta$ -d-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor complex
EDA-ID	Ectodermal dysplasia associated with immunodeficiency
eEF1A	Elongation factor 1 alpha
ES	Embryonic stem cell
Exo1	Exonuclease 1
FACS	Fluorescence activated cell sorting
FACT	Facilitates chromatin transcription complex
GADD45	Growth arrest and DNA damage-inducible 45
GANP	Germinal center-associated nuclear protein
GC	Germinal centers
GLT	Germline transcription
HIGM	Hyper-IgM syndrome
HoxC4	Homeobox C4
HP1	Heterochromatin protein 1
HR	Homologous recombination
hs	DNAse I hypersensitive sites
Id	Inhibitors of differentiation proteins
IGC	Immunoglobulin gene conversion
IgH	Immunoglobulin heavy chain locus

IgL	Immunoglobulin light chain locus
IL-4	Interleukin-4
IL-7R	Interleukin-7 receptor
IP	Immunoprecipitation
iPSC	Induced primordial stem cell
Iws1	Interacting with Spt6 homolog 1
KAP1	KRAB associated protein 1
Kb	Kilobases
Mb	Megabases
MBD4	Methyl-domain binding protein 4
MDC1	Mediator of DNA damage checkpoint protein 1
MEF	Mouse embryonic fibroblast
MLH1	MutL homolog 1
MMSET	multiple myeloma SET domain-containing protein
MMR	Mismatch repair pathway
MS	Mass spectrometry
MSH2	MutS protein homolog 2
MSH3	MutS protein homolog 3
MSH6	MutS protein homolog 6
MudPIT	Multi-dimensional protein identification technology
Nbs1	Nijmegen breakage syndrome protein 1
NELF	Negative elongation factor
NEMO	NF- $\kappa$ B essential modulator
NES	Nuclear export signal
NHEJ	Non-homologous end joining pathway
Nipbl	Nipped-B-like protein
NLS	Nuclear localization signal
nt	Nucleotide(s)
PAF	RNA polymerase II-associated factor complex
PARP1	Poly(ADP-ribose) polymerase 1
PARP3	Poly(ADP-ribose) polymerase 3
PCNA	Proliferating cell nuclear antigen
PGC	Primordial germ cell
PKA	Protein kinase A
PMS2	Postmeiotic segregation increased ( <i>S. cerevisiae</i> ) 2
PP2A	Protein phosphatase type 2A
PTBP2	Polypyrimidine tract binding protein 2
PTIP	Pax interaction with transcription-activation domain protein 1
RAG	Recombination-activated gene
REG $\gamma$	Proteasome activator complex subunit 3
RPA	Replication protein A
RPC	Replication factor C
RSS	Recombination signal sequences
SE	Signal end
SHM	Somatic hypermutation
Smc1	Structural maintenance of chromosomes 1
Smc2	Structural maintenance of chromosomes 2
Smc3	Structural maintenance of chromosomes 3
Smc4	Structural maintenance of chromosomes 4
Smc5	Structural maintenance of chromosomes 5
Smc6	Structural maintenance of chromosomes 6
Spt4	Suppressor of Ty homolog 4

Spt5	Suppressor of Ty homolog 5
Spt6	Suppressor of Ty homolog 6
Spt16	Suppressor of Ty homolog 16
SSB	Single stranded DNA break
ssDNA	Single stranded DNA
STAT6	Signal transducer and transcription activator 6
TCR	T cell antigen receptor
TGF $\beta$	Transforming growth factor $\beta$
Top I	Topoisomerase I
TSS	Transcription start site
UNG	Uracil-DNA glycosylase
Wapal	Wings apart-like homolog
WT	Wild-type
XLF	XRCC4-like factor (Cernunnos)
XRCC4	X-ray repair cross-complementing protein 4
YY1	Transcriptional repressor protein YY1

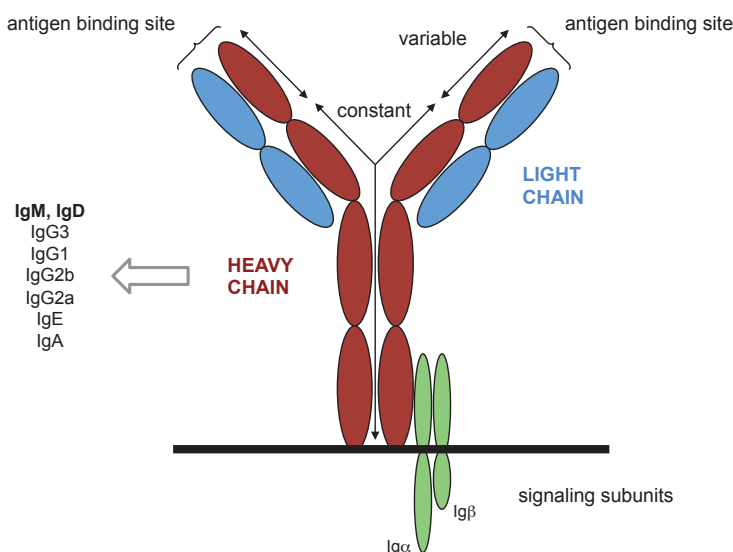
# **INTRODUCTION**

# I. B cell receptor diversification

## 1. The B cell receptor

The B lymphocytes originate in the bone marrow from hematopoietic precursors and, throughout their development, the important role these cells play within the immune response is dependent on the repertoire of B cell antigen receptors (BCRs) expressed in their membrane-bound form or in their soluble form, also known as antibodies (Abs) or immunoglobulins (Igs).

The BCR is a transmembrane protein composed of two identical heavy chains (IgH), two light chains (IgL) and additional subunits Ig alpha (Ig $\alpha$ ) and Ig beta (Ig $\beta$ , Figure 1). Each IgH chain is covalently bound to an IgL chain. In both IgH and IgL chains, the amino-terminal portion represents the variable (V) region of the receptor, responsible for the recognition of the antigen through the complementarity-determining region (CDR), which dictates the affinity and the clonal selection for the cognate antigen. On the other hand, the IgH carboxy-terminal portion represents the constant (C) or invariant region, which defines the isotype expressed (IgM, IgD, IgG, IgA or IgE) and the effector function of the Ig in terms of downstream pathways and responses activated. In mammals, the light chain can be either kappa ( $\kappa$ ) or lambda ( $\lambda$ ), whereas the heavy chain can be  $\mu$  (IgM),  $\delta$  (IgD),  $\gamma$  (IgG),  $\alpha$  (IgA) or  $\epsilon$  (IgE).

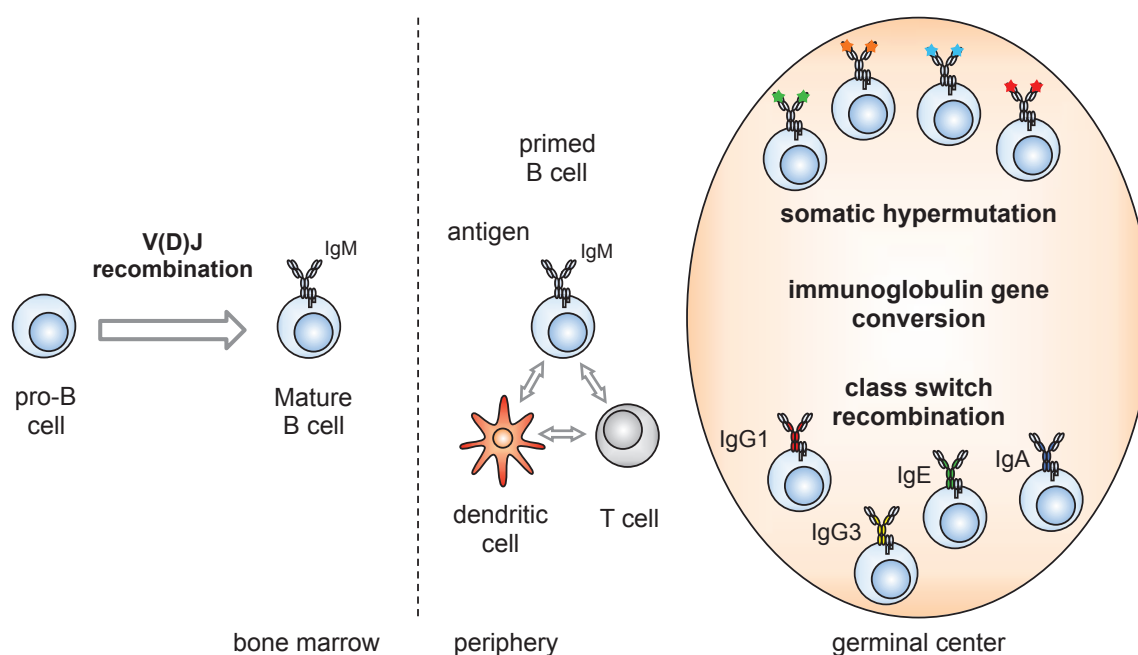


**Figure 1. Structure of the B cell receptor (BCR)**

The B cell receptor is composed of two heavy chains and two light chains, each one harboring a variable (V) region and a constant (C) region. V regions of the heavy and light chains represent the antigen binding site, whereas the C regions represent the isotype expressed and consequently exert a different effector function. The BCR is the membrane-bound form of the antibody (or immunoglobulin) and, upon antigen recognition, the receptor subunits Ig $\alpha$  and Ig $\beta$  mediate the signal transduction, which results in clonal expansion.



The Ab repertoire produced by B lymphocytes is estimated to be higher than  $10^{11}$  and is acquired through genomic rearrangements (recombination and mutation) at the IgH and IgL loci. In particular, four mechanisms have been identified to contribute to Ig diversification: V(D)J recombination, which occurs during the early stages of B cells development prior to antigen encounter; and antigen-dependent mechanisms such as class switch recombination (CSR), somatic hypermutation (SHM) and, in species such as chicken and rabbit, immunoglobulin gene conversion (IGC, Figure 2).



**Figure 2. Antibody diversification mechanisms**

B cells originate in the bone marrow from a hematopoietic precursor and, through V(D)J recombination, they rearrange the V, D and J genes on the Ig heavy and light chains in order to express a functional receptor on their cell surface. Mature B cells migrate to the periphery and, upon cognate antigen recognition and T cell-mediated activation, they start proliferating in the secondary lymphoid organs giving rise to structures named germinal centers. Within germinal centers, B cells further diversify their antibody repertoire through somatic hypermutation, which modifies the affinity for the antigen; class switch recombination, which modifies the antibody isotype expressed and, in some species, through gene conversion.

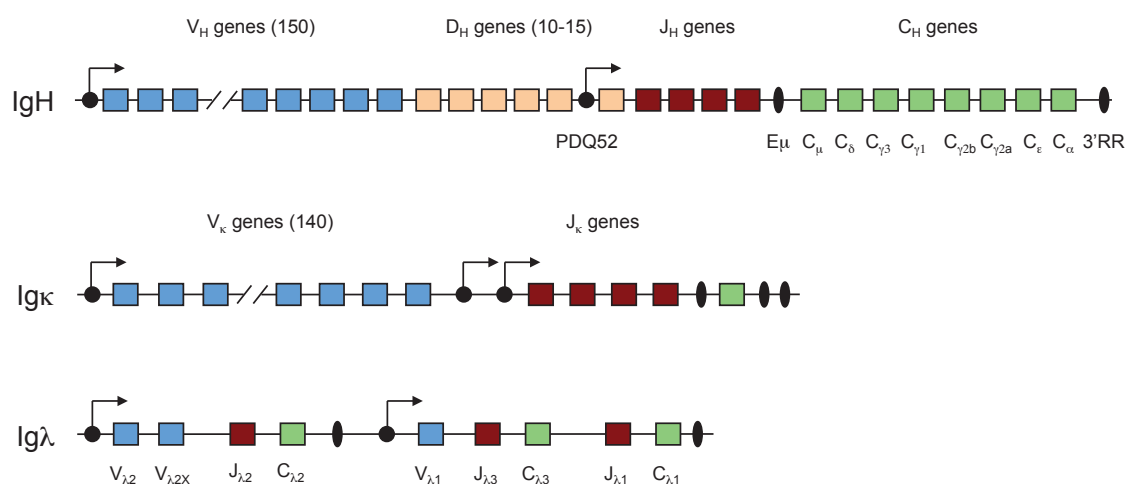
## 2. Antigen-independent Ig diversification: V(D)J recombination

V(D)J recombination is a recombination reaction which assembles the variable region exons of B and T cells antigen receptors (BCRs and TCRs) by providing a high variability in the antigen recognition domain from a limited number of exons. This process assembles the variable (V), diverse (D) and joining (J) gene segments at the Ig loci (IgH, Ig $\kappa$  and Ig $\lambda$ ) and at the TCR loci ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ ). V(D)J recombination reaction is strictly controlled: it is tissue-specific, occurs in primary lymphoid tissues (bone marrow and thymus); it is lineage-specific as Ig and TCR loci are rearranged only in B and T cells respectively and, furthermore, it is stage-specific as the IgH locus is rearranged before the IgL,

as well as occurs for the TCR $\beta$  before TCR $\alpha$ . However, as my work focused on B cells, this dissertation will be centered on this lineage.

The mouse IgH locus is located on chromosome 12 in proximity of the telomere and it spans for almost 3 megabases (Mb) (Chevallard et al., 2002). The V<sub>H</sub> genes are approximately 150 – depending on mouse strain – and classified in 16 families (Johnston et al., 2006); they are located at the 5' of the locus, upstream of 10-15 D<sub>H</sub> gene segments (Retter et al., 2007; Ye, 2004). Then, four J<sub>H</sub> genes precede the constant (C) region exons (C $\mu$ , C $\delta$ , C $\gamma$ 3, C $\gamma$ 1, C $\gamma$ 2b, C $\gamma$ 2a, C $\epsilon$  and C $\alpha$ ), coding for the different Ab isotypes. Within the IgH locus, three main *cis*-regulating elements have been identified: the promoter/enhancer PDQ52, located in proximity of the most 3' D<sub>H</sub> gene; the enhancer E $\mu$  located in the intron between J<sub>H</sub> and C $\mu$  and the 3' regulatory region (3'RR), at the 3' of the IgH locus, which harbors several DNase I hypersensitive sites.

The Ig light chain loci (IgL) display a slightly different organization by presenting only V, J and C gene segments. The  $\kappa$  locus spans over 3 Mb on mouse chromosome 6 and is composed by 140 V $\kappa$  gene segments and 4 functional J $\kappa$  exons, followed by a single C $\kappa$  exon. The light chain  $\lambda$  locus, instead, spans about 200 kilobases (Kb) on chromosome 16 and harbors 3 distinct units composed by V $\lambda$ /J $\lambda$  segments and C $\lambda$  exons (Figure 3).



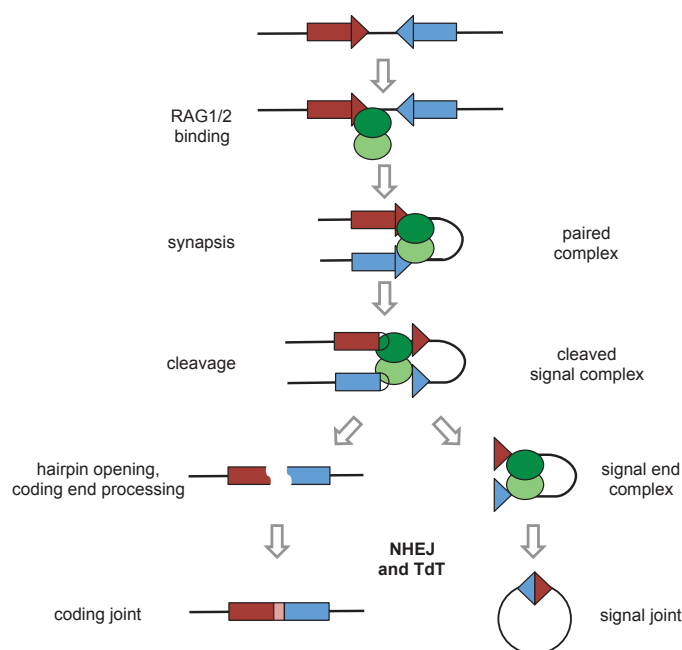
**Figure 3. Organization of the mouse Ig loci**

Schematic representation of the mouse IgH, Ig $\kappa$  and Ig $\lambda$  loci. The IgH locus harbors about 150 variable (V) genes, up to 15 diverse (D) genes, 4 joining (J) genes and 8 constant (C) genes; the Ig $\kappa$  locus displays 140 V genes and 4 J genes, which precede one C exon; the Ig $\lambda$  locus, instead, has a limited number of V exons, organized in distinct cassettes. Diagram not drawn in scale; adapted from Cobb et al., 2006.

V(D)J recombination allows the expression of the rearranged V coding region and of the downstream C region and depends on recombination-activating genes 1 and 2 (RAG1 and RAG2) (Oettinger et al., 1990; Schatz et al., 1989) which code for the RAG recombinase. This site-specific process occurs at recombination signal sequences (RSSs) flanking each gene segment: RSSs are composed by a conserved palindromic heptamer, a spacer sequences of 12 or 23 base pairs (bp) and an A-T rich

nonamer (Sakano et al., 1980). RAGs mediate the DNA cleavage, whereas RSSs dictate the order of the reaction, as the recombination occurs only between exons flanked by RSS harboring a 12 bp and a 23 bp spacer (also known as the 12/23 rule) (Tonegawa, 1983). The presence of 23 bp spacers flanking  $V_H$  and  $J_H$  genes at their 3' and 5' respectively, and of 12 bp spacers located at both ends of the  $D_H$  genes, allows the sequential recombination between D and J exons followed by the V and DJ rearrangement at the IgH locus (Alt et al., 1984).

The recombination is a multistep process and it starts with the expression of RAG1/2 in pre-pro B cells and the recognition of the RSS through the contact between RAG1 and the nonamer sequence (Figure 4) (Swanson and Desiderio, 1998). Then the RAG complex interacts with the heptamer (Swanson and Desiderio, 1999) where it introduces a ssDNA nick in the 12 bp RSS and, followed by synapsis with the 23 bp RSS, it generates a second nick resulting in a double stranded DNA break (McBlane et al., 1995; Schatz and Swanson, 2011). The hydroxyl groups free on both ends interact with the phosphate on the opposite end by generating on one side blunt ends, called signal ends (SEs), and on the other side DNA hairpins, named coding ends (CEs) as they lack RSSs (Gellert, 2002; Roth et al., 1993; Schlissel et al., 1993). The RAG complex is released with the SEs and both the hairpins at the CEs and the blunt ends at SEs are processed by the non-homologous end joining (NHEJ) repair pathway (Taccioli et al., 1994). The first molecular players which act in the repair step are Ku70 and Ku80, which recruit the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs), whose activation results in phosphorylation of target proteins such as Artemis and the histone variant H2AX. Artemis phosphorylation leads to the opening of the hairpins at CEs and the generation of palindromic sequences (named P elements), whereas its nuclease activity is responsible for the random deletion of nucleotides from the opened ends (Lafaille et al., 1989; Ma et al., 2002). This event, as well as the addition of nucleotides at the CEs mediated by the terminal-deoxynucleotidyl transferase (TdT) enzyme (Alt and Baltimore, 1982), further contributes to generate a pool of antibodies harboring a high variability in the V region. The final resolution of the DSBs generated at CEs and SEs is mediated by X-ray repair cross-complementing protein 4 (XRCC4) and the DNA ligase IV (Grawunder et al., 1997; Li et al., 1995), and allows the expression of the upstream V region and of the C exon in IgH and IgL loci. During the early stages of B cell development, pre-pro B cells undergo the first recombination event which occurs at the IgH locus, leading to the expression of the antibody heavy chains which will be combined with the surrogate light chains in order to express the pre-BCR on the cell surface (Hombach et al., 1990). This step is crucial to inhibit the rearrangement of the second allele (allelic exclusion) and in activating the signaling pathway leading to cell proliferation (Jung et al., 2006). The following re-expression of RAGs, the recombination occurring at the light chain loci,  $\kappa$  and  $\lambda$ , and the association of the light chain with the pre-assembled heavy chain will finally lead to the expression of the BCR harboring the  $\mu$  heavy chain on naïve B cells.



**Figure 4. The steps of V(D)J recombination**

The substrates (red and blue rectangles) are flanked by a 12-RSS (red triangle) and by a 23-RSS (blue triangle). The RAG proteins (green ovals) are the central players in the reaction which leads to the cleaved signal complex, and also cooperate with the non-homologous end joining (NHEJ) DNA repair factors in the end processing and joining steps. The final coding joints often contain non-templated nucleotides (pink rectangle) introduced by the TdT. Adapted from Schatz, 2004.

The order and specificity observed during V(D)J recombination raises many questions relative to the tight control of this reaction and, almost thirty years ago, Yancopoulos and Alt proposed the differential DNA accessibility within the locus as crucial feature to direct the RAGs-mediated lesions (Yancopoulos and Alt, 1985). The “accessibility hypothesis” is supported by evidence showing how transcription, 3D relocation within the nucleus and chromatin status influence the Ig and TCR loci during the early stages of the B and T cell development. Germline (non coding) transcription occurs at the Ig and TCR loci prior D-J and V-DJ rearrangements (Corcoran, 2010; Hesslein and Schatz, 2001). At the IgH locus, germline transcripts have been detected before  $D_H$ - $J_H$  recombination, dependent on the activity of the promoter/enhancer PDQ52 and on the intronic enhancer  $E_\mu$ , giving rise respectively to the  $I_\mu$  and  $\mu 0$  transcripts. At the same extent, transcription starting from the promoter interspersed within the V region precedes  $V_H$ - $D_HJ_H$  recombination, and is dependent on interleukin-7 receptor (IL-7R) signaling (Bertolino et al., 2005). Additionally, antisense transcription has been detected throughout the  $V_H$  genes prior V-DJ rearrangement and at the  $D_H$  and  $J_H$  gene segments before  $D_H$ - $J_H$  joining, in this latter case mediated by  $E_\mu$  (Bolland et al., 2007; Bolland et al., 2004; Chakraborty et al., 2007; Perlot et al., 2008). The controversial role of antisense transcription as process which could support the locus availability to RAG-mediated recombination or which could lead to gene silencing by hybridization with the corresponding sense sequences, still has to be fully elucidated.

The chromatin status of the Ig loci plays an important role in dictating the order of the rearrangements. Active histone marks as histone 3 acetylated on lysine 9 (H3K9ac), hyperacetylated histone 4 and histone 3 dimethylated on lysine 4 (H3K4me2) are present in the  $D_H$ - $J_H$  region, around the 5' most  $D_H$

gene and spread over the  $J_H$  genes in early pro-B cells undergoing  $D_H$  to  $J_H$  rearrangements (Chakraborty et al., 2007; Morshead et al., 2003). After this first step of rearrangement, hyperacetylation is detected at V genes as limited to the promoter, the V segment and the RSS (Johnson et al., 2003), and is lost after productive  $V_H$ - $D_H$ - $J_H$  recombination in order to make the locus inaccessible in pro-B cells (Chowdhury and Sen, 2003). Moreover, RAG2 has been shown to bind, through its PHD finger domain, H3K4me3 mark present at the IgH locus and the tryptophan in position 453 (W453) appears to be critical for the binding of the modified H3 histone as well as for efficient V(D)J recombination, thus providing the first direct link between epigenetic modulation of V(D)J recombination and RAGs accessibility (Liu et al., 2007; Matthews et al., 2007). Another feature displayed by the Ig loci undergoing recombination is the differential nucleosome packaging, and DNA sensitivity to restriction enzymes is an indication of accessibility or nucleosome free status. Before  $D_H$  to  $J_H$  rearrangements, the region between PDQ52 and  $E_\mu$  as well as the  $J_H$  RSS are DNase I sensitive, whereas the accessibility of the  $V_H$  regions is limited to the V-D $J_H$  rearrangement (Chowdhury and Sen, 2003; Maes et al., 2006). Methylation of cytosines at the CpG dinucleotides is an additional mechanism of gene silencing in mammals (Stein et al., 1982; Vardimon et al., 1982) and, at the Ig loci, modulates V(D)J recombination as well as allelic exclusion. Additionally, locus relocation and contraction are a hallmark of active recombination. In pro-B cells, the IgH locus is actively relocated towards the center of the nucleus (Kosak et al., 2002), and is regulated by Pax5, a transcription factor required for B cell commitment (Busslinger, 2004), which mediates locus contraction and distal  $V_H$ -D $J_H$  rearrangements (Fuxa et al., 2004).

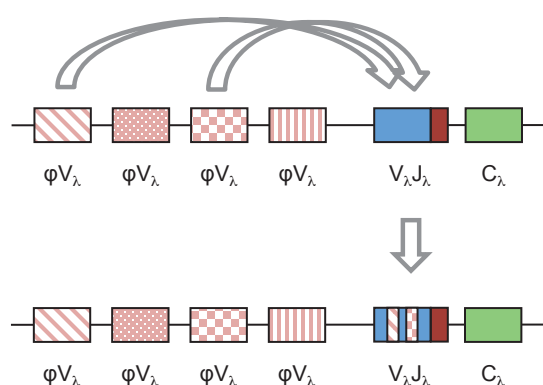
The productive V(D)J rearrangement at the IgH locus allows the expression of the V gene “chosen” during the recombination and of the downstream C exon; in primary B cells, the expression of the rearranged heavy chain and light chain ( $\kappa$  or  $\lambda$ ) leads to the exposure on cell surface of a functional IgM antibody which defines the primary repertoire. Although, the recombination being a random process, self-reactive antibodies can be generated and B cells expressing those antibodies are negatively selected in the bone marrow before entering the peripheral lymphoid organs (Wardemann et al., 2003). By migrating to the periphery, primary B cells will undergo further Ig genes diversification which is dictated by the antigen recognition and T cells-mediated activation, and which will adapt the response in order to provide a faster and more efficient antigen clearance.

### 3. Antigen-dependent Ig diversification

Naïve B cells that exit the bone marrow display a huge repertoire of antibodies that are potentially able to recognize every antigen. However, being the antigen binding of low affinity, upon antigen recognition and T cell-mediated activation, B cells undergo further Ig diversification mechanisms in order to modulate the immune response according to the pathogen and to the stimuli received. They actively proliferate in the secondary lymphoid organs and form structures named germinal centers (GC). Three additional Ig diversification mechanisms have been identified: immunoglobulin gene conversion, which occurs in some species such as chicken and rabbit, and is characterized by the use of pseudogenes to diversify the V(D)J sequence; somatic hypermutation, which introduces mutations and deletions in the V region of the Ig heavy and light chains and class switch recombination, which replaces the antibody isotype expressed while preserving the antigen specificity of the antibody. Overall, these mechanisms contribute to a more specific and adapted immune response.

#### 3.1. Gene conversion

The antibody repertoire generated upon V(D)J recombination is dependent on the choice of V, D and J gene segments assembled during the reaction. But whereas mammals produce a highly diversified repertoire, some species such as the chicken do not. The chicken IgL locus displays unique  $V_L$  and  $J_L$  exons, as well as the IgH locus and in consequence B cell precursors provide a limited specificity. Thus, antibody diversity occurs through a mechanism called immunoglobulin gene conversion (IGC), based on the usage of pseudogenes located upstream of the  $V_L$  and  $VD_H$  exons to replace the coding sequence of rearranged loci (McCormack et al., 1991; Reynaud et al., 1985; Reynaud et al., 1987; Reynaud et al., 1991; Reynaud et al., 1989; Thompson and Neiman, 1987). The lesions introduced in the V genes are repaired through homologous recombination (HR) by using the upstream pseudogenes as template.

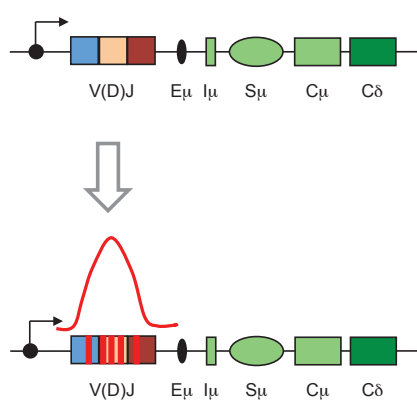


**Figure 5. Model of immunoglobulin gene conversion (IGC) at the chicken  $Ig\lambda$  locus**

Schematic representation of the chicken  $Ig\lambda$  locus undergoing immunoglobulin gene conversion. Chickens harbor only one V and J exon and diversify their Ig repertoire by using upstream  $V\lambda$  pseudogenes (indicated as  $\phi$ ) as template. Adapted from Maizels, 2005.

### 3.2. Somatic hypermutation

Somatic hypermutation introduces point mutations, and occasionally insertions and deletions, in the V genes of the Ig heavy and light chain loci and, by modifying the CDR, this mechanism gives rise to a higher affinity antibody repertoire (Saribasak and Gearhart, 2012). Mutations occur at a frequency of  $10^{-5}$ - $10^{-3}$  mutations/bp/generation, definitely higher than the basal level of mutations in the genome which is estimated being approximately  $10^{-9}$  (Peled et al., 2008), and the mutations extend about 1-2 Kb downstream of the promoter (Saribasak and Gearhart, 2012). B cells expressing high affinity antibodies which do not recognize self-antigens undergo proliferation and further differentiate into plasma cells or memory B cells (LeBien and Tedder, 2008).



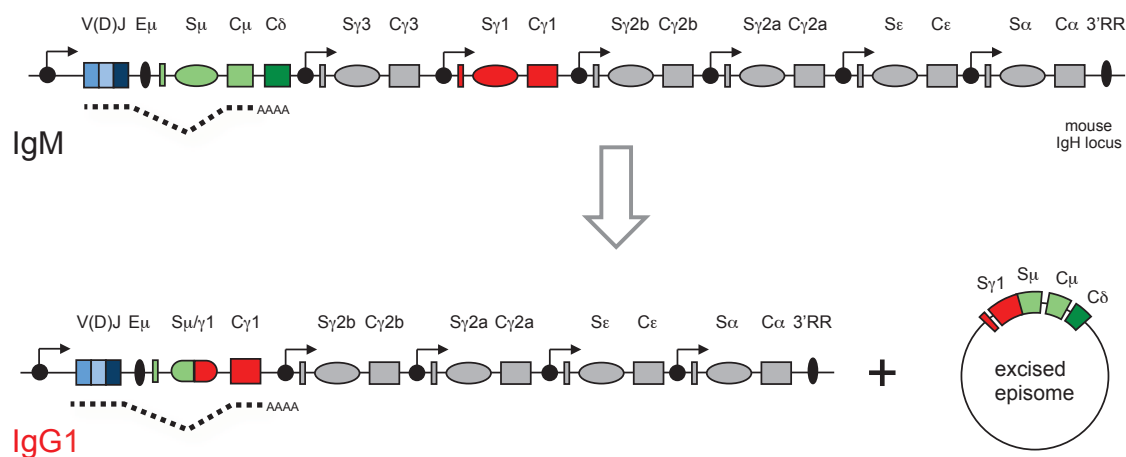
**Figure 6. Somatic hypermutation (SHM)**

Somatic hypermutation modifies the rearranged V genes of the Ig heavy and light chain in order to express antibodies with higher affinity for the cognate antigen. The V genes are diversified by the introduction of non-templated point mutations, insertions and deletions. Mutation frequency is higher closer to the promoter spanning up to 2 Kb downstream and gradually decreases, as shown by the red peak. Diagram not in scale.

### 3.3. Class switch recombination

Class switch recombination is a region-specific recombination reaction that occurs at the IgH locus and joins two switch (S) regions, by deleting the intervening sequence and replacing the isotype of the antibody expressed (from IgM/IgD to IgG, IgE or IgA; Figure 7). This mechanism allows a different effector function of the antibodies while preserving their antigen specificity (Chaudhuri et al., 2007). The recombination takes place at the switch regions (S regions), which are repetitive and non-homologous regions of 3-12 Kb located upstream of each C exon (with the exception of C $\delta$ , whose expression occurs by alternative RNA splicing). Each switch region is preceded by a promoter, and transcription at the donor (S $\mu$ ) and acceptor (S $\gamma$ , S $\epsilon$  or S $\alpha$ ) S regions is activated prior to recombination and mediated by helper T cells and cytokine stimuli (Chaudhuri et al., 2007; Stavnezer et al., 2008b). The introduction of lesions in the DNA and the intermediate DSBs generated are then repaired and allow the expression of the downstream C $_H$  gene after the excision of the intervening

region (Chaudhuri et al., 2007). As the S regions are located far apart within the IgH locus, has been proposed that long-range interactions could contribute to class switching, as the generation of a loop could place in close proximity the donor and acceptor S regions allowing the recombination (Kenter et al., 2012).



**Figure 7. Class switch recombination (CSR)**

Class switch recombination is a region-specific recombination reaction that takes place at the IgH locus. The recombination involves the S regions, repetitive and non-homologous sequences located upstream of each C $_H$  exon. As depicted in the figure, the donor (S $\mu$ ) and acceptor (S $\gamma_1$ ) S regions are involved in the recombination reaction, which finally results in the expression of a different antibody isotype (IgG1) and consequently a different effector function exerted, while preserving the antigen specificity.

More than ten years ago, the work performed by Anne Durandy and Tasuku Honjo led to the identification of activation-induced cytidine deaminase (AID) and represented a breakthrough in the understanding of the CSR, SHM and IGC mechanisms. AID was identified in humans and mice, respectively, as the factor able to mediate CSR and SHM (Muramatsu et al., 2000; Revy et al., 2000) and, from further studies, IGC (Arakawa et al., 2002). These findings allowed a completely new point of view on the regulation of these physiological mechanisms and, most importantly, on their misregulation.

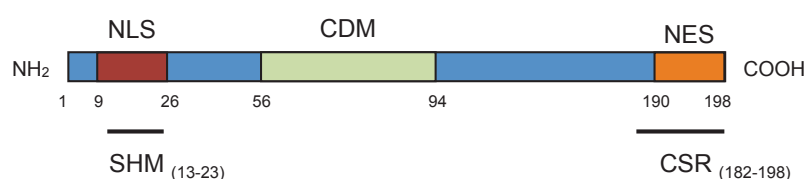


## II. Activation-induced cytidine deaminase (AID)

Activation-induced cytidine deaminase (AID) was identified more than a decade ago through a subtractive cDNA screen performed on mouse CH12 B cells unstimulated and stimulated to undergo CSR (Muramatsu et al., 1999). Furthermore, the characterization of patients affected by class switch recombination-immunodeficiency due to a loss of AID (CSR-ID, also known as type 2 hyper-IgM syndrome), as well as of mice deficient of AID, clarified its role as master regulator of CSR and SHM in B cells (Muramatsu et al., 2000; Revy et al., 2000). Moreover, two years later, further studies described AID as required for immunoglobulin gene conversion in the chicken DT40 B cell line (Arakawa et al., 2002; Harris et al., 2002). AID is a small protein of 198 amino acids, which harbors a nuclear localization signal (NLS) at the N-terminus, a nuclear export signal (NES) at the C-terminus and a cytidine deaminase motif (Figure 8).

AID belongs to the APOBEC family of deaminases, which includes also APOBEC1, APOBEC2 and APOBEC3 subgroups. APOBECs deaminate cytidine in RNA and/or cytosine residues in DNA and regulate different mechanisms. APOBEC1 deaminates the cytidine 6666 on apolipoprotein B RNA, by creating an in frame stop codon which leads to the expression of a shorter RNA (Navaratnam et al., 1993; Teng et al., 1993). APOBEC2 function has still not been clarified, whereas APOBEC3G and APOBEC3F act in the innate immunity protection by retroviruses (Rosenberg and Papavasiliou, 2007). Although AID displays the highest homology to APOBEC1, phylogenetic sequence analysis showed that AID and APOBEC2 are the most ancient members of the family, and that APOBEC1 and APOBEC3 appeared later and are restricted to mammals (Conticello et al., 2005).

Studies based on sequence analysis contributed to clarify the functional domains of AID whereas mutagenic analysis gave insights into AID's function. In particular, its NLS and NES not only determine the subcellular localization (Ito et al., 2004; Patenaude et al., 2009) but also play a pivotal role in SHM and CSR. While loss of the AID N-terminal domain specifically impairs SHM (Shinkura et al., 2004), the C-terminal domain has been shown to be specifically required for CSR (Barreto et al., 2003; Durandy et al., 2007; Geisberger et al., 2009; Imai et al., 2005; McBride et al., 2004; Ta et al., 2003).



**Figure 8. AID domains organization**

*Schematic representation of the domain structure of AID. The protein harbors a nuclear localization signal (NLS) at the N-terminus (NH<sub>2</sub>), a nuclear export signal (NES) at the C-terminus (COOH) and a cytidine deaminase motif (CDM). Many studies have shown that the N-terminus (13-23) and the C-terminus (182-198) are specifically required for SHM and CSR, respectively. Adapted from Muramatsu et al., 2007.*

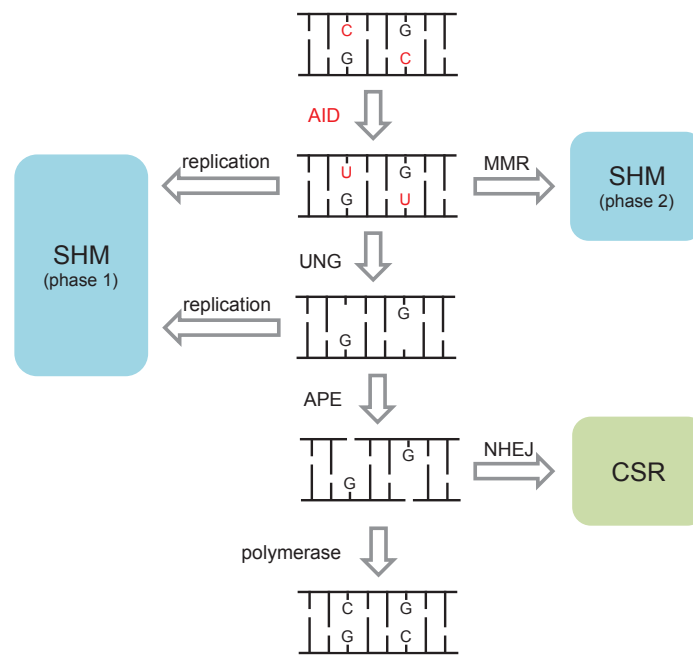
## 1. The RNA editing model

The sequence homology between AID and APOBEC1 allowed a better understanding of the functionality of its domains but, somehow, also contributed to a misleading interpretation of its mechanism of action. As APOBECs are RNA editing enzymes, AID was initially proposed to be an RNA editing enzyme, in light of its ability to regulate two independent mechanisms, such as CSR and SHM. As AID shares 34% homology with APOBEC1 (Muramatsu et al., 1999), it was hypothesized – according to the RNA editing model - that it could modify a putative mRNA precursor coding for a recombinase, which could cleave the DNA on Ig genes to lead to isotype switching and mutations on the V regions resulting in somatic hypermutation; moreover, the specificity of the editing would be dependent on an unknown AID cofactor (Muramatsu et al., 2000). This hypothesis was supported by similarities between AID and APOBEC1: a) APOBEC1 requires APOBEC1 complementation factor (ACF) for its function (Mehta et al., 2000); b) the subcellular localization, with APOBEC1 able to shuttle between nucleus and cytoplasm (Chester et al., 2003); c) APOBEC1 homodimerization (Chester et al., 2003; Teng et al., 1993) which has been predicted for AID by sequence analysis and d) the evidence that *de novo* protein synthesis is required for CSR (Begum et al., 2004; Doi et al., 2003; Muramatsu et al., 1999). However, as it will be presented in the next section, genetic studies supported an alternative model based on DNA deamination, and Fritz et al. recently showed by RNA-Seq analysis that AID is not editing polyadenylated RNA in activated B cells (Fritz et al., 2013).

## 2. The DNA deamination model

The DNA deamination model suggests that AID is able to act on the DNA, and proofs in support of this model come from the experiments performed by Petersen-Mahrt and Neuberger in *E. coli*, where they found that the target nucleic acid of the enzymatic reaction mediated by AID was indeed the DNA and not the RNA. Furthermore, the deamination and mutation profile associated to a loss of uracil-DNA glycosylase (UNG) contributed to dissect the mutagenesis and consequent DNA repair mechanisms that lead to recombination and mutation (Petersen-Mahrt et al., 2002). The DNA deamination model obtained further support from the work performed by other laboratories, which showed that AID is able to deaminate cytosines on ssDNA *in vitro* (Bransteitter et al., 2003; Chaudhuri et al., 2003; Dickerson et al., 2003; Morgan et al., 2004; Pham et al., 2003; Yu et al., 2004) and that it associates with S regions in cells undergoing CSR through interaction with replication protein A (RPA) (Chaudhuri et al., 2004; Nambu et al., 2003).

AID-mediated deamination converts cytosines to uracil residues on DNA, thus introducing a dU:dG mismatch. At this initial step of the reaction, the different options to resolve these lesions can result in SHM and CSR (Figure 9).

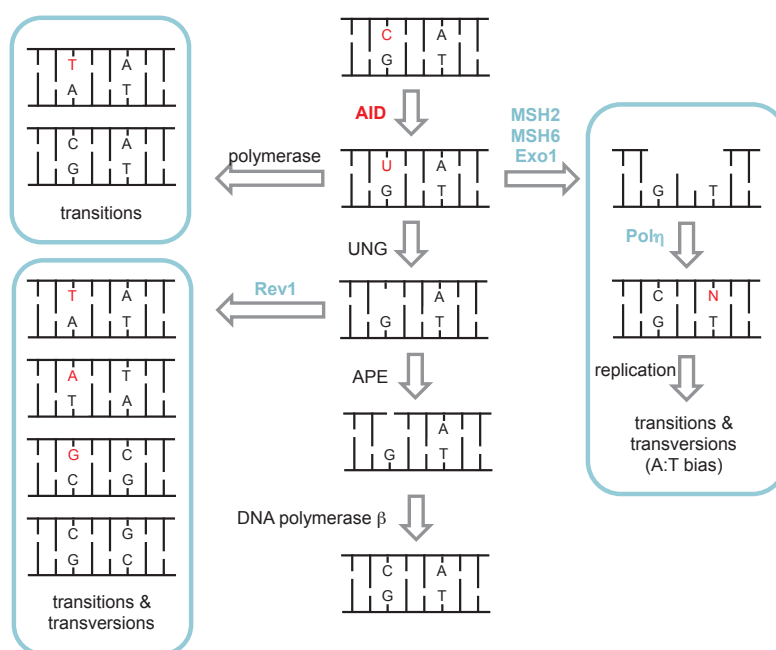


**Figure 9. The DNA deamination model**

*AID deaminates the cytosine to uracil at the V genes of the Ig heavy and light chain and at the S regions of the heavy chain, introducing a mismatch in the primary sequence. The lesion can be processed by the base excision repair pathway (BER): UNG generates an abasic site by removing the uracil whereas APE induces a nick into the DNA, allowing the DNA polymerase  $\beta$  to successfully repair the lesion. If replication occurs after deamination over the dU:dG mismatch or after UNG base removal, a mutagenic profile can be identified, as occurring during SHM. Additionally, the dU:dG mismatch can be processed by the mismatch repair pathway (MMR) which introduces biased mutations at the A:T pairs. If deamination occurs at cytosines located in close proximity, double stranded DNA breaks, intermediates of the CSR reaction, can be generated and repaired by the non-homologous end joining (NHEJ) pathway. Adapted from Petersen-Mahrt et al., 2002.*

### 3. Role of AID in somatic hypermutation

Somatic hypermutation occurs in germinal center B cells upon cognate antigen recognition. According to the deamination model, cytosine deamination mediated by AID introduces the mismatch dU:dG at the Ig genes; although no consensus sequences have been described for AID targeting, deamination occurs at “hot spots” which correspond to the sequence WRCY (W=A/T; R=A/G; Y=C/T) which have been identified in the CDR of the antigen binding site (Betz et al., 1993; Rogozin and Kolchanov, 1992; Sharpe et al., 1991). Furthermore, has been observed that at the Ig V genes mutations are more frequent about 100-200 bp downstream of the transcription start site (TSS) and span up to 1.5-2 Kb downstream of the promoter (Lebecque and Gearhart, 1990; Rada and Milstein, 2001) and that AID is able to mutate both DNA strands (Rada et al., 2004; Shen, 2007; Xue et al., 2006).



**Figure 10. AID-mediated mutagenesis during SHM**

During SHM, AID deaminates the cytosine to uracil at the V genes of the Ig heavy and light chain and introduces a dU:dG mismatch. If replication occurs over the deaminated DNA through a high fidelity polymerase, transitions are introduced in the DNA sequence. Alternatively, replication occurring over the abasic site generated by UNG will introduce transitions and transversions. On the other hand, the processing of the dU:dG mismatches mediated by MSH2, MSH6 and Exo1, components of the MMR pathway, leads to the removal of the strand harboring the mismatch which will be resynthesized by the DNA polymerase  $\eta$ , resulting in the introduction of biased transitions and transversions. Adapted from Petersen-Mahrt et al., 2002.

The uracil introduced by AID deamination can lead to a non-mutagenic profile if the base excision repair (BER) is involved in the resolution of the lesion. Upon uracil excision by uracil-DNA glycosylase (UNG), the apurinic/aprimidinic endonuclease (APE) introduces a single stranded DNA break (SSB) in the corresponding abasic site. Successful repair is mediated by the DNA polymerase  $\beta$  which is able to resynthesize the DNA strand (Figure 10).

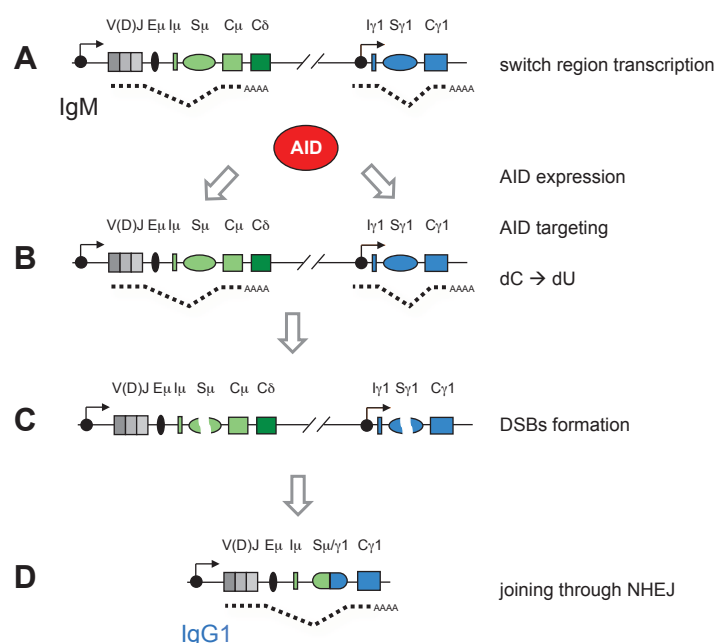
Otherwise, the dU:dG mismatches can be processed through two different pathways, defined as phase 1 and phase 2, according to the mutation profile displayed by the targeted sequence (Petersen-Mahrt et al., 2002; Rada et al., 1998). In the phase 1, replication over the deaminated DNA sequence through a high fidelity polymerase leads to a biased mutation towards transitions C→T or G→A (Figure 10) (Petersen-Mahrt et al., 2002); alternatively, replication can occur over the abasic site generated by UNG and will introduce transitions as well as transversions (replacement of a purine with a pyrimidine and *vice versa*), which can be mediated by Rev1 polymerase (Figure 10) (Jansen et al., 2006; Petersen-Mahrt et al., 2002; Zan et al., 2012).

The phase 2 of mutagenesis, instead, is mediated by MutS protein homolog 2 and 6 (MSH2/MSH6) and processed through the mismatch repair pathway (MMR), which introduces a biased mutagenic profile as the absence of MSH2 or MSH6 lead to an altered mutagenesis and reduced numbers of A:T mutations (Bertocci et al., 1998; Frey et al., 1998; Li et al., 2004; Martomo et al., 2004; Phung et al.,

1998; Rada et al., 1998; Wiesendanger et al., 2000). The MSH2/MSH6 heterodimer recruits the exonuclease 1 (Exo1), which removes the strand harboring the mismatch, and then the DNA will be resynthesized by the DNA polymerase  $\eta$  (Di Noia and Neuberger, 2007). In conclusion, SHM modifies the coding regions located at the antigen binding site and B cells expressing high affinity antibodies will be positively selected for cognate antigen recognition.

## 4. Role of AID in class switch recombination

Class switch recombination is a multi-step process which takes place in germinal centers upon B cell activation, which replaces the antibody isotype expressed and thus the effector functions according to the stimuli received. In this section I am going to dissect the CSR mechanism and to discuss each step of the reaction (Figure 11).



**Figure 11. Class switch recombination reaction**

Class switch recombination is a multistep process which starts with (A) germline transcription at the donor and acceptor S regions (black dotted lines). (B) Upon AID expression and targeting to the DNA, the deamination of the cytosines in the DNA sequence will introduce a mismatch, which will be (C) processed into DSBs through the BER or MMR pathway. (D) The DSBs will be repaired through the NHEJ pathway, resulting in the expression of a different antibody isotype (from IgM to IgG1).

### 4.1. Transcription at the IgH locus

The IgH locus displays transcription units, composed of an I exon, a S region and a downstream C exon coding for the different isotypes. S regions are repetitive and non-homologous regions, located upstream of each C exon, with the exception of C $\delta$ ; they contain GC-rich sequences which display different length and sequence similarity between each other. Primary transcripts generated along the

locus are spliced to remove the sequence corresponding to the S region and polyadenylated (Chaudhuri and Alt, 2004). Why transcription should be important for CSR? Three roles have been proposed so far: first, it provides the substrate for AID-mediated cytosine deamination, as predicted by the “accessibility model” (Stavnezer-Nordgren and Sirlin, 1986; Yancopoulos et al., 1986). Strikingly, ssDNA exposure would occur during mRNA elongation and through the formation of R-loops, structures represented by an RNA:DNA hybrid and the displaced non-template ssDNA (G-rich), which is identical to the newly-synthesized RNA (Aguilera and Garcia-Muse, 2012). Although original hypotheses described AID as able to target only the non-template strand, further studies showed that deamination occurs instead on both strands of DNA (Rada et al., 2004; Shen, 2007; Xue et al., 2006), and this is supported by the recent involvement of the RNA exosome in the process of switching (Basu et al., 2011).

A second role proposed for germline transcription requirement during CSR involves the targeting of AID. Initial studies both *in vitro* or aimed to characterize the phenotype of patients affected by CSR-ID have emphasized the specific requirement of the C-terminal domain for CSR (Barreto et al., 2003; Durandy et al., 2007; Geisberger et al., 2009; Imai et al., 2005; McBride et al., 2004; Ta et al., 2003), giving rise to the “CSR cofactor(s) hunt”. Furthermore, the study from Nambu and co-workers showed that AID co-immunoprecipitates with the RNA polymerase II in splenocytes (Nambu et al., 2003) and the identification of the transcription elongation factors Spt5 and Spt6 as CSR regulators (Okazaki et al., 2011; Pavri et al., 2010) supported the importance of transcription as process which involves the recruitment of AID cofactors; however, as the identification of CSR-specific AID interactors is one of the main points of my dissertation, I am going to discuss it in detail later.

The third and last consequence of germline transcription at the IgH locus is represented by chromatin remodeling through histone post-translational modifications. The donor S $\mu$  region presents numerous activating histone marks, such as H3K9ac/K14ac, H3K27ac, H4ac, and H3K4me3 and H3K36me3 detected even in naïve B cells (Chowdhury et al., 2008; Daniel et al., 2010; Kuang et al., 2009; Nambu et al., 2003; Stanlie et al., 2010; Wang et al., 2006a; Wang et al., 2009; Yamane et al., 2011) and which suggest that S $\mu$  is in a state which constitutively allows recombination. Upon activation, the acceptor S region is made accessible by removal of repressive marks such as H3K27me3 (Chowdhury et al., 2008) and active transcription allows the recruitment of factors which modify the histones to make the DNA accessible for recombination, as occurs with the facilitates chromatin transcription complex (FACT) complex (SSRP1/Spt16) (Stanlie et al., 2010). The discovery that stalled RNA polymerase II and Spt5 association with AID is required for CSR further supports this model, leading the histone modifying enzymes to be “carried” by the RNA polymerase II and to exert their function (Li et al., 2013; Pavri et al., 2010). Moreover, H3ac and H3K4me3 are enriched at the acceptor S regions (Wang et al., 2006a; Wang et al., 2009), while H3K9me3 is still present at S $\mu$  region, and recruits KRAB domain associated protein 1 (KAP1)/ heterochromatin protein 1 (HP1) complex to the IgH locus, which will tether AID to the donor S region (Jeevan-Raj et al., 2011). Furthermore, histone modifications are also important to recruit DNA repair proteins such as 53BP1, which is stabilized at DSBs by H3K20me2 and is pivotal for DSBs resolution, as loss of the methyl

transferase multiple myeloma SET domain-containing protein (MMSET) impairs 53BP1 recruitment and results in defective CSR (Pei et al., 2013).

#### **4.2. Sequence specificity and IgH locus regulatory elements**

As transcription *per se* has been discussed being indispensable for CSR, the sequences of the S regions and of the IgH locus regulatory elements have been extensively investigated in order to identify any consensus which would justify why AID is extensively deaminating cytosines at the Ig loci and not at all the transcribed genes in the cell at that particular developmental stage.

The I exon promoters integrity is required for efficient CSR (Bottaro et al., 1994; Harriman et al., 1996; Jung et al., 1993; Seidl et al., 1998; Zhang et al., 1993) as well as the presence of S regions (Daniels and Lieber, 1995b; Kinoshita et al., 1998; Lepse et al., 1994; Leung and Maizels, 1992; Ott and Marcu, 1989; Petry et al., 1999; Stavnezer et al., 1999). On the other hand, replacement of the S $\alpha$  sequence with the S $\gamma$ 1 or S $\epsilon$  in the CH12 B cells which switch to IgA does not impair CSR efficiency (Kinoshita et al., 1998). Moreover, the observation that in *Xenopus laevis* AT-rich S regions support CSR suggests that the G content within these sequences is not a limiting factor, whereas the fact that they are palindromic it can be, by supporting the secondary structures generated during transcription (Tashiro et al., 2001).

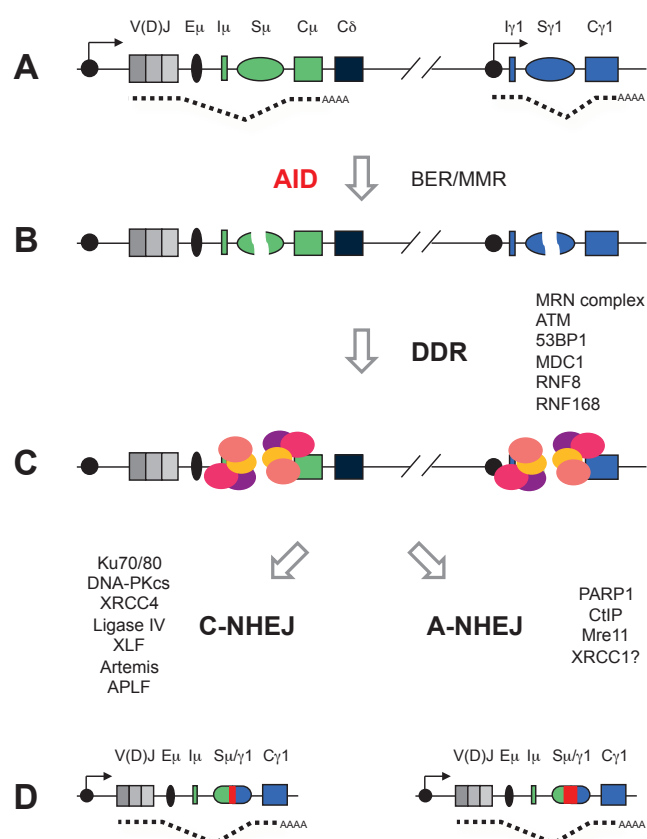
The IgH locus contains two enhancer elements: the intronic enhancer E $\mu$  and the 3'RR. E $\mu$  is located between the J $\mu$ 4 exon and the 5' of S $\mu$  and its targeted deletion reduces CSR (Sakai et al., 1999), although a debate concerning the system used (knockout of the enhancer core, which contains the I $\mu$  promoter) questions whether or not the enhancer itself or more specifically the promoter deletion in this experimental system leads to a reduced recombination efficiency. The 3'RR, instead, is located downstream of the C $\alpha$  exon, it spans for approximately 40 Kb and is composed of DNase I hypersensitive sites (hs): hs3A, hs1,2, hs3B, hs4, hs5, hs6 and hs7. While deletion of hs1,2 and hs3A does not affect CSR, the loss of hs3B and hs4 reduces GLT levels and consequently recombination to all the isotypes except IgG1 (Manis et al., 1998b; Pinaud et al., 2001). Furthermore, the hs3B and hs4 have been involved in the regulation of the locus rearrangement required to bring into close proximity the donor and acceptor S regions involved in the recombination (Wuerffel et al., 2007). Additionally, the proper splicing of germline transcripts has also been shown to be required for efficient CSR, as deletion of splicing donor and acceptor sites impairs efficient recombination (Hein et al., 1998; Lorenz et al., 1995).

### **4.3. Formation of double stranded DNA breaks at the IgH locus**

According to the DNA deamination model, the uracils introduced into the DNA upon AID-mediated deamination can be processed by either the base-excision repair or the mismatch repair pathway, which will lead to a break into the DNA. The involvement of the BER components has been evidenced by loss-of-function experiments and from the study of patients harboring mutations in the *UNG* gene, pointing out the importance of the uracil excision into the mutagenic profile of CSR and SHM (Imai et al., 2003b; Rada et al., 2002b; Schrader et al., 2005). The single nucleotide gap created by APE is then filled by the DNA polymerase  $\beta$  (Stavnezer et al., 2008a) with efficient repair, as deficiency of the polymerase has been shown to increase CSR (Wu and Stavnezer, 2007).

The result of the BER processing is usually a high fidelity repair of the lesion, although the single strand DNA breaks generated by APE, if occurring on both strands and in close proximity, may lead to DSBs and favor switching. However, this might not occur at a frequency that could sustain massive recombination at the S regions, implying a parallel pathway in the formation of DSBs at S regions. The mismatch repair pathway has been implicated in inducing DSBs at the IgH locus as well as a biased mutagenic profile on V genes at the IgH and IgL loci (Chahwan et al., 2012). The MSH2/MSH6 heterodimer recognizes dU:dG mismatches into the DNA and recruits MutL homolog 1 (MLH1) and postmeiotic segregation increased 2 (PMS2); the complex then recruits replication factor C (RPC), the proliferating cell nuclear antigen (PCNA) and Exo1 for excision of the single stranded DNA patch containing the mismatch, which will be then resynthesized by low-fidelity polymerases such as DNA polymerase  $\eta$  (Chahwan et al., 2012). Deficiency of any of the MMR molecular players (MSH2, MSH6, MLH1, PMS2 or Exo1) leads to a reduction in CSR and/or SHM, with the exception of MSH3 which does not seem to be involved in dU:dG mismatches processing at the Ig loci (Bardwell et al., 2004; Ehrenstein and Neuberger, 1999; Ehrenstein et al., 2001; Li et al., 2004; Martin et al., 2003; Martomo et al., 2004; Schrader et al., 1999).





**Figure 12. Formation and repair of DSBs during CSR**

(A) During CSR germline transcription at the donor ( $S_{\mu}$ ) and acceptor ( $S_{\gamma 1}$ ) S regions leads to the exposure of ssDNA; (B) AID targeting and cytosine deamination introduces a dU:dG mismatch that can be processed by the BER or MMR pathway and lead to the formation of DSBs. (C) The breaks are recognized by the MRN complex (Mre11, Rad50 and Nbs1) which binds the breaks and allows the recruitment of ATM which phosphorylates Nbs1, 53BP1, MDC1, KAP1 and histone H2AX, resulting in the further recruitment of effectors and formation of protein foci at the breaks. (D) DSBs can be repaired through the classical NHEJ pathway (C-NHEJ), which results in the presence of short microhomology (red rectangle) or blunt ends at the S junctions, or through the alternative NHEJ pathway (A-NHEJ), whose signature is the presence of junctions displaying longer microhomologies.

#### 4.4. Processing of double stranded DNA breaks: DNA damage response and repair

##### 4.4.1. "Sensing" the lesions: the DNA damage response (DDR)

After formation of DSBs at the donor and acceptor S regions, the recombination occurs through a signaling cascade that starts with the sensing of the lesion. This step is crucial to activate protein kinases and signal transduction cascade which overall is defined as DNA damage response (DDR, Figure 12, (Harper and Elledge, 2007). The sensor complex for DNA DSBs is the MRN complex, composed by Mre11, Rad50 and Nbs1 (Lee and Paull, 2005). Its binding at the DNA breaks allows the recruitment of ataxia telangiectasia mutated (ATM) (Shiloh, 2003), a serine/threonine kinase which,

after activation, is able to undergo autophosphorylation and to phosphorylate substrates such as Nijmegen breakage syndrome protein 1 (Nbs1) (Falck et al., 2005), p53 binding protein 1 (53BP1) (Anderson et al., 2001; Rappold et al., 2001), mediator of DNA damage checkpoint protein 1 (MDC1) (Lou et al., 2006), KAP1 (White et al., 2006; Ziv et al., 2006) and the histone variant H2AX on serine 139 (Burma et al., 2001; Paull et al., 2000; Rogakou et al., 1998). ATM-mediated phosphorylation of H2AX (so named  $\gamma$ H2AX) leads to the recruitment of 53BP1, MDC1 and Nbs1, which accumulate at the site of damage by generating protein “foci” (Kobayashi et al., 2002; Stewart et al., 2003; Stucki et al., 2005; Ward et al., 2003). MDC1 phosphorylation induces an additional “feedback” by recruiting the E3 ubiquitin ligase RNF8 (Huen et al., 2007; Kolas et al., 2007; Mailand et al., 2007), responsible of the ubiquitinylation of the H2A-type histones, which in turns recruit another E3 ligase, RNF168 (Doil et al., 2009; Stewart et al., 2009), involved in 53BP1 stabilization at the DSBs. By comparing V(D)J recombination to CSR, it appears that main components of the DDR are not crucial for the former, whereas knockout mice studies of Nbs1, ATM, Mre11, H2AX, 53BP1, MDC1, RNF8, RNF168 and PARP1 showed their clear involvement in the Ig S regions recombination (Dinkelmann et al., 2009; Franco et al., 2006; Kracker et al., 2005; Li et al., 2010; Manis et al., 2004; Ramachandran et al., 2010; Reina-San-Martin et al., 2004; Reina-San-Martin et al., 2005; Robert et al., 2009; Santos et al., 2010).

#### 4.4.2. The DSBs repair through non-homologous end joining pathway

The DNA DSB repair in mammals can occur through two alternative pathways: HR and NHEJ. Whereas the former is the preferential solution during the S/G<sub>2</sub> phase of the cell cycle, when the homologous sequence is available, the latter is active through out the cell cycle and is the main pathway involved in resolving DSBs generated during CSR, considering that S regions display no or short homologies and that AID-mediated deamination takes place during the G<sub>1</sub> phase (Petersen et al., 2001; Schrader et al., 2007). The NHEJ pathway displays seven components: Ku70, Ku80 (also known as Ku86) and DNA-PKcs (which form the DNA-PK holoenzyme), the ligase complex XRCC4/DNA Ligase 4, XRCC4-like factor (XLF)/Cernunnos and Artemis (Kotnis et al., 2009). The first molecular players recruited to the breaks are Ku70 and Ku80, which in turn recruit the DNA-PKcs and originate the holocomplex (Lieber, 2010); then XRCC4 and the DNA ligase IV are recruited to the complex and catalyze the ligation of the DNA ends (Figure 12) (Lieber, 2008; Lieber et al., 2003).

The role of NHEJ pathway in resolving DSBs at S regions has been investigated by loss of function studies or patient analysis. Ku70 and Ku80 deficiency leads to a severe CSR defect (Casellas et al., 1998; Manis et al., 1998a; Reina-San-Martin et al., 2003), whereas the involvement of DNA-PKcs is still under debate (Bosma et al., 2002; Cook et al., 2003; Manis et al., 2002; Rooney et al., 2005). XRCC4 and DNA Ligase IV are absolutely required for V(D)J recombination (Frank et al., 1998; Gao et al., 1998; Li et al., 1995; Taccioli et al., 1998) but not for CSR. XRCC4 depletion in mice is embryonic lethal, and transgenic or IgH/IgL knock-in mice display reduced but not abolished switching (Soulas-Sprauel et al., 2007; Yan et al., 2007). A similar profile is observed for DNA ligase IV:

peripheral blood cells isolated from patients harboring a DNA ligase IV deficiency as well as a hypomorphic mutation in the Ligase IV gene (Y288C) display reduced CSR (Nijnik et al., 2009; Pan-Hammarstrom et al., 2005). XLF/Cernunnos, on the other hand, is essential for DSBs repair through the NHEJ pathway but its loss leads to a partial reduction in switching (Li et al., 2008; Zha et al., 2007); however, further studies have demonstrated that the role played by XLF in NHEJ is partially overlapping with the one exerted by ATM, as combined deficiency of these two factors impairs switching and NHEJ-mediated repair (Zha et al., 2011). Strikingly, analysis of switch junctions derived from patients harboring mutations in the gene encoding Cernunnos displayed an altered repair pathway, supporting the function of this factor in NHEJ (Du et al., 2012). Artemis deficiency in B cells isolated from knock-in mice for the IgH and IgL chains results in relatively normal rates of switch recombination (Rooney et al., 2005); however, patient analysis and conditional deletion of Artemis in mature B cells showed an impaired switching to certain isotypes, and a bias towards the use of long microhomologies for the repair of S junctions (Du et al., 2008; Rivera-Munoz et al., 2009). Moreover, the newly described Aprataxin and PNK-like factor (APLF) (Macrae et al., 2008) has been involved in retaining XRCC4/DNA ligase IV to the site of the breaks, in association with PARP3, but APLF deficiency does not impair CSR although it affects the pathway involved in the repair of the DNA breaks (Rulten et al., 2011).

#### 4.4.3. The alternative non-homologous end joining pathway

The observations that mutations in key components of NHEJ pathway still allowed substantial CSR lead to the hypothesis that an alternative pathway could be involved in the resolution of AID-mediated DSBs (Boboila et al., 2010; Pan-Hammarstrom et al., 2005; Soulas-Sprauel et al., 2007; Yan et al., 2007). The alternative NHEJ pathway (A-NHEJ) can be distinguished from the classical one (C-NHEJ) by the analysis of S-S junctions: C-NHEJ-mediated repair leads to very short microhomologies (<4 nt) or blunt ends, whereas the A-NHEJ favors longer microhomologies and loss of blunt ends (Stavnezer et al., 2010). So far, a few proteins have been implicated in this alternative pathway: PARP1, CtIP, Mre11, XRCC1 and DNA ligase III (Figure 12) (Boboila et al., 2012a). PARP1 deficiency leads to a normal switching frequency but to an altered S-S junctions repair, biased towards the usage of short microhomologies (Robert et al., 2009); on the other hand, the involvement of Mre11 could be linked to end processing, as depletion of Mre11 in XRCC4-deficient cells affects end resection (Xie et al., 2009). A similar hypothesis has been proposed for CtIP, whose knockdown in CH12 B cells results in increased microhomologies at S-S junctions (Lee-Theilen et al., 2011). However, this is in contrast with what has been shown for primary B cells depleted of CtIP, whose microhomology pattern at  $\Sigma\mu$ - $\Sigma\gamma$ 1 junctions is similar to CtIP-proficient cells (Bothmer et al., 2013), suggesting that CtIP might not be essential for A-NHEJ. The DNA ligase III, as well as its cofactor XRCC1, has been considered as part of the A-NHEJ based on biochemical assays and plasmid joining assays (Audebert et al., 2004; Wang et al., 2005; Wang et al., 2006b). However, the XRCC1 contribution to this alternative repair pathway seems controversial: whereas heterozygous mice display increased blunt ends and reduced overlaps

at S junctions (Saribasak et al., 2011), conditional inactivation in WT and XRCC4-deficient B cells does not impact the A-NHEJ-mediated repair of S regions in cells undergoing CSR to IgG1 or IgE (Boboila et al., 2012b). Moreover, no defect in CSR is detected after XRCC1 depletion in DNA ligase IV deficient CH12 B cells (Han et al., 2012), suggesting that it is not a critical factor for this repair pathway. The role of DNA ligase III, instead, has been further investigated by depletion in primary B cells or CH12 B cells lacking DNA ligase IV and, surprisingly, no effect on CSR efficiency as well as on switch regions microhomology length or IgH/c-myc translocation frequency was observed, suggesting that is not a crucial factor for this repair pathway or that, as alternative, residual protein level of DNA ligase III can efficiently mediate DNA repair through A-NHEJ pathway (Boboila et al., 2012b).

Alternative NHEJ has been proposed to proceed through end resection, 5'→3' DNA excision in order to generate ssDNA; this step usually occurs during HR, when the homologous sequence can be used as template for the repair. However, if resection occurs at DSBs during CSR, this may lead to intra-switch region recombination which would no longer allow productive joining between donor and acceptor switch region. Recent studies have tried to assess whether the C-NHEJ and A-NHEJ factors might be implicated in mechanism, and both ATM and 53BP1 have been involved in the choice between the classical repair pathway (which allows productive CSR) and the alternative one, biased towards microhomologies and intra-S regions mutations (Bothmer et al., 2010). The ability of 53BP1 to prevent end resection appears as dependent on the distance between the paired DSBs (Bothmer et al., 2011), and this is in line with the putative role of 53BP1 in the synapsis between the two S regions involved in the recombination (Manis et al., 2004; Reina-San-Martin et al., 2007). Additionally, Rif1 interaction with phosphorylated 53BP1 contributes to protect DNA ends from resection and to favor C-NHEJ-mediated repair and productive CSR (Di Virgilio et al., 2013). On the other hand, CtIP and Exo1 have been implicated in positively regulating end resection in primary B cells (Bothmer et al., 2013) as well as RPA, which is bound to DSBs during the G<sub>1</sub> or S-G<sub>2</sub>/M phase of the cell cycle, and is able to promote end resection within the A-NHEJ or HR pathway respectively (Yamane et al., 2013).

## 5. Role of AID outside the immune system

Although the role of AID has been extensively described within the immune system, where the protein exerts its main function, AID-mediated deamination has been described to contribute to other processes, such as genes demethylation and epigenetic reprogramming. AID was found to mediate 5-methylcytosine (5-meC) deamination in mammalian germ cells, introducing a dT:dG mismatch which can either be repaired, with subsequent DNA demethylation, or can lead to C→T transitions on methylated DNA (Morgan et al., 2004). A “demethylation through deamination” mechanism was also described in zebrafish embryos, where the dG:dT mismatch dependent on the 5-meC deamination by AID is processed by the thymine glycosylase methyl-domain binding protein 4 (MBD4), as the interaction between AID and MBD4 is possibly mediated by growth arrest and DNA damage-inducible

45 (GADD45) (Rai et al., 2008). Furthermore, the genome-wide map of cytosine methylation status showed that AID deficiency leads to a hypermethylated status in mouse primordial germ cells (PGCs); as most of the DNA methylation is erased during normal development of PGCs to limit inheritance of epigenetic marks, this suggests that AID could modulate not only the antibody repertoire of mature B cells but also the inheritance of mutations in the early stages of the development (Popp et al., 2010). The DNA methylation status is predictive of gene expression and differentiation state of the cells, and whereas demethylation is required for reprogramming of differentiated cells to a pluripotent stage, AID could represent an intriguing factor in light of future therapies and approaches based on regenerative medicine. Bhutani et al. focused on demethylation by using heterokaryons as system: they fused human fibroblasts with mouse embryonic stem (ES) cells to induce reprogramming and they observed, after AID knockdown, impaired DNA demethylation which resulted in a lower expression of pluripotent markers such as OCT4 and NANOG, consistent with increased methylation of the genes' promoters upon AID downregulation (Bhutani et al., 2010); additionally, in a recent work they showed that AID is required for the generation of induced primordial stem cells (iPSC) by reprogramming mouse embryonic fibroblasts (MEFs) (Bhutani et al., 2013). The study performed by Kumar et al., instead, showed a different profile compared to the one just described. Strikingly, AID does not seem to be required for the induction of pluripotency, as murine fibroblasts lacking AID expression are initially hyper-responsive to the reprogramming stimuli, but rather for establishing/maintaining this feature, thus suggesting that AID might be important in maintaining the stem cell state (Kumar et al., 2013). Moreover, AID-mediated demethylation is still under debate as Fritz et al. showed that there is no significant change in methylation in B cells proficient or deficient for AID (Fritz et al., 2013), thus suggesting that further studies will be required to clarify whether AID exerts "extra functions" within the cell.

Although AID plays an important role in antibody diversification and besides its emerging role in demethylation, it is a dangerous protein as AID-mediated lesions can lead to pathogenesis.

## 6. AID and pathogenesis: B cell lymphomas

In the Western world, most of the lymphomas diagnosed are of B cell origin and derived from GC and post-GC cells (Kuppers, 2005), and these observations are not surprising considering that B cells undergo controlled DNA damage mediated by AID during CSR and SHM in the GC. These two processes have been implicated in the development of tumorigenesis, due to the observations that reciprocal chromosomal translocations, a hallmark of mature B cell lymphomas, present breakpoint at the V and S regions of the Ig loci (Kuppers and Dalla-Favera, 2001). Thus, while experiments involving plasmacytoma cells generated by injection of mineral oil or overexpressing IL-6 ruled out the cause/effect relationship between SHM and CSR and the malignant transformations (Potter and Wiener, 1992; Suematsu et al., 1992), AID-mediated deamination was described as the reaction initiating this uncontrolled genome instability (Ramiro et al., 2006; Ramiro et al., 2004; Unniraman et

al., 2004). Chromosomal translocations can occur during CSR, like those detected between *c-myc* and the IgH locus in Burkitt's lymphoma (BL): in this case the proto-oncogene is relocated in proximity of Ig regulatory regions resulting in its constitutive expression (Casellas et al., 2009).

Furthermore, aberrations involving *Bcl-2* and the IgH locus have been identified in diffused large B cell lymphoma (DLBCL) or follicular lymphoma (FL) cases (Kuppers, 2005). In this latter case, *Bcl-2*/IgH translocations appear to be the consequence of mistakes occurring during V(D)J recombination during the early B cells development, and account for 85% of follicular lymphoma cases (Marculescu et al., 2006; Roulland et al., 2011). Strikingly, healthy individuals carry "FL-like" B cells which harbor *Bcl-2*/IgH translocations and undergo clonal expansion within the GC; moreover, these cells undergo CSR on both alleles despite they display an IgM memory B cells phenotype, and multiple rounds of GC reaction, and thus AID-mediated mutagenesis, might account for the progression to a pathological state and FL development (Roulland et al., 2011; Roulland et al., 2006).

Although DSBs are not intermediate of SHM reaction, they might result from AID activity at the V regions, and contribute to genome instability. This has been detected in some cases of BL where *c-myc* was translocated to the IgL locus (Kuppers, 2005). Recent studies which applied high-throughput sequencing technologies to the analysis of AID-mediated translocations all along the genome evidenced as most of the DSB intermediates of translocations occur at highly transcribed genes (Chiarle et al., 2011; Klein et al., 2011). Additionally, the proximity of transcribed genes to the IgH locus – possibly being located into the same transcription factory – increases the chances of translocations and makes these genes AID "hotspots" (Rocha et al., 2012).

In addition to the deleterious effect of additional DSBs in the genome, B cell lymphomas can also derive from mutations of non-Ig genes, such as *Bcl-6*, *c-myc*, *Pim1*, *RhoH/TFF1* and *Pax5* (Gordon et al., 2003; Migliazza et al., 1995; Pasqualucci et al., 2001; Shen et al., 1998). AID is partially responsible of this scenario, as it is able to mutate 25% of transcribed genes in germinal center cells in absence of UNG and MSH2 (Liu et al., 2008). The majority of the B cell lymphomas, such as FL, DLBCL and BL, express AID (Okazaki et al., 2007) and AID overexpression in transgenic mice induces hypermutation of non-Ig genes (Robbiani et al., 2009). Moreover, when AID constitutive expression is driven by the CAG promoter, it leads to the development of T cell lymphomas, lung microadenomas and adenocarcinomas (Okazaki et al., 2003). Furthermore, AID has been associated to gastric cancer and to oral squamous cell carcinoma (Nakanishi et al., 2013; Takeda et al., 2012). This evidence highlights the "dark side" of cytosine deamination resulting from AID activity and, whereas genomic aberrations represent the first injury which can lead to cell transformation and resulting cancer, is it clear that B cells, which express AID during the germinal center reaction, tightly regulate its expression, localization and activity.

## 7. AID regulation: how B cells limit AID-mediated DNA damage

According to the deamination model and to the mechanisms regulating CSR and SHM, is not surprising that a mutagen and (potentially) dangerous protein as AID has to be tightly regulated at different levels: transcription, post-translational modification, localization and through a plethora of interactors differentially required for Ig diversification mechanisms.

### 7.1. Transcriptional regulation

The *Aicda* locus presents four regulatory regions that are bound by positive and negative regulators. Region 1 is located upstream of the transcription start site (TSS) and contains putative transcription factor-binding motifs for signal transducer and transcription activator 6 (STAT6), NF- $\kappa$ B, Sp elements and HoxC4 (Dedeoglu et al., 2004; Gonda et al., 2003; Park et al., 2009; Tran et al., 2010; Yadav et al., 2006); region 2 is located within the first intron and contains sites for Pax5, E proteins and NF- $\kappa$ B (Gonda et al., 2003; Tran et al., 2010); region 3 is located downstream of exon 5 and has been considered as a putative enhancer (Crouch et al., 2007); although, it has been recently reported the binding of basic leucine zipper transcriptional factor ATF-like (BATF) which regulates AID expression and S regions transcription (Ise et al., 2011). The region 4 is located 8 Kb upstream of the TSS and is recognized by STAT6, NF- $\kappa$ B, Smad3/4 and CCAAT/enhancer-binding protein (C/EBP) (Tran et al., 2010; Yadav et al., 2006). Transcription factors are regulated by the stimuli received by B cells and which can derive from helper T cells (as for CD40L and NF- $\kappa$ B), BCR crosslinking and cytokine production, as occurs for STAT6 and Smad3/4 which are induced by IL-4 and TGF $\beta$ , respectively (Xu et al., 2012). Negative regulation of AID expression is exerted by MYB, E2F and inhibitor of differentiation protein 2 (Id2), which have been proposed to interact with the region 2 (Tran et al., 2010; Xu et al., 2012); additionally, Id1, 2 and 3 proteins inhibit AID by binding to E47 and Pax5 (Goldfarb et al., 1996; Gonda et al., 2003; Quong et al., 1999; Sayegh et al., 2003). Additional modulation of AID expression has been attributed to estrogen and progesterone: the former is enhancing its expression by upregulating HoxC4 (Mai et al., 2010; Pauklin et al., 2009) whereas the latter displays an inhibitory activity (Pauklin and Petersen-Mahrt, 2009).

AID transcripts' stability is deregulated by microRNAs (miRs) such as miR-181b and miR-155 by binding to the 3'UTR of AID (de Yebenes et al., 2013; de Yebenes et al., 2008; Dorsett et al., 2008; Teng et al., 2008). Whereas miR-181b is expressed in resting B cells and has been proposed to impair an inappropriate expression in cells which are not stimulated (de Yebenes et al., 2008), miR-155 is expressed in activated B cells and might limit excessive accumulation of AID transcripts and potential off-target effects, as suggested by increased number of IgH/c-myc translocations detected in mice carrying a mutation in the AID 3'UTR miR-155 binding site (Dorsett et al., 2008; Teng et al., 2008). Moreover, a recent study proposed miR-93 as additional negative regulator of AID. By using the MCF-7 breast carcinoma line, which expresses AID, Borchert et al. propose that miR-93 and miR-155 inhibit AID translation, thus limiting its oncogenic potential (Borchert et al., 2011).

## 7.2. Post-translational modifications

Post-translational modifications represent an additional level to control protein functions and, concerning AID, phosphorylation has been described as pivotal for its function in antibody diversification. AID displays several phosphorylation sites: serine 3 (S3), threonine 27 (T27), serine 38 (S38), serine 41 (S41), serine 43 (S43), threonine 140 (T140) and tyrosine 184 (Y184) (Basu et al., 2005; Gazumyan et al., 2011; McBride et al., 2006; McBride et al., 2008; Pasqualucci et al., 2006; Pham et al., 2008). Phosphorylation of T27 and S38 is mediated by protein kinase A (PKA) and regulates AID interaction with replication protein A (RPA) (Basu et al., 2005; Pasqualucci et al., 2006): as RPA has been described to enhance AID binding to transcribed DNA (Chaudhuri et al., 2004), is not surprising that mutations in these two residues significantly impair CSR, SHM and IGC, although T27 has been suggested as minor site (Basu et al., 2005; Chatterji et al., 2007; McBride et al., 2006; Pasqualucci et al., 2006; Vuong et al., 2009). Additionally, S38 can be phosphorylated by the PKC *in vitro*, as well as T140, and mutation in the latter residue (T140A) has been described to impair specifically SHM, leading to the hypothesis that T140 phosphorylation might be important for AID and SHM-specific factors (McBride et al., 2008). The Y184 does not seem to be a critical residue, as mutation to alanine does not impair CSR (Basu et al., 2005); on the other hand, S41 and S43 have been identified in *Sf9* insect cells expressing human GST-AID but no role has been described so far (Pham et al., 2008), implying that further studies will be required to address whether these sites are important in modulating AID activity. In spite of the positive regulation of AID mediated by phosphorylation, the recently described S3 site seems to be important in limiting AID activity: mutation of S3 to alanine leads to enhanced CSR and a significant increase in IgH/c-myc translocations (Gazumyan et al., 2011). Furthermore, the serine/threonine phosphatase PP2A has been involved in regulating phosphorylation rate at this site, suggesting that the balance between the PKC-mediated phosphorylation and PP2A-mediated dephosphorylation of AID at S3 might represent a fast solution to “turn off” AID in germinal centers (Gazumyan et al., 2011).

## 7.3. Subcellular localization

AID function as deaminase is exerted in the nucleus but its localization is predominantly cytoplasmic (Rada et al., 2002a; Schrader et al., 2005), and this restriction – as well as the factors-mediated regulation – can be considered as a strategy to limit AID's off targeting. Mutational analyses have revealed that AID C-terminus contains an anchor sequence for cytoplasmic retention and that AID is actively imported into the nucleus through importin- $\alpha$ 3 (Patenaude et al., 2009); here its retention can be limited by either ubiquitylation and proteasome-mediated degradation (Aoufouchi et al., 2008) or export through CRM-1, which interacts with the NES present at the C-terminus of the protein (Brar et al., 2004; Geisberger et al., 2009; Ito et al., 2004; McBride et al., 2004). Thus, AID shuttles between



the nucleus and the cytoplasm and the retention in one or the other compartment can be modulated by numerous factors.

#### **7.4. AID and its cofactors**

AID is a relatively small protein but when looking at the numerous screening performed in order to identify its interactome it looks like is part of a complex with an excessively huge number of partners. In the last ten years numerous regulators of AID – direct or indirect – in the process of CSR, SHM or IGC have been identified, and what is surprising and exciting in the meantime is that many of these factors play additional roles when compared to their main function within the cell. What emerges by these studies is that AID is tightly regulated in each step of its “life” (Figure 13): it is stabilized in the cytoplasm, translocated into the nucleus, targeted to ssDNA, phosphorylated and retained to the targeted DNA the time required to deaminate cytosines and allow the lesion processing, which will lead in turn to mutations at the Ig V regions or generation of DSBs at the S regions.

It is well known that AID is active into the nucleus but mostly retained into the cytoplasm, consistent with its mutagenic and thus potentially dangerous activity. The laboratory of Michael Neuberger recently identified elongation factor 1 alpha (eEF1A), a factor involved in protein synthesis, as responsible of the cytoplasmic retention of AID (Hasler et al., 2011). eEF1A interacts with the residue in position 187 of AID, as the AID mutant D187A displays increased nuclear localization and, in association to a rapid degradation, a better ability to rescue CSR in AID-deficient splenocytes (Hasler et al., 2011). This is consistent with previous observations which described the residues located at AID C-terminus (D187 and D188) as critical for the deaminase’s cytoplasmic retention (Patenaude et al., 2009). Moreover, AID is stabilized in the cytoplasm by histone chaperones, Hsp90 and Hsp40 Dnaja1, as described by the work of Orthwein et al. (Orthwein et al., 2010; Orthwein et al., 2012). It has been proposed that Hsp40 Dnaja1 stabilizes cytoplasmic AID whereas Hsp90 prevents proteasomal AID degradation (Orthwein and Di Noia, 2012). The role of AID in the “chaperone network”, as well as the possibility that Hsp90 could be part of the same complex of AID and eEF1A, needs to be further investigated, but these evidence show that AID is “preserved” and made available in the cytoplasm.

Upon translocation within the nucleus, AID is stabilized by the transcriptional repressor protein YY1: knockout of YY1 impairs switching to all antibody isotypes except IgE, and its interaction with AID has been assessed *in vitro* and *in vivo* (Zaprazna and Atchison, 2012). On the other hand, REG $\gamma$  activity counteracts the one of YY1. REG $\gamma$  mediates the proteasome-dependent AID nuclear destabilization, as REG $\gamma$ -depleted splenocytes display a higher abundance of AID (Uchimura et al., 2011), in line with the shorter half-life of AID within the nucleus compared to the cytoplasm (Aoufouchi et al., 2008).

However, AID presence within the nucleus is not sufficient to induce antibody diversification, as the concerted availability of DNA template and AID recruitment have to be tightly coordinated. As I already mentioned, transcription is required for CSR and, in line with this, there are many transcription-related factors described to regulate AID activity. AID was identified in the same complex with the RNA polymerase II as far as ten years ago (Nambu et al., 2003), and transcription makes available ssDNA

template for AID-mediated deamination through the formation of R-loops. The negative regulation of Topoisomerase I (Top I) has been proposed to favor the R-loop formation, as lower levels of the enzyme would allow a negative supercoiling at the rear of the transcription machinery and thus the formation of non-B DNA structures (Kobayashi et al., 2009). Interestingly, AID has been shown to inhibit Top I translation, and the topoisomerase threshold expression has been described to be important for CSR and SHM (Kobayashi et al., 2009; Kobayashi et al., 2011). In this scenario of ongoing transcription and ssDNA exposure to AID can be included the RNA exosome, implicated in RNA processing/degradation, which targets AID to template and non-template ssDNA (Basu et al., 2011). The RNA polymerase II processivity has also been subject of study, as stalled polymerase seems to be pivotal in recruiting factors which build a sort of "AID binding platform": it is the case for Spt5 (Pavri et al., 2010), as it will be discussed later. Germinal center-associated nuclear protein (GANP), on the other hand, is a factor involved in RNA export from the nucleus and has been proposed to translocate AID into the nucleus and to regulate its binding to the V regions (Maeda et al., 2010).

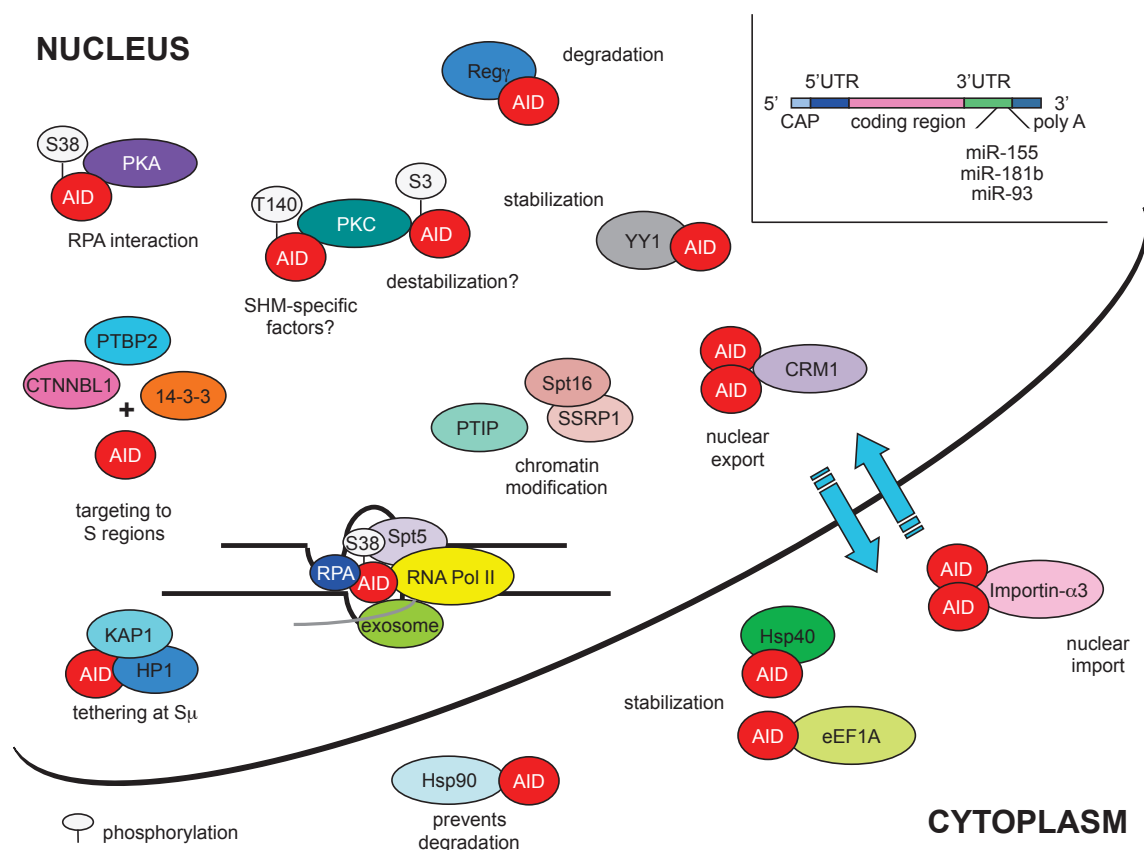
The adaptor 14-3-3, as well as polypyrimidine tract binding protein 2 (PTBP2) and CTNNBL1 are also implicated in AID targeting/tethering to the S regions (Conticello et al., 2008; Nowak et al., 2011; Xu et al., 2010). The 14-3-3 proteins are upregulated in stimulated splenocytes and recruit AID to the 5'-AGCT-3' sequences at the S regions through interaction with AID C-terminal domain (Xu et al., 2010); the splicing regulator PTBP2, on the other hand, interacts with AID in mouse B cell and has been proposed to mediate AID targeting to the S regions through its interaction with the RNA generated during transcription (Nowak et al., 2011). The role played by CTNNBL1, instead, is still under debate. By using the yeast two hybrid screening for human splenocyte cDNA library in order to identify AID cofactors, CTNNBL1 was described to interact with the N-terminal domain of the deaminase (Conticello et al., 2008). Its depletion in DT40 cell lines leads to an impairment in immunoglobulin gene conversion, while the reconstitution of AID-deficient splenocytes with a mutant AID unable to bind CTNNBL1 displays reduced class switching (Conticello et al., 2008). Additionally, ChIP experiments performed on DT40 cells overexpressing tagged CTNNBL1 showed that it binds to chromatin at the Ig $\lambda$  loci, suggesting a role in AID targeting (Conticello et al., 2008). However, a recent report from Han et al. describes CTNNBL1 as not essential for CSR, upon knockout in the mouse CH12 B line (Han et al., 2010). This latter report questioned the involvement of CTNNBL1 in CSR, and suggested a specific role in SHM and IGC, but it did not address fully the different steps of CSR to rule out any potential secondary effect due to CTNNBL1 depletion. Thus, it appears that more investigation is needed to rule out the role played by CTNNBL1 in AID regulation.

In addition to the ongoing transcription and to the physical interactions, which can bring AID to the S regions, histone modifications at the Ig loci play as well a pivotal role in regulating antibody diversification. In this scenario can be included Pax interaction with transcription-activation domain protein 1 (PTIP) and the FACT complex (Spt16/SSRP1), both modulators of the chromatin status (Daniel et al., 2010; Stanlie et al., 2010). ChIP-Seq experiments revealed that PTIP is able to regulate the association of the RNA polymerase II to the acceptor switch region in activated splenocytes; moreover, PTIP regulates most of the histone modifications detected at the activated S regions (such

as H3K4me3, H3K27ac, H3K8ac and H3K36me3) while it seems to be dispensable for modifications occurring at S $\mu$ , S $\delta$  and 3'RR (Daniel et al., 2010). The FACT complex, instead, is responsible of maintaining H3K4me3 at the S regions, as SSRP1 knockdown impairs histone post-translational modifications at the donor and acceptor S regions (Stanlie et al., 2010). Our laboratory addressed the role of KAP1 in CSR regulation: KAP1, in association with HP1, mediates AID tethering at the donor S $\mu$  region, through interaction with H3K9me3 mark *in vivo* (Jeevan-Raj et al., 2011).

The recruitment of AID to the target regions has as consequence its activation mediated by PKA (Basu et al., 2005; Pasqualucci et al., 2006; Vuong et al., 2009), which phosphorylates S38 enhancing the binding with RPA, that in turn stabilizes AID on ssDNA (Chaudhuri et al., 2004). The ability of AID S38A, harboring a mutation which impairs PKA-mediated phosphorylation, to bind to the S regions (Vuong et al., 2009) suggests that the interaction with the PKA may occur after AID recruitment to the Ig loci, although further investigations will better define the order of these important steps which allow AID to deaminate cytosines at the Ig loci. On the other hand, ChIP-Seq experiments revealed that RPA is recruited specifically to the IgH loci, dependent on AID expression (Yamane et al., 2010).

Moreover, DNA repair factors such as DNA-PKcs, UNG and MSH2/MSH6 have also been described as AID interactors (Ranjit et al., 2011; Wu et al., 2005). In this case has been proposed that AID, instead, is involved in their recruitment to the Ig loci to favor the formation of DSBs and an efficient repair; interestingly, all these factors are able to interact with the C-terminus of AID: more specifically, DNA-PKcs requires the cytidine deminase motif and the NES (Wu et al., 2005), whereas for UNG and MSH2/6 the presence of the last 10 residues of AID (AA 189-198) is critical (Ranjit et al., 2011). MDM2, instead, is a negative regulator of DNA repair and, upon identification as partner of AID through yeast two hybrid screening, its controversial role in IGC has led to the hypothesis that it might play a role as negative AID regulator as well, which competes with other factors for AID binding (MacDuff et al., 2006). The E3 ligase RNF126 was identified upon AID co-expression with a Ramos B cell cDNA library and is shown to mono-ubiquitinylate AID *in vitro* and in HEK293T cells; although the role of RNF126 should be addressed in B cells, this work shows as the alternative strategy of protein co-expression - which allows to increase the solubility of the "insoluble partner" of the complex - can be considered as good alternative to the previously described systems (Delker et al., 2013).



**Figure 13. Cellular regulation of AID**

*AID mRNA stability is regulated by miR-155, miR-181b and miR-93 (inset). AID cellular localization is modulated by importin- $\alpha$ 3 and CRM1, while its stability within the cytoplasm depends on Hsp40 Dnaja1, eEF1A Hsp90 which prevents proteasomal degradation. Within the nucleus, AID is stabilized by YY1 or destabilized by Reg $\gamma$ . AID activity is modulated also through post-translational modifications and through factors involved in chromatin modification, targeting/tethering to the DNA or transcription.*

The regulation of AID in activated B cells is still a matter of debate, and further studies will be required to elucidate the exact mechanism of AID recruitment and retention at the targeted DNA. However, a huge contribution to what is known so far about AID and antibody diversification in general derives from the study of patients affected by class switch recombination-immunodeficiencies or, as also known, hyper IgM syndromes. Moreover, for our study we disposed of cell lines derived from patients harboring a CSR defect whose cause has not been identified yet, as it will be presented in the next chapter.

### III. Class switch recombination-immunodeficiencies

The class switch recombination-immunodeficiencies (CSR-ID), also known as hyper-IgM syndromes (HIGM), are rare primary immunodeficiencies whose frequency is about 1 in 500000 births in the Caucasian population. They are all due to a defect in CSR, resulting in normal or higher IgM serum levels associated to a lower or absent IgG, IgA or IgD. Mature B cells display expression of IgM and IgD or IgM alone, and the defect in CSR can be associated or not to an impairment in SHM (Durandy et al., 2007). The characterization of patients affected by CSR-ID allowed the delineation of the exact role of key molecular players in antibody diversification mechanisms as well as the repair and signaling pathways which contribute to a more efficient and specific immune response. The impaired switching can be the consequence of an impaired cell-cell interaction and intracellular pathway activation, as occurs for the CD40/CD40 ligand signaling components, or can depend on a B cell intrinsic defect, as listed in Table 1.

	GENE	TRANSMISSION	CSR IMPAIRMENT	SHM IMPAIRMENT
<b>CD40 SIGNALING DEFECT</b>				
	CD40L	X-L	yes	yes
	CD40	AR	yes	yes
	NEMO	X-L	yes	yes/no
<b>B CELL INTRINSIC DEFECT</b>				
	AID	AR	upstream of DSBs	yes
	AID N-terminus	AR	?	yes
	AID C-terminus	AR	?	no
	AID C-terminus (NES)	AD	downstream of DSBs	no
	UNG	AR	upstream of DSBs	no (bias)
	PMS2	AR	upstream of DSBs	no
	AID cofactor?	?	upstream of DSBs	no
	DNA repair?	?	downstream of DSBs	no

**Table 1. Class switch recombination immunodeficiencies (CSR-ID)**

List of CSR-ID due to a defect in the CD40/CD40L signaling or to an intrinsic B cell defect. X-L: X-linked; AR: autosomal recessive; AD: autosomal dominant. Adapted from Kracker et al., 2010a.

#### 1. CSR-ID due to a CD40 signaling defect

The interaction between B and T cells in secondary lymphoid organs is an essential step to activate B cells and to initiate the process of CSR and SHM, and it occurs through the recognition of the CD40 ligand (CD40L, CD154), expressed on the surface of follicular helper T ( $T_H$ ) cells (Breitfeld et al., 2000), by CD40 which is expressed on B cells and monocytes. Thus, the loss of any of the CD40 signaling components does not enable B cells to further diversify their antibody repertoire.

### **1.1. CD40 ligand deficiency**

The characterization of patients harboring mutations in the gene coding for CD40L contributed to the dissection of the CD40-mediated B cells activation pathway (Allen et al., 1993; Aruffo et al., 1993; DiSanto et al., 1993; Korthauer et al., 1993; Kroczek et al., 1994). The deficiency of CD40L has an X-linked inheritance (Castle et al., 1993; Fuleihan et al., 1993; Nonoyama et al., 1993) and B cells from patients are unable to proliferate and form germinal centers in secondary lymphoid organs, displaying a CSR defect *in vivo*; although, they are intrinsically normal and able to undergo switching to different isotypes *in vitro* (Durandy et al., 1993). As the defect is upstream of the signaling for differentiation, SHM is also affected (Agematsu et al., 1998). The impaired production of IgG and IgA immunoglobulin leads to susceptibility to recurrent bacterial infections. However, some patients display serum IgA immunoglobulin and some level of SHM, suggesting an alternative pathway for antibody diversification: CSR to IgA could be mediated by a proliferation-inducing ligand (APRIL) in the gut lamina propria (He et al., 2007), whereas T cell-independent SHM could reflect an innate immune defense (Scheeren et al., 2008; Weller et al., 2003). In addition to the loss of activation of B cells, CD40L-deficient patients are also prone to opportunistic infections because T cells are not able to interact with monocytes and dendritic cells, resulting in inefficient dendritic cells maturation and T cells priming (Lougaris et al., 2005; Notarangelo et al., 1992) and, in this case, Ig substitution therapies are inefficient.

### **1.2. Loss of CD40**

Class switch recombination defect due to a CD40 deficiency has been diagnosed in a few patients lacking CD40 expression on B cells and monocytes. The disease has an autosomal recessive inheritance and the clinical and immunological profiles are basically identical to the one caused by a deficiency of CD40L; the only difference is the inability of B cells stimulated with CD40L to undergo CSR, due to the lack of the receptor (Ferrari et al., 2001).

### **1.3. Impaired activation of NF- $\kappa$ B pathway**

The binding of CD40L to CD40 receptor on B cells leads to the activation of NF- $\kappa$ B pathway, whose deficiency leads to ectodermal dysplasia associated with immunodeficiency (EDA-ID) (Doffinger et al., 2001; Jain et al., 2001; Zonana et al., 2000). EDA-ID is caused by a X-linked hypomorphic mutation of NF- $\kappa$ B essential modulator (NEMO), a scaffolding protein in the activation pathway (Hanson et al., 2008). The lack of NEMO expression results in low levels of serum IgG and IgA and impaired antibody responses, with susceptibility to mycobacterial infections. As EDA-ID is heterogeneous, CSR and SHM can be either defective or occur at normal frequencies *in vitro* (Jain et al., 2004; Kracker et al., 2010a).

## 2. CSR-ID due to an intrinsic B cell defect

While CD40 signaling defect extends the pathological consequences to humoral and cell-mediated immune responses, mutations in factors directly involved in CSR and SHM have an impact on humoral immunity and define the intrinsic B cells immunodeficiencies. Altered B cell functionality results in susceptibility to bacterial infections, which can be controlled by intravenous Ig substitutions. CSR does occur neither *in vivo* nor *in vitro*, upon B cells stimulation, while SHM can be defective as well in some cases.

### 2.1. AID deficiency

The characterization of patients harboring AID deficiency represented a breakthrough discovery in the study of antibody diversification mechanisms, and has been pivotal for the deaminase identification. Loss of AID represents the most frequent autosomal recessive form of CSR immunodeficiency (Caratao et al., 2013; Catalan et al., 2003; Minegishi et al., 2000; Revy et al., 2000; Zhu et al., 2003). Mutations in the *AICDA* gene are scattered all along the gene and lead to a complete block in CSR and SHM, although CD27<sup>+</sup> memory B cells are present at normal levels. Mutations are missense, in-frame small insertions or deletions, or a large deletion which lead to a reduced or undetectable protein level in EBV-immortalized B cell lines, and no mutation hotspots have been observed (Durandy et al., 2007). The detection of lymphoid hyperplasia suggests high B cell proliferation (Minegishi et al., 2000; Quartier et al., 2004; Revy et al., 2000; Zhu et al., 2003), and some patients display IgM-mediated autoimmune manifestations, such as autoimmune hemolytic anemia and thrombocytopenia or, more rarely, systemic lupus erythematosus or diabetes (Durandy et al., 2007). By dissecting the switching reaction, it has been observed that germline transcription at the IgH locus occurs at normal levels, whereas the downstream steps are disturbed: in fact, it is not possible to detect DSBs at the S regions, nor to amplify the recombined sequences and the intervening sequences which are excised upon recombination (Catalan et al., 2003; Revy et al., 2000). When assessed in CD19<sup>+</sup> and CD19<sup>+</sup>/CD27<sup>+</sup> memory B cells, SHM is found impaired or drastically reduced compared to age-matched controls (Revy et al., 2000). Although these represent the general features displayed by AID deficiency, further investigations revealed slight differences in the clinical and immunological profiles of some patients harboring *AICDA* mutations located at the C-terminal or N-terminal domain.

#### 2.1.1. AID C-terminal mutations

In some patients, the *in vivo* and *in vitro* defect in CSR is associated to a normal frequency and pattern of SHM; sequence analysis displayed that the AID mutations are located at the C-terminus of its coding sequence (Durandy et al., 2007; Ta et al., 2003). Although most of these mutations are homozygous, consistent with an autosomal recessive inheritance, some of those located in the NES are found to be heterozygous and to exert a dominant negative effect by leading to a truncated form of

the protein (Imai et al., 2005; Kasahara et al., 2003; Ta et al., 2003). Interestingly, in this latter case the switching defect is located downstream of DSBs generated at the S regions, and junction sequencing revealed a bias towards the use of long microhomologies, suggesting the involvement of a DNA repair factor or a deleterious accumulation of the truncated protein within the nucleus (Imai et al., 2005; Kracker et al., 2010b). The intriguing phenotype observed in patients harboring *AICDA* mutation in its C-terminus uncouples CSR and SHM, and suggests that AID could interact with CSR-specific cofactors through its C-terminal domain. These observations were confirmed by the analysis of an AID artificial mutant lacking the last 10 residues and which was able to catalyze SHM and IGC but not CSR (Barreto et al., 2003).

### 2.1.2. AID N-terminal mutations

Although it has been described that mutations in the AID NLS lead to a defect in SHM and normal CSR in mouse (Shinkura et al., 2004), this phenotype has not been confirmed yet by the study of CSR-ID patients, where the two antibody diversification mechanisms are both impaired; this suggests that either the mutated protein is unable to translocate into the nucleus or that the AID cofactors requirement in mouse and humans might be different (Durandy et al., 2007).

## **2.2. Loss of UNG**

The UNG deficiency is a rare cause of CSR-specific defect, as only three patients have been described so far with inactivating mutations in its coding sequence (Imai et al., 2003b). These patients display recurrent bacterial infections of the respiratory tract, lymphadenopathies (2/3 patients) and one developed Sjögren syndrome, an autoimmune disease. The defect in CSR is located upstream of the generation of DSBs at the S regions, whereas SHM is detected as occurring at normal frequencies but is characterized by a mutation bias towards transitions, as the uracil introduced upon AID-mediated cytosine deamination is not removed by UNG, and DNA replication occurs over the mismatch introduced (Imai et al., 2003b). Four mutations have been identified so far: one homozygous and two heterozygous small deletions located in the catalytic domain of *UNG1* and *UNG2* (mitochondrial and nuclear isoforms, respectively), which lead to a premature stop codon, and a homozygous missense mutation (Kracker et al., 2010a). As UNG is part of the BER pathway which removes the uracil introduced into the DNA upon AID-mediated deamination, and thus protects the genome from spontaneous mutations, it represents an anti-mutagenic factor as has been confirmed by the development of B cell lymphomas in mice *UNG*<sup>-/-</sup> (Nilsen et al., 2003). Thus it is possible that inactivating mutations in the *UNG* gene may predispose to such malignancies, although the limited number of cases analyzed does not allow drawing any conclusion.



### **2.3. Deficiency of PMS2**

The MMR pathway, as well as the BER, plays an important role in CSR and SHM. Two main MMR complexes have been identified: MutS homolog (MSH1-6) and MutL homolog (PMS2/MLH1/PMS1). MSH1/2 are involved in dU:dG mismatch recognition in absence of UNG, and allow SHM and CSR through the downstream action of Exo1 and DNA polymerase  $\eta$  (Delbos et al., 2007; Kracker et al., 2010a; Kratz et al., 2008; Peron et al., 2008); PMS2/MLH1, instead, have been proposed to convert single- to double-stranded DNA breaks to favor recombination (Stavnezer and Schrader, 2006). Additionally, six patients displaying a variable CSR-immunodeficiency harbor homozygous nonsense mutations in PMS2 gene, leading to a truncated protein or reduced expression (De Vos et al., 2006). Recurrent bacterial infections are not a common feature of this defect, whereas reduced serum levels of IgG and IgA are detected in all the patients; most of them develop cancer, such as colon carcinoma, and SHM is found normal in frequency and pattern. The CSR defect seems to be located upstream of DSBs at the S regions, as PMS2 might not be able to convert SSBs into DSBs after UNG-mediated uracil excision; additionally, as some patients display few IgA<sup>+</sup> switched B cells, the sequencing of S $\mu$ -S $\alpha$  junctions reveals a bias towards the use of long microhomologies (Peron et al., 2008), defining a role for PMS2 in the generation of DSBs required for efficient CSR.

### **2.4. CSR-ID due to a known DNA repair defect**

As the components of DNA repair pathways such as MMR, BER and NHEJ play a key role in favoring the proper repair at the S regions upon recombination, it is not surprising that a deficiency in one of these molecules might also have an impact on isotype switching. This is the case for the DNA-damage response components ATM, MRE11 and Nbs1, as well as for the NHEJ components, such as Cernunnos, DNA ligase IV or Artemis (de Miranda et al., 2011).

### **2.5. CSR-ID due to an unknown defect and associated to normal SHM**

Half of the B cell-intrinsic CSR-immunodeficiencies identified so far cannot be attributed to a deficiency of AID, UNG or PMS2 and display an autosomal recessive inheritance pattern. They display a clinical profile similar to the one caused by loss of AID: increased susceptibility to bacterial infections, mild lymphoid hyperplasia and possible autoimmune manifestations. CSR defect is milder compared to patients lacking AID expression, as residual IgGs are present in the serum of some patients, but due to an intrinsic B defect as patients' B cells are not able to switch *in vitro* upon stimulation. Expression of AID and UNG as well as germline transcription is normally detected,

suggesting that the IgH locus is poised to be targeted by AID. The defect is limited to CSR, as mutations at the V region occur at normal frequency and pattern in memory B cells (Durandy et al., 2007). A further analysis of patients displaying these features revealed the existence of two sub-groups, depending on the presence or not of DSBs at the S regions.

#### 2.5.1. Defect upstream of DSBs: AID CSR-specific cofactor hypothesis

This subset of patients displays a phenotype similar to the one of patients carrying mutations in the C-terminal domain of AID. SHM is occurring and memory B cells are present at normal frequency; although, undetectable DSBs at S $\mu$  region localize the block in CSR downstream of germline transcription and upstream of AID-mediated DNA lesion, suggesting that AID might not be targeted to the S regions (Durandy et al., 2007) and that the lack of expression of AID cofactor(s) specifically required for CSR could explain these CSR-ID. The cofactor has not been identified yet and the characterization of this defect in CSR is one of the main goal of my thesis project, as it will be developed later.

#### 2.5.2. Defect downstream of DSBs: DNA repair factor hypothesis

This second subset of patients, unlike the one described above, displays detectable DSBs at the S $\mu$  regions, thus locating the CSR defect downstream of deamination-mediated DNA lesion. *In vitro* B cells activation confirms the intrinsic inability of patients' B cells to undergo switching (Durandy et al., 2007; Imai et al., 2003a; Peron et al., 2007). S regions analysis reveals the usage of longer microhomologies in patients' S $\mu$ -S $\alpha$  sequences and a lower number of insertions; moreover, the impairment in DNA repair is confirmed by increased radiosensitivity of patients-derived fibroblasts submitted to increasing doses of  $\gamma$ -irradiation (Peron et al., 2007). Interestingly, the memory B compartment appears decreased in some cases when compared to controls (Imai et al., 2003a), and occurrence of B cell lymphomas is reported (Durandy et al., 2007; Imai et al., 2003a; Peron et al., 2007). Concerning SHM, it seems to occur at normal or slightly reduced frequency: in this latter case, a bias in G:C substitutions in favor of transitions is found (Peron et al., 2007). This intriguing phenotype suggests that a mutation in a DNA repair protein might be responsible of this phenotype, but so far NHEJ components as well as UNG, Rev1, Rev3, Rev7 and the MMR components (Msh2, Msh5, Msh6, Exo1, Mlh1 and Pms2) have been excluded from the list of candidates, in light of normal gene sequence and protein expression (Peron et al., 2007), making this defect still uncharacterized.

The CSR-ID characterization has heavily contributed to dissect the process of switching and the mechanism of action of AID; although, as already mentioned, one of the key steps indispensable for efficient recombination at the IgH locus is transcription of the S regions and the relevant role played by transcription-related factors will be developed in the next chapter.

## IV. Spt5 and RNA polymerase II: the breakthrough

Transcription is one of the most extensively studied processes in eukaryotes, and plays a pivotal role in antibody diversification mechanisms. From the “AID point of view”, taking into account its interaction with the RNA polymerase II (Nambu et al., 2003), one could predict a correlation between transcription and AID recruitment, and this prediction has found a confirm from the work of Yamane et al., who showed by ChIP-Seq experiments that AID localizes mainly at actively transcribed genes, although transcription *per se* is not a hallmark to predict AID-mediated cytosine deamination (Yamane et al., 2010). However, the genome-wide AID occupancy has somehow revolutioned the approach applied to unravel where, when and how AID is targeted (or not) to Ig (or non-Ig) genes, but when I undertook my thesis project the RNA polymerase II dynamics were not as clear as now.

However, recent evidence suggests that switching and the RNA polymerase II dynamics are strongly related. In addition to the “processive” transcription steps – initiation, elongation and termination – it has been described that the RNA polymerase II undergoes a slow down step after transcription of about 100 nt from the promoter, and this promoter-proximal stalling seems to be prevalent in metazoans as required to further regulate transcription output. Pol II stalling can be dependent on the DNA sequence which is transcribed and also by many factors which dictate its dynamics (Nechaev and Adelman, 2011). Although the role of the template DNA in transcription dynamics has not been fully elucidated yet, it appears that influences the processivity of the RNA polymerase. More specifically, the stability of the RNA/DNA hybrid can influence elongation (Nechaev and Adelman, 2011). It has been proposed that the Pol II, after proceeding through low stability regions (such as those AT-rich), might slide backward to a more thermodynamically stable sequence, such as GC-rich (Komissarova and Kashlev, 1997), and this movement can displace the 3' of the newly synthesized RNA from the polymerase active site and thus, by the time the RNA is re-aligned, resulting in a blockage. This model would fit with transcription occurring at the IgH locus during germline transcription, if the RNA polymerase moves back to the S regions, which are known to be GC-rich, and gets stacked over there. However, most likely the repetitive G-rich sequences at the S regions may facilitate DNA distortion and thus favor the formation of R-loops, impairing the forward motion of Pol II (Daniels and Lieber, 1995a; Rajagopal et al., 2009; Ronai et al., 2007; Tian and Alt, 2000; Wang et al., 2009; Yu et al., 2003). The stalled polymerase could then represent a sort of “platform” for AID recruitment and retention, the time required to allow cytosine deamination. Additionally, in 1996 Peters and Storb proposed that SHM may be mediated by a “mutator factor” which is bound to the stalled RNA polymerase II within the target sequence: in their visionary work they speculated that this factor is present only in B cells, and is loaded into the transcription initiation complex, that allows pausing and recruits transcription-coupled DNA repair factors (Peters and Storb, 1996). Most of these hypotheses are now supported by experimental evidence, but this scenario is not complete without the factors that have been recently involved in binding stalled Pol II and in promoting transcription elongation.

## 1. Spt5: the missing link between AID and transcription

The work published by Pavri and Nussenzweig almost three years ago, describing the role of the Suppressor of Ty 5 homolog (Spt5) in antibody diversification, represents a breakthrough in the knowledge of transcription and AID-mediated mutagenesis (Pavri et al., 2010). Spt5 was initially identified in *Saccharomyces cerevisiae* and described as elongation factor (Swanson et al., 1991); further studies showed its interaction with Spt4, by forming the 5,6-dichloro-1- $\beta$ -d-ribofurano-sylbenzimidazole (DRB) sensitivity-inducing factor (DSIF) complex (Wada et al., 1998; Yamaguchi et al., 1999b) which, in association with the negative elongation factor (NELF), is able to induce RNA polymerase II pausing *in vitro* (Wada et al., 1998; Yamaguchi et al., 1999a).

Indeed, Spt5 was shown to bind the Pol II and to induce stalling *in vitro* and *in vivo* (Lis, 2007; Rahl et al., 2010; Yamaguchi et al., 1999a) and, furthermore, the association between Spt5 and the exosome (Andrulis et al., 2002) and the role of the exosome complex in targeting AID to the S regions (Basu et al., 2011) suggested that Spt5 could be involved as well in antibody diversification. Now, its role in class switching by mediating AID association to the transcription machinery has been extensively elucidated (Pavri et al., 2010). Not only Spt5 associates with AID in a DNA and RNA-independent manner, but it mediates the binding of AID to the RNA polymerase II; furthermore, ChIP-Seq experiments showed a correlation between Spt5 genome-wide occupancy and the stalled Pol II as well as a higher mutation frequency of those genes targeted by Spt5. Thus, these results reinforce the idea that the association between AID and the stalled Pol II would favor cytosine deamination by providing, on one hand, ssDNA exposed during transcription to AID and, on the other hand, by retaining AID – through stalling – at the Ig loci the time required to mutate the target sequences (Pavri et al., 2010).

Moreover, the role of the DSIF complex has been further characterized in CSR. Spt4 and Spt5 depletion has an impact on CSR efficiency, on H3K4me3 presence at the donor and acceptor S regions and on DNA repair, although further analysis are needed to clarify the different behavior observed for the two members of the complex (Stanlie et al., 2012). Additionally, the mRNA export factor GANP has also been associated to Spt5: has been described as part of the same complex with the RNA polymerase II and Spt5 in Ramos B cells, and has been proposed to mediate histone modifications at the IgV regions and the recruitment of the DSIF complex, which might mediate the RNA polymerase II stalling and allow the exposure of ssDNA substrate for AID deamination and consequent mutagenesis (Singh et al., 2013).

However, when I started my thesis project nothing was known about these mechanisms and, in the course of my experiments, two additional players appeared to be relevant for antibody diversification: the histone chaperone Spt6 and the PAF complex.

## V. Transcription and chromatin-regulating factors

### 1. Spt6: more than a chaperone

The Suppressor of Ty 6 homolog (Spt6), as Spt5, was identified in *Saccharomyces cerevisiae*, described as essential for viability and involved in transcription initiation (Clark-Adams and Winston, 1987; Winston et al., 1984). However, Spt6 can be considered as a “multi-tasking” factor involved in many processes linked to chromatin modification and transcription regulation.

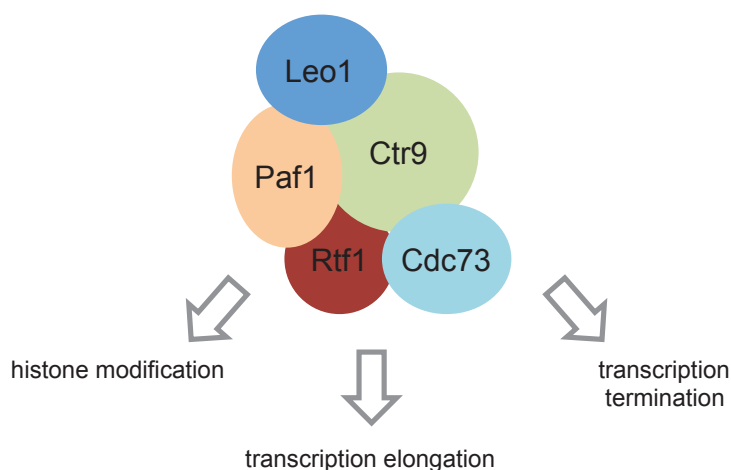
The study of Swanson and collaborators showed, by immunoprecipitation experiments, a physical association between Spt4, Spt5 and Spt6 and the comparison of mutant strains for *spt4*, *spt5* and *spt6* with histone mutants led to the hypothesis that they might be involved in chromatin remodeling (Swanson and Winston, 1992). Further experiments on yeast provided evidence that Spt6 interacts with histones H3 and H4 and acts as histone chaperone *in vitro* (Bortvin and Winston, 1996; Compagnone-Post and Osley, 1996). Spt6 cooperates with the methyltransferase Set2 in regulating histone modifications, such as H3K36me2 and H3K36me3 (Carrozza et al., 2005; Youdell et al., 2008). The chromatin changes and histone deposition behind the processing polymerase lead to gene repression and reflect on transcription dynamics. A possible role in elongation has been proposed by *spt6* mutants unable to reassemble nucleosomes (Adkins and Tyler, 2006; Ivanovska et al., 2011; Jensen et al., 2008) and, most importantly, by Spt6 interaction with the RNA polymerase II (Endoh et al., 2004; Yoh et al., 2007). Spt6 harbors tandem SH2 domains at its C-terminus, which are involved in the binding to the Serine 2-phosphorylated C-terminal domain (CTD) of the largest RNA polymerase II subunit Rpb1 (Close et al., 2011; Dengl et al., 2009; Diebold et al., 2010b; Liu et al., 2011; Mayer et al., 2010; Sun et al., 2010). Furthermore, Spt6 localizes to transcribed genes, proportionally to their transcription rate (Andrulis et al., 2000; Ivanovska et al., 2011; Kaplan et al., 2005; Kaplan et al., 2000; Krogan et al., 2002; Mayer et al., 2010), and has been described to repress yeast cryptic promoters located within coding genes during elongation (Cheung et al., 2008; Kaplan et al., 2003). Moreover, the Interacting with Spt6 homolog 1 (Iws1) protein has been proposed as “bridge factor” connecting Spt6 and Set2 to the Pol II CTD, in light of its interaction with Spt6 (Diebold et al., 2010a; Yoh et al., 2007; Yoh et al., 2008). Thus, it appears that characterizing the Spt6 interactome could contribute to a better understanding of its role in transcription regulation and chromatin remodeling.

Spt6 is also regulating mitotic recombination in yeast (Malagon and Aguilera, 2001), signal transduction in mammals (Baniahmad et al., 1995; Shen et al., 2009), as well as HIV transcription regulation and mRNA processing in human cells (Vanti et al., 2009; Yoh et al., 2007). Moreover, developmental studies have shown that Spt6 expression is regulated during zebrafish embryogenesis, *Drosophila* development and *C. elegans* gut morphogenesis (Ardehali et al., 2009; Keegan et al., 2002; Kok et al., 2007; Nishiwaki et al., 1993). Furthermore, Spt6 has also been involved in mRNA export and surveillance (Andrulis et al., 2002; Estruch et al., 2009). This function, in particular, appears intriguing as the interaction between Spt6 and the exosome has been described in *Drosophila*

and, as Spt5 is part of this complex (Andrulis et al., 2002) and the exosome is able to target AID to template and non-template ssDNA (Basu et al., 2011), this evidence let us wonder whether Spt6 is able to regulate AID as well. This question represents one of those we have addressed within this work and now it has an answer, provided by the work of Okazaki and collaborators (Okazaki et al., 2011), as Spt6 is shown to be required for CSR and dispensable for SHM but, as part of my work has been focused on Spt6, I will discuss later these results.

## **2. The PAF complex: the “transcription platform”**

As for the Spt proteins, the identification of the PAF complex was derived from studies performed on yeast, which allowed the identification of its five subunits: Paf1, Ctr9, Leo1, Rtf1 and Cdc73 (Figure 14) (Krogan et al., 2002; Wade et al., 1996). However, a small difference is present in the human PAF complex (hPAF), as it displays, in addition to the five subunits described in yeast, hSki8 which is a component of the human SKI complex, involved in 3'-5' mRNA degradation (Carpten et al., 2002; Kim et al., 2010; Rozenblatt-Rosen et al., 2005; Yart et al., 2005; Zhu et al., 2005). The PAF complex associates with the promoter and coding regions of transcribed genes (Kaplan et al., 2005; Kim et al., 2010; Kim et al., 2004; Mueller et al., 2004; Pokholok et al., 2002; Qiu et al., 2006; Rozenblatt-Rosen et al., 2009) and is able to associate with Pol II in three different states, regarding to the phosphorylation of its CTD: the unphosphorylated form, the serine 5 phosphorylated (S5-P) which is generally associated to transcription initiation and the S2-P in the elongation step to then detach from Pol II in proximity of the poly(A) site (Jaehning, 2010). The intriguing aspect of this complex is that it has been considered like a “platform”, able to recruit additional factors required for histone modification during transcription (Jaehning, 2010). Whereas recombinant Cdc73 is able to interact with purified Pol II (Shi et al., 1997), the PAF complex itself is recruited to the chromatin by Spt5: the C-terminal phosphorylation of the elongation factor appears to be required for PAF recruitment in yeast (Liu et al., 2009) while it seems to be dispensable in human (Chen et al., 2009). The PAF subunit involved in the interaction with Spt5 has not been identified yet, but one likely candidate is Rtf1 (Squazzo et al., 2002). In addition to Spt5, which interacts with PAF as member of the DSIF complex (Spt4/Spt5), other interactors may be required for PAF recruitment during transcription elongation, such as Spt6 (Kaplan et al., 2005) and Spt16, as part of the FACT complex (Pavri et al., 2006). Thus, considering that Spt5, Spt6 and the FACT complex have been addressed as regulators of CSR and SHM, the PAF complex seems to be the central element where all these factors converge and these observations justify the interest in addressing its role in antibody diversification, as it will be presented in the results and discussion section.



**Figure 14. Main roles of the PAF complex**

Schematic representation of the hPAF complex and its main roles presented in this dissertation. Adapted from Jaehning, 2010.

## 2.1. Role of the PAF complex in histone modifications and transcription

As associated to the Pol II during transcription, the PAF complex contributes to histone modification and seems to be indirectly required for H3K4me3 and H3K36me3 (Chu et al., 2007; Krogan et al., 2003a; Ng et al., 2003), two modifications found in yeast in proximity of the promoter and 3' half of the transcription unit, respectively (Jaehning, 2010). On one hand, the PAF complex coordinates H2B ubiquitylation Rad6/Bre1-mediated which is critical for H3K4me3 formation (Jaehning, 2010; Kim et al., 2009; Wood et al., 2003; Xiao et al., 2005); on the other hand is required for S2 phosphorylation at the CTD of the RNA polymerase II (Mueller et al., 2004; Nordick et al., 2008) which in turns mediates the recruitment of the methyltransferase Set2 and the consequent H3K36 trimethylation (Krogan et al., 2003b; Xiao et al., 2003).

Concerning its role in transcription, although the components of the complex have been identified with other elongation factors (Krogan et al., 2002; Squazzo et al., 2002), it seems that in yeast and flies loss of PAF does not impair transcription, in terms of Pol II density and distribution (Adelman et al., 2006; Mueller et al., 2004). However, hPAF and hSpt5 are described to stimulate transcription elongation *in vitro* (Chen et al., 2009), and a recent work shows as the hPAF has an intrinsic activity in elongation, which is not dependent on histone modifications (Kim et al., 2010).

Another step of transcription, which seems to be regulated by the PAF complex, is termination and, more specifically, the generation of the 3' end of the newly transcribed RNA. The identification of yeast genes whose expression was PAF-dependent revealed that, upon PAF depletion, these genes were transcribed at lower rate due to the generation of unstable transcripts consequent to the changes in the use of 3' end formation sites (Penheiter et al., 2005). In line with these results, loss of PAF leads to a reduction in the length of the poly(A) transcripts (Mueller et al., 2004). Whether these effects are secondary to the PAF-mediated histone modifications or direct consequences of PAF components

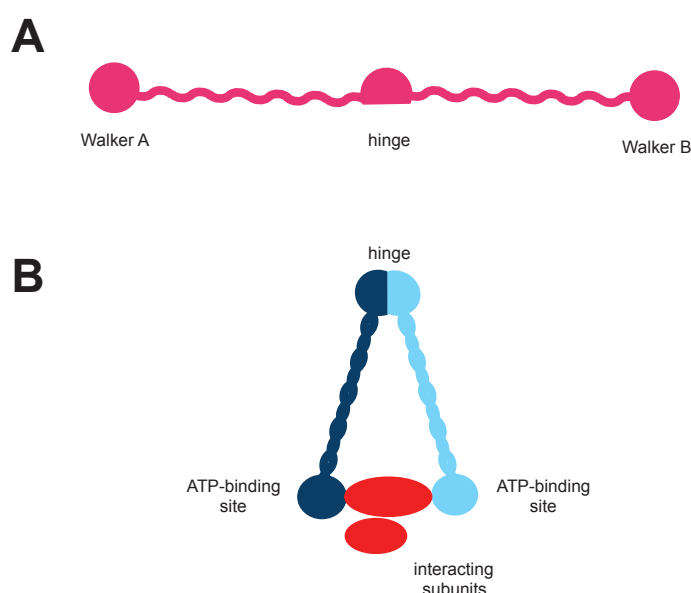
depletion is still under debate, but the direct interaction observed in yeast between the PAF complex and the polyadenylation factor Ctf1, independent on the Pol II (Nordick et al., 2008), suggests that this complex might be directly involved even in the final step of transcription.

As presented in these chapters, transcription seems to be pivotal in regulating AID-mediated antibody diversification and in particular class switching. However, by looking at the S regions, being poised for transcription is not enough to justify the efficient recombination between the donor and acceptor S region, in light of the huge intervening sequence within the locus. Thus, has been proposed that 3D rearrangements might contribute to juxtapose the S regions involved in the recombination and the Structural maintenance of chromosome (Smc) proteins have emerged as new “hot topic” in antibody diversification, as it will be presented in the next chapter.



## VI. The Structural maintenance of chromosomes (Smc) complexes in genome regulation

The structural maintenance of chromosomes (Smc) protein family represents a group of highly conserved proteins, from bacteria to eukaryotes, which share a common structure and organization and are classified in three distinct complexes, according to their core components: cohesin (Smc1 and Smc3), condensin (Smc2 and Smc4) and the Smc5/6 complex (Losada and Hirano, 2005). In eukaryotes, each complex is composed by an Smc heterodimer, which represents the functional core, and by accessory proteins, which dictate the dynamics of association and release from the chromosomes (Losada and Hirano, 2005). The interaction of the Smc complexes with chromosomes is dependent on their structure: the Smc proteins display Walker A and Walker B motifs, responsible of nucleotide binding and located respectively at the N-terminal and C-terminal domain, whereas the central domain is composed by a hinge sequence flanked by two coiled-coil motifs (Figure 15A)(Hirano, 2006). The two coiled-coil motifs are associated in an anti-parallel fashion, bringing the Walker A and B motifs in close proximity and forming an ATP-binding site (Haering et al., 2002; Hirano and Hirano, 2002; Melby et al., 1998). While the ATPase activity is required for Smcs function, the hinge domain allows the dimerization with the partner Smc protein and provides flexibility to the complex (Figure 15B, (Arumugam et al., 2003; Foustero and Lehmann, 2000; Hirano et al., 2001; Hirano and Hirano, 1998). The Smc complexes share a common structure and their main function is related to cell cycle regulation and chromosome dynamics; however, as my work focused on the cohesin complex and on the Smc5/6 complex, I will introduce the specialized functions that these proteins exert.

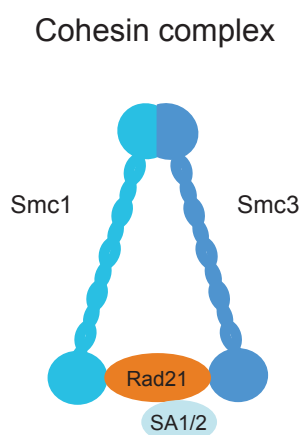


**Figure 15. Domain organization of the Smc protein and Smc complex**

(A) Smc protein structure, displaying a Walker A motif at the N-terminus, two coiled-coiled motifs with a hinge in between and a Walker B motif at the C-terminus. (B) Model of Smc complex and additional subunits.

## 1. The cohesin complex

The cohesin complex is composed by Smc1, Smc3, Rad21 and one of the stromal antigen proteins SA1 or SA2 (Figure 16 and Table 2), and its canonical role consists in regulating chromosome cohesion during mitosis and meiosis: precocious separation of sister chromatids is observed upon mutation of the cohesin subunits (Guacci et al., 1997; Michaelis et al., 1997), while the impairment of cohesin removal inhibits chromosomes separation and delays the progression through the earlier stages of mitosis (Kueng et al., 2006). In yeast, the cohesin complex is loaded onto chromatids during the S phase of the cell cycle until the anaphase (Guacci et al., 1997; Michaelis et al., 1997), whereas in higher eukaryotes the loading occurs during the telophase/G<sub>1</sub> phase until the next prophase, when just a little amount is maintained at the centromeres, to ensure a correct segregation, and then be released at anaphase (Losada et al., 1998; Waizenegger et al., 2000). The loading of cohesins onto chromosomes depends on ATP hydrolysis and on the complex formed by Nipbl/Sccl, which binds chromosomes and thus allows their interaction with cohesins (Dorsett and Strom, 2012); although the exact mechanism of cohesin binding is not fully elucidated, the ability of the hinge interface to bind DNA suggests that the complex might be opened at this interface to allow the DNA to be “embraced” (Dorsett and Strom, 2012). The establishment of cohesin requires, instead, the lysine acetyl transferase Establishment of cohesion 1 (Esccl), which acetylates Smc3 on K112 and K113, two residues highly conserved among eukaryotes (Rolef Ben-Shahar et al., 2008; Zhang et al., 2008). This modification of the core complex counteracts the activity of Pds5/Wapl, involved in cohesin dissociation from chromosomes (Dorsett and Strom, 2012; Kueng et al., 2006). Dissociation occurs once chromosomes are aligned on metaphase plate and chromatids captured by the mitotic spindle, and depends on Smc3 deacetylation mediated by the deacetylase HDAC8 (Beckouet et al., 2010; Borges et al., 2010; Dorsett and Strom, 2012; Xiong et al., 2010). This starts with the activity of the ubiquitin ligase APC/cyclosome, which degrades securin, the inhibitory partner of the cysteine protease separase. Once separase is active, it degrades Sccl and the cohesin complex is released from the DNA (Feeney et al., 2010).



**Figure 16. Cohesin complex**  
Schematic representation of the cohesin complex and its interacting subunits, Rad21 and SA1/2.

	<i>S. cerevisiae</i>	<i>S. pombe</i>	<i>D. melanogaster</i>	<i>X. laevis</i>	<i>H. sapiens/ M. musculus</i>
<b>SMC</b>	Smc1	Psm1	Smc1	Smc1	<b>Smc1<math>\alpha</math></b> , Smc1 $\beta$ (M)
	Smc3	Psm3	Smc3	Smc3	<b>Smc3</b>
<b><math>\alpha</math>-KLEISIN</b>	Scs1/Mcd1	Rad21	Rad21	Rad21	Scs1/Rad21
	Rec8 (M)	Rec8 (M)	C(2)M	-	Rad21L, Rec8 (M)
<b><math>\alpha</math>-KLEISIN INTERACTING SUBUNITS</b>	Scs3	Psc3	SA	SA1, SA2	SA1/Stag1, SA2/Stag2
	-	Rec11 (M)	-	-	SA3/Stag3
<b>REGULATORY FACTORS</b>	Pds5	Pds5	Pds5	Pds5A, Pds5B	Pds5A, Pds5B/APRIN
	Rad61/Wapl	Wapl	Wapl	Wapl	<b>Wapl/Wapal</b>
	-	-	Dalmatian	Sororin	Sororin
<b>LOADING COMPLEX</b>	-	Mis4	Nipped-B	Scs2	<b>Nipbl/Scs2</b>
	Scs4	Ssl3	Scs4	Scs4	Mau2/Scs4
<b>ACETYL TRANSFERASES</b>	Eco1/Ctf7	Eso1	Deco, San	Eso1, Eso2	Eso1, Eso2
<b>DEACETYLASES</b>	Hos1	-	-	-	HDAC8

**Table 2. Cohesin subunits and regulatory factors**

List of the cohesin complex components and regulatory factors in *S. cerevisiae*, *S. pombe*, *D. melanogaster*, *X. laevis*, *H. sapiens* and *M. musculus*. In red are evidenced the factors analyzed in this study; (M): meiosis-specific components. Adapted from Remeseiro and Losada, 2013.

As already mentioned, in addition to ensure a proper chromosome segregation cohesins play also additional roles, such as gene expression regulation. Cohesins bind heterochromatic regions at centromeres and telomeres, and in *S. cerevisiae* limit the silencing boundaries at the mating type loci (Dorsett and Strom, 2012). Studies conducted on *Drosophila*, instead, showed that Nipped-B is required for the expression of selected genes (Rollins et al., 1999), that cohesins bind to genomic loci occupied by the RNA polymerase II and regulate genes with paused polymerase II (Fay et al., 2011; Panigrahi and Pati, 2012). Additionally, the cohesin complex colocalizes with the insulator CCCTC-binding factor (CTCF) (Parelho et al., 2008; Wendt et al., 2008) and, along with CTCF, can either bring enhancers and promoter in close proximity or impair this communication, with a consequent negative effect on gene expression (Panigrahi and Pati, 2012). Furthermore, cohesins can recruit directly transcription factors (Schmidt et al., 2010) and regulate c-myc transcription in zebrafish through an evolutionarily conserved mechanism (Rhodes et al., 2010). Moreover, ChIP-Seq experiments performed in mouse ES cells revealed that Smc1 $\alpha$ , Smc3 and Nipbl colocalize with Mediator, a transcriptional regulator which facilitates the interaction between transcription factors and the RNA polymerase II, and the Pol II, and chromosome conformation capture (3C) assays revealed the formation of a loop, which leads to the interaction between the core promoter and the enhancer of the loci analyzed (Kagey et al., 2010). Thus, all these results show as, in addition to its canonical role, the cohesin complex is also important in transcription regulation and genomic loci accessibility.

Strikingly, the cohesin complex is also involved in DNA repair. Most likely, cohesins promote DSBs repair by keeping the chromatids together and thus assuring the presence of the template required for HR (Hagstrom and Meyer, 2003). This is supported by observations in *S. pombe*, where Scs1 was first identified as Rad21, a gene involved in DSBs repair (Birkenbihl and Subramani, 1992), and in *S. cerevisiae*, where mutations in Smc1, Smc3, Scs3 and Pds5 lead to sensitivity to  $\gamma$ -irradiations

(Hagstrom and Meyer, 2003). Additionally, Smc1 has been proposed to direct the choice between HR and NHEJ (Schar et al., 2004), and is phosphorylated by ATM upon DNA damage and S phase checkpoint activation (Kim et al., 2002; Yazdi et al., 2002). Furthermore, has been described that cohesin depletion leads to an impaired recruitment of 53BP1 to irradiation-induced DSBs (Watrin and Peters, 2009). Thus, these observations suggest a global control of cell cycle not only by ensuring the proper chromosomes separation but also allowing the fidelity of the repair.

### **1.1. Cohesins and Ig loci reorganization**

The “multi-tasking” activity of the cohesin complex has been the object of particular interest in the last years, in light of the long-range interactions described to occur at the Ig loci during B cells diversification. Concerning CSR, one of the main unanswered questions was: how can the donor and acceptor S regions, which are separated more than 100 Kb, relocate close to each other for efficient recombination and repair? Wuerffel et al. showed that the IgH locus displays a loop configuration which undergoes dynamic changes upon B cells activation: in resting B cells, the intronic enhancer  $E_{\mu}$  interacts with the 3'RR, and upon activation and AID expression the acceptor S region, poised for transcription as well as  $S_{\mu}$ , is actively relocated to the  $E_{\mu}/3'RR$  synapsis, and this is cytokine-specific as B cells stimulation with either LPS or LPS and IL-4 would recruit  $S_{\gamma 3}$  or  $S_{\gamma 1}$  regions, respectively (Wuerffel et al., 2007). Furthermore, a study focused on the role of the cohesin complex on early T cells development revealed as cohesins colocalize at regulatory regions of the  $TCR\alpha$  loci, and that Rad21 depletion in the developmental transition between the CD4-CD8 double negative to double positive thymocytes reduced the H3K4me3 mark, RAGs recruitment and DNA DSBs, due to a reduced enhancer/promoter interaction (Seitan et al., 2011). Moreover, the observations that Rad21 colocalizes with CTCF binding sites at IgH locus in pro-B cells, and that the depletion of Rad21 and CTCF affects the interaction between  $E_{\mu}$  and the  $D_H$  region (Degner et al., 2011; Degner et al., 2009), further support the idea that the cohesin complex plays an active role in the IgH reorganization during B cells differentiation. However, the precise role of cohesins during CSR has not been clarified yet and it represents one of the questions we have addressed during this study.

### **1.2. Cohesin deficiency and pathological consequences**

The deficiency of the cohesin complex components is the cause of human syndromes, such as Cornelia de Lange syndrome (CdLS) and Roberts-SC phocomelia syndrome. CdLS is caused by point mutations or small deletions or insertions in the genes coding for *NIPBL*, *SMC1* or *SMC3*; *NIPBL* mutations are more frequent and lead to a more severe form of the syndrome, whereas patients affected by *SMC* mutations represent mild cases (Deardorff et al., 2007; Horsfield et al., 2012). The symptoms include behavioral and cognitive defects, characteristic facial features and a common

gastroesophageal dysfunction. All patients have neurodevelopmental delay and variable mental retardation (Horsfield et al., 2012), and cells derived from CdLS patients are sensitive to the DNA-damaging agent mitomycin C (Vrouwe et al., 2007). Additionally, mutations in the *RAD21* gene also lead to a cohesinopathy syndrome: the patients affected display even milder phenotype compared to mutations detected in *SMC* genes, and patient-derived lymphoblastoid cell lines exhibit radiation sensitivity (Deardorff et al., 2012).

Roberts/SC phocomelia syndrome, instead, is caused by homozygous mutations in *ESCO2* gene, which lead to a truncated protein or to an inactive form (Vega et al., 2010). As *ESCO2* is expressed in embryonic tissues, patients harboring *ESCO2* mutations display both upper and lower limb defects, craniofacial defects and mental retardation (Vega et al., 2010). Moreover, cells from patients affected by this syndrome are hypersensitive to DNA damage caused by mitomycin C, camptothecin and etoposide (Bose and Gerton, 2010). Although *Eco1* is an essential gene in yeast, in human the two orthologs *ESCO1* and *ESCO2* are both involved in chromosome cohesion, and the redundant function between these two members might explain why mutations in *ESCO2* are compatible with life (Bose and Gerton, 2010).

## 2. The Smc5/6 complex

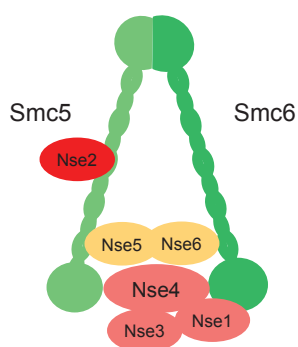
The Smc5/6 complex consists of Smc5 and Smc6 core subunits and additional non-Smc elements (Nse) proteins: Nse 1-6 in yeast and Nse 1-4 in human (Table 3). Studies conducted in yeast led to the identification of the complex components and, in *S. pombe*, has been described an additional subunit, Rad60, whose homolog in *S. cerevisiae* is named Esc2, which is associated to the complex (Kegel and Sjogren, 2010). The Smc5/6 complex displays the same domain organization of the other Smc proteins and the ATPase activity retained by the WalkerA/B interacting domains is stimulated by dsDNA (Fousteri and Lehmann, 2000); if this activity is impaired, it leads to cell sensitivity to DNA-damaging agents (Fousteri and Lehmann, 2000; Verkade et al., 1999). Three of the Nse subunits associate in proximity of the ATPase domains (Figure 17): Nse4 interacts with the ATPase head domains of Smc5 and Smc6, while Nse1 and Nse3 interact with each other and, in turn, with Nse4; Nse2, instead, is bound to the coiled-coil region of Smc5 through its N-terminal domain (Palecek et al., 2006; Sergeant et al., 2005). In yeast have been identified two additional Nse proteins, Nse5 and Nse6, and the connection with the complex core is different in *S. cerevisiae* and *S. pombe* (Kegel and Sjogren, 2010; Stephan et al., 2011b) (Figure 17). Interestingly, two of the additional components of the Smc5/6 complex display enzymatic activity: Nse1 has E3 ubiquitin ligase activity and acts with Nse3 (Doyle et al., 2010), whereas Nse2 displays small ubiquitin-like modifier (SUMO) ligase activity, which exerts on itself and also on Smc5 and Smc6 (Andrews et al., 2005; Potts, 2009; Potts and Yu, 2005; Zhao and Blobel, 2005).

Although the Smc5/6 complex is less characterized compared to cohesins and condensins, its main role appears to be the DSBs repair, as Smc5, Smc6 and Nse 1-4 genes are essential for viability in

yeast (Kegel and Sjogren, 2010). While these observations are consistent with the embryonic lethality observed in Smc6 knockout mice (Ju et al., 2013), Smc5 knockout in the DT40 chicken B cell line displays normal viability (Stephan et al., 2011a). Moreover, inactivation of the Smc5/6 complex in *S. cerevisiae*, plants, chickens and humans leads to sister chromatid HR defects (Wu and Yu, 2012). Additionally, the human Smc5/6 complex has been proposed to recruit cohesins specifically to DSBs, as RNAi-mediated depletion of MMS21 (Nse2) and Smc5 impairs the recruitment of Smc1 and Rad21 (Potts et al., 2006) and leads to a premature chromosome separation (Behlke-Steinert et al., 2009). Interestingly, this function exerted by Smc5 and MMS21 seems to be independent on Smc6 (Behlke-Steinert et al., 2009), suggesting that further investigations are required to understand whether the two core components of the Smc5/6 complex can also exert independent function. The Smc5/6-mediated cohesin recruitment could fit with a model where cohesins recruitment allows to hold sister chromatids together for efficient HR, and is supported by the observation that Nse2 is able to sumoylate Rad21 and SA2 (Potts et al., 2006). However, this is in contrast with what described in yeast, as *S. pombe smc6* mutants display persistent cohesins retention at DSBs, and the Smc5/6 complex does not seem to be required for cohesins recruitment (Outwin et al., 2009). This suggests that further investigations are required to delineate the exact relationship between the Smc5/6 complex and the cohesin complex. Moreover, studies conducted in yeast showed that Smc6 is involved in the G<sub>2</sub>/M checkpoint activation (Verkade et al., 1999) and that its recruitment to DSBs depends on Mre11 (Lindroos et al., 2006).

The Smc5/6 complex is also involved in the repair of collapsed replication forks, which occurs through HR, and in maintaining rDNA integrity, as mutation of the Smc5/6 complex leads to a defective segregation of rDNA during mitosis (Potts, 2009). Moreover, the Smc5/6 complex has been described to maintain telomere length in alternative lengthening of telomeres (ALT) cells, namely cancer cells which are unable to upregulate telomerase expression and thus regulate telomere length through recombination (Cesare and Reddel, 2010). Consistent with their involvement in HR, Smc5, Smc6 and MMS21 (Nse2) knockdown leads to telomere shortening and senescence in ALT cells, and the three proteins have been identified in the ALT-associated promyelocytic leukemia bodies (APBs), specialized compartments within the nucleus where the presence of HR components might facilitate recombination (Cesare and Reddel, 2010; Potts and Yu, 2007). Interestingly, the MMS21 SUMO ligase activity seems to be pivotal for telomere recruitment to APBs, possibly through modification of shelterin components (Cesare and Reddel, 2010). Thus, Smc5/6 complex appears, on one hand, as regulator of genome integrity and cell cycle progression, while on the other hand facilitates proliferation of cancer cells and understanding, in this latter case, the molecular targets of MMS21-mediated sumoylation and the effects exerted within the cells will contribute to further elucidate the role played by this complex.

## Smc5/6 complex



**Figure 17. Smc5/6 complex**

Schematic representation of the Smc5/6 complex and its interacting subunits. Note that the Smc5/6 complex depicted is the one characterized in *S. pombe* as, in *S. cerevisiae*, Nse5 and Nse6 interact with the hinge domain whereas in higher organisms these last two components are lacking.

	<i>S. cerevisiae</i>	<i>S. pombe</i>	<i>H. sapiens</i>
<b>SMC</b>	Smc5	Spr18/Smc5	<b>Smc5</b>
	Rhc18/Smc6	Rad18/Smc6	<b>Smc6</b>
<b>NSE</b>	Nse1	Nse1	Nse1
	Mms21/Nse2	Nse2	Nse2
	YDR228W/Nse3	Nse3	Nse3
	Qri2/Nse4	Rad62/Nse4	Nse4
	YML023C/Nse5	Nse5	-
	Kre29/Nse6	Nse6	-

**Table 3. Smc5/6 complex subunits and associated non-Smc elements (Nse)**

List of the Smc5/6 complex components and regulatory factors in *S. cerevisiae*, *S. pombe*, and *H. sapiens*. In red are evidenced the factors analyzed in this study. Adapted from Wu and Yu, 2012.

## VII. Working hypothesis

Despite the considerable progress of the last years towards a better understanding of the mechanisms which dictate antigen diversification, how class switch recombination is regulated is not well understood. In particular, the targeting of AID to the IgH locus, the juxtaposition the S regions involved in the reaction and the repair pathways which mediate the resolution of DSBs are not fully elucidated. In order to clarify these mechanisms, we focused on different aspects of CSR regulation and the working hypothesis of my thesis was built on several observations regarding the requirement of the C-terminal domain of AID, deriving from the phenotype displayed by patients affected by a CSR-ID due to an unknown defect. These patients display a specific defect in CSR, which is not associated to a defect in SHM, and express AID at levels comparable to controls. Moreover, the undetectable DSBs at the S $\mu$  regions suggest that AID-mediated deamination is not occurring (Durandy et al., 2007).

This phenotype is reminiscent of mutations or truncations located in the C-terminal domain of AID (Ta et al., 2003) and of AID<sup>-/-</sup> B cells reconstituted with a C-terminal deleted mutant, which are able to undergo IGC, SHM but not CSR (Barreto et al., 2003).

This intriguing phenotype and the absence of breaks at the S $\mu$  region suggest an impaired AID targeting and, based on these observations, we hypothesized that AID is able to associate, through its C-terminal domain, with factors specifically involved in the CSR reaction. We wanted to identify proteins required for the specific regulation of AID during CSR; thus my thesis focused on the identification of candidate factors and on their functional characterization in CSR.



# RESULTS

## I. Overview of thesis work

In order to better understand the mechanisms which regulate CSR, the aim of my thesis was to identify factors specifically required for this process. Therefore, we established collaboration with the laboratory of lymphocyte interactions and lymphocytes B terminal maturation headed by Dr. Anne Durandy (Necker Hospital, Paris) in order to characterize the molecular defect identified in patients affected by CSR-ID. These patients displayed impairment in CSR, associated to normal AID expression and, surprisingly, normal frequency and pattern of SHM (Durandy et al., 2007). This phenotype resembles to the one associated to mutations in the C-terminus of AID, which suggests that CSR-specific factors would preferentially interact with this domain (Barreto et al., 2003; Durandy et al., 2007; Geisberger et al., 2009; Imai et al., 2005; McBride et al., 2004; Ta et al., 2003). As no DSBs were detected at the donor S $\mu$  region in CSR-ID patient-derived B cells, the block in switching would most likely occur before AID-mediated deamination. In light of this evidence, we hypothesized that the CSR-ID patients are deficient for a factor required to target AID to the switch regions, and that this factor interacts with the C-terminal domain of AID. For this study, we used seven EBV-immortalized B cell lines isolated from peripheral blood mononuclear cells of three healthy donors (Ctr) as positive controls, one patient affected by CSR-ID due to a loss of AID (AID<sup>-/-</sup>) as negative control and three patients affected by CSR-ID due to a specific defect in class switch recombination (CSR-ID, Table 4).

Cell line	Type	AID expression	SHM	CSR	other
Ctr	healthy donor	+	+	+	
Pat	CSR-ID (unknown AID factor)	+	+	-	no DSBs at S $\mu$
AID <sup>-/-</sup>	CSR-ID (AID deficiency)	-	-	-	

**Table 4. Characteristics of the human B cell lines analyzed in this study**

In order to identify this factor, we applied a multi-approach strategy: while our collaborators were focusing on the genome, by looking at mutations in the patients through linkage analysis and exome sequencing, we performed a transcriptome profiling by Digital Gene Expression-tag profiling (DGE) of the CSR-ID cell lines and relative controls, to identify deregulated genes which would be good candidates for sequencing in patients. We obtained a list of deregulated genes in CSR-ID patients, but no mutations in our candidate genes were found by sequencing. Although, if the mutation would not affect the expression of the gene, we would not be able to detect it by DGE but, instead, the impaired targeting of AID to the switch regions could be dependent on lower protein expression. To identify proteins under-represented or missing in CSR-ID patients, we undertook a proteome screening on B cells isolated from CSR-ID patients, healthy donors and AID<sup>-/-</sup> patient. We extracted nuclear proteins and we analyzed the proteome composition by mass spectrometry (MS). Furthermore, we

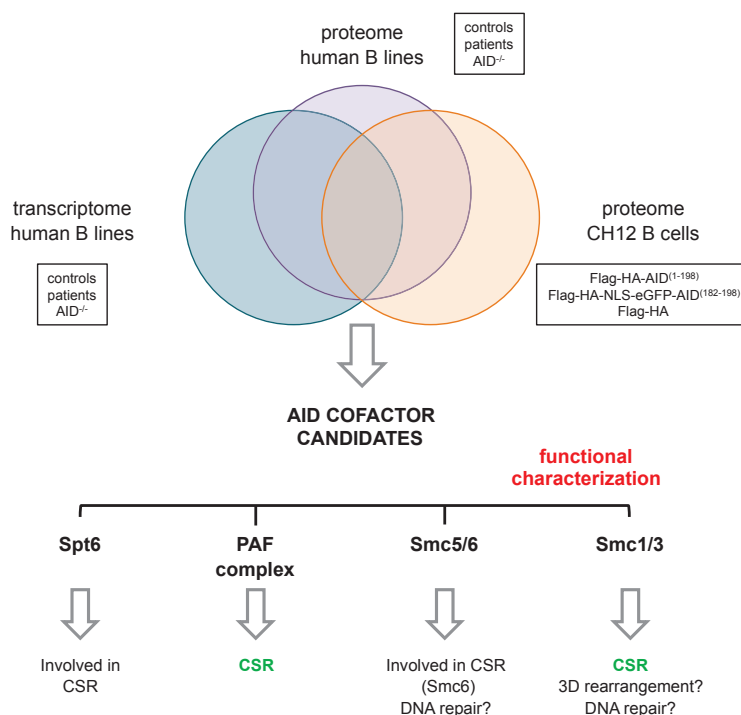
complemented our analysis on human B lymphocytes with the proteome identified in the mouse CH12 B cell line overexpressing either full length tagged-AID (Flag-HA-AID<sup>(1-198)</sup>) or the C-terminal domain of the protein (Flag-HA-NLS-eGFP-AID<sup>(182-198)</sup>). As AID C-terminus has been proposed to interact with factors specifically involved in CSR, by using CH12 B cells expressing this domain we would be able to focus on CSR-specific AID partners (Figure 18).

In addition to factors previously described to play a role in CSR such as eEF1A (Hasler et al., 2011), YY1 (Zaprazna and Atchison, 2012), RPA (Chaudhuri et al., 2004), 14-3-3 (Xu et al., 2010), CTNBL1 (Conticello et al., 2008) and Spt5 (Pavri et al., 2010), we identified the histone chaperone and elongation factor Spt6. As this factor was identified in the proteome of healthy donor-derived lines while it was lacking in CSR-ID patients, and western blot analysis suggested that Spt6 protein levels were reduced in CSR-ID patients when compared to controls, we undertook its functional characterization to understand whether Spt6 could play any role in CSR. We silenced *Spt6* gene expression in mouse CH12 B cells by using retroviral vectors expressing shRNAs targeting Spt6, but our pilot experiments showed that this system was not as efficient as expected. Thus, we decided to use lentiviruses to knockdown Spt6 expression. However, by the time we were performing our experiments, a publication from the laboratory of Tasuku Honjo showed that Spt6 was involved in CSR regulation (Okazaki et al., 2011), confirming the hypothesis we aimed to address. Nevertheless, part of the results we obtained while investigating Spt6 have supported the work performed in collaboration with the laboratory of Dr. Svend Petersen-Mahrt and focused on the role of the PAF complex, known to regulate histone modifications during transcription, in CSR. By performing co-immunoprecipitation experiments in CH12 B cells overexpressing tagged AID, we showed that nuclear AID is in the same complex than Spt5, Spt6, the RNA polymerase II and three subunits of the PAF complex: Paf1, Leo1 and Ctr9.

Moreover, as a mutation in the gene coding for *SMC5* in a CSR-ID patient was found, and identifying Smc5 in the proteome of one healthy donor-derived B line as well as interacting with the full length AID in CH12 cells, we verified its expression in CSR-ID and controls B cells. Smc5 is a member of the structural maintenance of chromosomes (Smc) family, and is mainly involved in DSBs repair through homologous recombination (Kegel and Sjogren, 2010). We observed variable expression levels of Smc5 in CSR-ID patients when compared to controls and we decided to address whether it could play a role in CSR. We knocked down *Smc5* gene expression in CH12 B cells and we observed contradictory results, which suggested that it might not be required for efficient switching. Furthermore, as Smc5 forms a heterodimer with Smc6, we addressed whether Smc6 could be required for switching. By using recombinant lentiviruses, we knocked down its expression in CH12 cells and we observed a significant impairment in CSR. However, further investigations will be required to clarify the role played by Smc5 and Smc6 in antibody diversification.

The proteome analysis we performed on human and mouse B cells represented a powerful tool to identify AID interactors and, in addition to the Smc5/6 complex, we found two additional complexes belonging to the Smc family: Smc2/4 (condensins) and Smc1/3 (cohesins). The latter has been investigated in our lab, as cohesins have been described to mediate long-range interactions at the Ig and TCR loci (Degner et al., 2011; Seitan et al., 2011). As during switching the juxtaposition of donor

and acceptor S regions is a mandatory step to allow efficient recombination and the expression of the downstream exon coding for a different antibody isotype, and taking into account the size of the IgH locus which suggests a 3D reorganization (Kenter et al., 2012), we hypothesized that cohesins may be involved in mediating the interaction between donor and acceptor S regions. By focusing on Smc1 and Smc3, as well as their loading and unloading factors into the DNA, Nipbl and Wapal, we observed by co-immunoprecipitation that they exist in a complex with AID. Thus we performed ChIP-Seq experiments on resting and activated mouse B cells, and we observed that Smc1 and Smc3 are actively recruited at the S $\mu$ -C $\mu$  region of the IgH locus upon activation. Furthermore, we characterized the role of cohesins in CSR by silencing their expression in CH12 B cells. We observed impairment in CSR upon Smc1, Smc3 and Nipbl knockdown, which was not due to a lower AID expression nor to reduced germline transcription at the donor and acceptor S regions. As cohesins are involved in homologous recombination during meiosis, we decided to address whether they could be involved in the repair step of CSR. We sequenced the switch junctions and analyzed the microhomology usage during repair. We observed that, upon Smc1, Smc3 and Nipbl depletion, the usage of microhomologies was biased in favor of longer microhomologies, which is a hallmark of DNA repair mediated by the A-NHEJ pathway. Our results suggest that cohesins could regulate switching by mediating the long-range interactions at the IgH locus and also by influencing the choice of the repair pathway involved in the resolution of AID-mediated DSBs.



**Figure 18. Workflow of identification and functional characterization of CSR-specific regulators**

Scheme depicting the workflow applied to identify factors specifically involved in CSR regulation. The integration of data obtained from the transcriptome profiling of EBV-immortalized human B cells and from the proteomic screening performed on human and mouse B cells allowed the identification of potential candidates in CSR regulation. Functional characterization for Spt6, the PAF complex, Smc5/6 complex and Smc1/3 complex has been performed, leading to the conclusions indicated below.

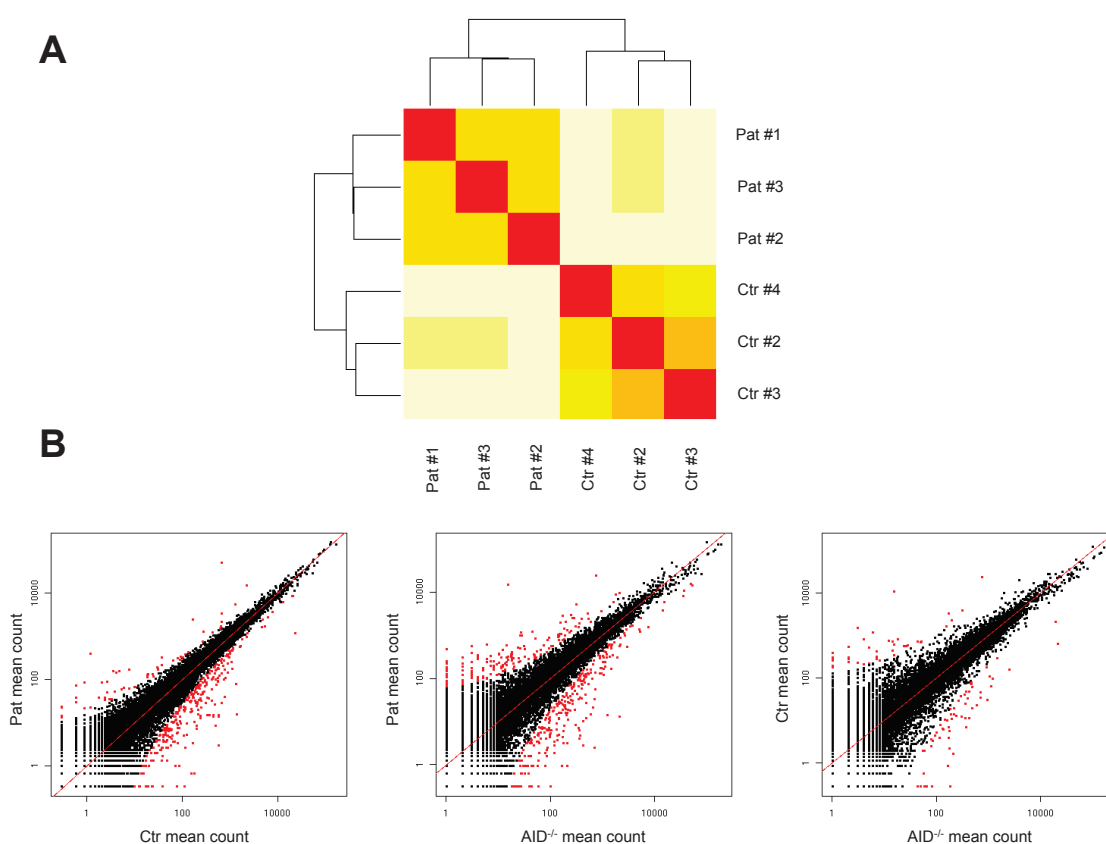
## II. Identification of CSR-specific factors

### 1. Transcriptome profiling of human B cell lines derived from healthy donors, CSR-ID and AID deficient-patients

My thesis project started with the characterization of the CSR defect identified in CSR-ID patients by the laboratory of Dr. Anne Durandy. These patients express AID at normal levels and display normal frequency and pattern of SHM, and their phenotype resembles to the one due to AID C-terminal mutations, suggesting that the specific impairment in CSR is due to a lack of a CSR cofactor able to interact with the C-terminus of AID. Furthermore, as in CSR-ID patients no DSBs were detected at  $S_{\mu}$ , we hypothesized that the missing AID cofactor would be required for efficient targeting of AID at the S regions. In order to identify the misregulated factor(s), we decided to analyze the transcriptome of three CSR-ID patients (Pat), three healthy donors (Ctr) and one AID<sup>-/-</sup>-derived (AID<sup>-/-</sup>) B lines by Digital Gene Expression – tag profiling (DGE) and compare deregulated genes in CSR-ID patients to controls. This technique allows the quantification of transcripts by sequencing 16 nt tags derived by two sequential digestions of polyadenylated total mRNA in the cells. After sequencing, the data is automatically filtered according to the fluorescence signal detected and to the tags sequence, in order to unequivocally identify the transcript. Upon data normalization and coverage calculation (number of transcripts detected over genome), we performed a quality check to validate our analysis by comparing controls and patients and calculating the correlation between samples. Correlation analysis can be represented as a heat map, where the colour spectrum is ranging from red (high correlation) to beige (low correlation; Figure 19A). We analyzed each group of samples (healthy donor controls and CSR-ID patients) to verify whether we could pool them and compare to the other two conditions as a group and not as individual units (Figure 19). As shown by the heat map, the CSR-ID lines (Pat) cluster together, and the same is observed for the group of healthy donor controls (Ctr); on the other hand, each control poorly correlate with the patients' lines and *vice versa*, as indicated by the light yellow/beige squares (Figure 19A). This result suggests that the group of patient-derived cell lines can be compared with the group of controls, as they appear to be independent data sets.

To avoid false positives in our analysis, we filtered the list of expressed genes according to the p value adjusted (padj), which represents p values adjusted for multiple testing with the Benjamini-Hockberg procedure, which controls the false discovery rate. We filtered our data according to the padj<0.05 for controls vs. patients and controls vs. AID<sup>-/-</sup> and padj<0.01 for patients vs. AID<sup>-/-</sup>. In this latter case, as we disposed of only one negative control (AID<sup>-/-</sup>), we increased the confidence interval to avoid false discoveries. Then, to identify deregulated genes, we filtered our data according to the log<sub>2</sub>fold change (log<sub>2</sub>FC)<1 (downregulated) or >1 (upregulated) and we plotted them according to the mean tags count for each transcript as scatter plot (Figure 19B). The scatter plots display two data sets (for instance controls vs. patients) and the x,y coordinates of each dot, which represents one transcript, depend on the mean expression levels of that particular transcript within the two conditions: equal expression in both conditions will locate the dot along the diagonal, whereas dots located below or

above the diagonal indicate deregulated genes in the condition indicated on the x and y axes, respectively. In Figure 19B, differentially expressed genes, according to the conditions we applied, are represented by red dots. We observed that most of the data scattered along the diagonal, indicating a similar expression level between patients and controls, controls and AID<sup>-/-</sup>, patients and AID<sup>-/-</sup> (Figure 19B). However, we identified some differentially expressed genes, indicated by the red dots: these were mainly downregulated in patients vs. controls, both upregulated and downregulated in patients vs. AID<sup>-/-</sup> and, to a lower extent, upregulated and downregulated in controls vs. AID<sup>-/-</sup> (Figure 19B). This result indicates that there are differentially expressed genes in each of the three conditions analyzed, thus we obtained three lists of deregulated genes (Tables 5, 6 and 7).



**Figure 19. Comparison of transcriptomic data obtained from control, patient and AID<sup>-/-</sup> EBV-immortalized B cell lines**

(A) Heat map displaying the correlation between CSR-ID patient (Pat) and healthy donor (Ctr) groups of samples calculated through the variance stabilizing transformation (VST) function. Red colour represents high correlation, whereas lighter colour represents low correlation. (B) Scatter plots displaying gene expression comparison between the three different data sets (Pat vs. Ctr; Pat vs. AID<sup>-/-</sup> and Ctr vs. AID<sup>-/-</sup>). Red dots located below or above the diagonal represent differentially expressed genes in the condition indicated on x or y axis, respectively.

Deregulated genes in patients when compared to AID<sup>-/-</sup>

Gene symbol	Description	Pat#1	Pat#2	Pat#3	AID <sup>-/-</sup>	log <sub>2</sub> FC	padj
JUP	junction plakoglobin	1	0	7	682	-8,29	1,78E-50
CNTNAP4	contactin associated protein-like 4	3	0	0	204	-7,93	3,29E-21
TP53I11	tumor protein p53 inducible protein 11	1	0	0	57	-7,67	1,16E-08
IGFL2	IGF-like family member 2	0	0	3	162	-7,64	7,17E-18
<b>PSMA4</b>	<b>proteasome (prosome, macropain) subunit, alpha type, 4</b>	<b>14</b>	<b>16</b>	<b>12</b>	<b>2053</b>	<b>-7,53</b>	<b>1,16E-91</b>
MYEOV	myeloma overexpressed (in a subset of t(11;14) positive multiple myelomas)	0	1	0	43	-7,44	3,02E-06
KRT17	keratin 17	6	0	3	413	-7,37	2,50E-36
LGALS2	lectin, galactoside-binding, soluble, 2	0	1	0	38	-7,26	1,60E-06
TNNI1	troponin I type 1 (skeletal, slow)	0	1	0	27	-6,77	4,91E-05
GABRB2	gamma-aminobutyric acid (GABA) A receptor, beta 2	4	1	8	357	-6,66	5,97E-33
<b>MRPS6</b>	<b>mitochondrial ribosomal protein S6</b>	<b>25</b>	<b>6</b>	<b>15</b>	<b>1220</b>	<b>-6,6</b>	<b>2,52E-71</b>
AZGP1	alpha-2-glycoprotein 1, zinc-binding	0	1	0	23	-6,53	0
IGKJ5	immunoglobulin kappa joining 5	3	195	47041	15	9,74	3,60E-42
<b>SPARC</b>	<b>secreted protein, acidic, cysteine-rich (osteonectin)</b>	<b>382</b>	<b>468</b>	<b>708</b>	<b>1</b>	<b>8,7</b>	<b>1,22E-18</b>
UGT2B17	UDP glucuronosyltransferase 2 family, polypeptide B17	196	324	267	1	7,7	3,74E-12
<b>P2RX1</b>	<b>purinergic receptor P2X, ligand-gated ion channel, 1</b>	<b>161</b>	<b>253</b>	<b>310</b>	<b>1</b>	<b>7,58</b>	<b>1,66E-11</b>
NGFRAP1	nerve growth factor receptor (TNFRSF16) associated protein 1	181	227	1396	3	7,34	1,62E-20
GUCY1A3	guanylate cyclase 1, soluble, alpha 3	255	264	68	1	7,28	7,16E-09
SLC12A7	solute carrier family 12 (potassium/chloride transporters), member 7	87	409	43	1	7,1	5,09E-09
<b>IL1A</b>	<b>interleukin 1, alpha</b>	<b>331</b>	<b>454</b>	<b>201</b>	<b>2</b>	<b>7,02</b>	<b>8,41E-13</b>
ARSD	arylsulfatase D	167	141	170	1	7	2,02E-08
DEPDC7	DEP domain containing 7	231	181	57	1	6,96	3,03E-08
GBP5	guanylate binding protein 5	677	411	160	3	6,8	1,68E-16
ASCL1	achaete-scute complex homolog 1 (Drosophila)	218	37	145	1	6,77	6,21E-07

Table 5. Deregulated genes in patients when compared to AID<sup>-/-</sup>

Example of deregulated genes in the CSR-ID patients (Pat) when compared to the AID<sup>-/-</sup> negative control. Gene symbol, description, normalized tags number for each sample, log<sub>2</sub>FC and padj are shown. In red are indicated those genes whose expression has been verified by RT-qPCR; the line divides the set of downregulated genes (top) from the upregulated ones (bottom).

Deregulated genes in controls when compared to AID<sup>-/-</sup>

Gene symbol	Description	Ctr#3	Ctr#4	Ctr#2	AID <sup>-/-</sup>	log <sub>2</sub> FC	padj
<i>SERPINA9</i>	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 9	0	0	1	177	-9,45	5,23E-11
<i>MAGEA1</i>	melanoma antigen family A, 1 (directs expression of antigen MZ2-E)	0	1	0	80	-8,27	1,04E-06
<i>LY6D</i>	lymphocyte antigen 6 complex, locus D	1	0	0	64	-7,99	7,67E-06
<i>TP53I11</i>	tumor protein p53 inducible protein 11	0	0	1	57	-7,82	1,78E-05
<i>IGFL2</i>	IGF-like family member 2	1	2	0	162	-7,72	2,18E-10
<i>CNTNAP4</i>	contactin associated protein-like 4	4	1	1	204	-7,07	3,12E-12
<i>F5</i>	coagulation factor V (proaccelerin, labile factor)	0	4	4	173	-6,4	8,86E-11
<i>MAFA</i>	v-maf musculoaponeurotic fibrosarcoma oncogene homolog A (avian)	0	0	2	29	-5,84	0,02
<i>ISLR2</i>	immunoglobulin superfamily containing leucine-rich repeat 2	2	0	1	32	-5,4	0,02
<i>AC096579.3</i>	Ig kappa chain C region	58	604	1617	21035	-5,18	8,56E-21
<i>PPP4R4</i>	protein phosphatase 4, regulatory subunit 4	1	3	16	176	-5,12	4,60E-06
<b><i>TESC</i></b>	<b>tescalcin</b>	<b>18</b>	<b>64</b>	<b>26</b>	<b>913</b>	<b>-5,04</b>	<b>2,19E-30</b>
<i>CRYBA4</i>	crystallin, beta A4	0	4	3	58	-5,01	0
<i>SGK493</i>	protein kinase domain containing, cytoplasmic homolog (mouse)	2	10	2	111	-4,95	0
<b><i>PRF1</i></b>	<b>perforin 1 (pore forming protein)</b>	<b>0</b>	<b>0</b>	<b>5</b>	<b>37</b>	<b>-4,87</b>	<b>0,02</b>
<i>IGKJ5</i>	immunoglobulin kappa joining 5	2	64	39316	15	9,38	6,74E-22
<i>NGFRAP1</i>	nerve growth factor receptor (TNFRSF16) associated protein 1	0	2031	3479	3	8,87	1,80E-14
<i>SPARC</i>	secreted protein, acidic, cysteine-rich (osteonectin)	414	309	692	1	8,49	3,30E-07
<i>EPS8</i>	epidermal growth factor receptor pathway substrate 8	42	886	175	1	8,15	9,24E-07
<i>UGT2A3</i>	UDP glucuronosyltransferase 2 family, polypeptide A3	1	1042	15	1	8,1	1,22E-06
<i>GUCY1A3</i>	guanylate cyclase 1, soluble, alpha 3	6	565	485	1	8,08	1,26E-06
<b><i>P2RX1</i></b>	<b>purinergic receptor P2X, ligand-gated ion channel, 1</b>	<b>612</b>	<b>25</b>	<b>368</b>	<b>1</b>	<b>7,99</b>	<b>1,06E-05</b>
<i>UGT2B17</i>	UDP glucuronosyltransferase 2 family, polypeptide B17	131	697	104	1	7,9	2,31E-05
<i>GUCY1B3</i>	guanylate cyclase 1, soluble, beta 3	8	1629	1267	4	7,54	2,08E-11
<i>CPXM1</i>	carboxypeptidase X (M14 family), member 1	477	3384	395	6	7,51	1,06E-10
<i>SLC12A7</i>	solute carrier family 12 (potassium/chloride transporters), member 7	208	96	380	1	7,44	0
<i>SLC25A24</i>	solute carrier family 25	98	510	42	1	7,39	0
<b><i>IL1A</i></b>	<b>interleukin 1, alpha</b>	<b>399</b>	<b>681</b>	<b>24</b>	<b>2</b>	<b>7,14</b>	<b>8,99E-06</b>

**Table 6. Deregulated genes in controls when compared to AID<sup>-/-</sup>**

Example of deregulated genes in the healthy donors (Ctr) when compared to the AID<sup>-/-</sup> negative control. Gene symbol, description, normalized tags number for each sample, log<sub>2</sub>FC and padj are shown. In red are indicated those genes whose expression has been verified by RT-qPCR; the line divides the set of downregulated genes (top) from the upregulated ones (bottom).



## Deregulated genes in patients when compared to controls

Gene symbol	Description	Pat#1	Pat#2	Pat#3	Ctr#3	Ctr#4	Ctr#2	log <sub>2</sub> FC	padj
<i>TNFRSF11B</i>	tumor necrosis factor receptor superfamily, member 11b	0	1	0	0	366	13	-8,62	9,90E-18
<i>TMEM176A</i>	transmembrane protein 176A	1	1	0	0	585	8	-8,18	3,27E-25
<i>TMEM176B</i>	transmembrane protein 176B	1	0	0	0	255	2	-7,89	4,22E-13
<i>MMP7</i>	matrix metalloproteinase 7 (matrilysin, uterine)	24	16	8	0	4604	8	-6,54	3,58E-88
<i>ACSS3</i>	acyl-CoA synthetase short-chain family member 3	0	1	0	2	73	2	-6,32	2,38E-05
<i>SLPI</i>	secretory leukocyte peptidase inhibitor	1	0	0	5	67	7	-6,19	1,86E-05
<i>NR2F2</i>	nuclear receptor subfamily 2, group F, member 2	9	11	6	0	1628	24	-5,96	1,54E-51
<b><i>PSMA4</i></b>	<b>proteasome (prosome, macropain) subunit, alpha type, 4</b>	<b>14</b>	<b>16</b>	<b>12</b>	<b>14</b>	<b>1939</b>	<b>11</b>	<b>-5,51</b>	<b>3,51E-57</b>
<i>NETO1</i>	neuropilin (NRP) and tolloid (TLL)-like 1	0	1	7	1	99	285	-5,51	1,10E-12
<i>AGTR2</i>	angiotensin II receptor, type 2	0	4	0	2	155	5	-5,4	2,91E-07
<i>MRPS6</i>	mitochondrial ribosomal protein S6	25	6	15	9	1914	17	-5,32	9,06E-34
<i>PRRX1</i>	paired related homeobox 1	8	33	0	3	1472	0	-5,19	1,56E-29
<i>DTNA</i>	dystrobrevin, alpha	1	1	2	4	120	1	-4,92	1,24E-05
<i>KAL1</i>	Kallmann syndrome 1 sequence	5	6	5	2	418	2	-4,68	1,34E-12
<i>LAMA1</i>	laminin, alpha 1	11	13	21	14	1086	18	-4,59	1,28E-25
<i>NFIB</i>	nuclear factor I/B	5	9	1	7	321	28	-4,56	3,23E-12
<i>IGKV4-1</i>	immunoglobulin kappa variable 4-1	3	2908	641	10	76853	9	-4,47	1,30E-80
<i>MMP9</i>	matrix metalloproteinase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	2	7	0	0	182	14	-4,46	1,90E-06
<i>GRB10</i>	growth factor receptor-bound protein 10	1	4	3	2	161	3	-4,36	5,69E-07
<i>IGLV6-57</i>	immunoglobulin lambda variable 6-57	67	23	31	2619	12	3	-4,34	2,57E-39
<i>RAB3C</i>	RAB3C, member RAS oncogene family	35	4	0	1	787	13	-4,27	4,06E-21
<i>EEF1A2</i>	eukaryotic translation elongation factor 1 alpha 2	35	8	38	9	1582	3	-4,22	4,99E-31
<b><i>PFN2</i></b>	<b>profilin 2</b>	<b>111</b>	<b>83</b>	<b>41</b>	<b>1128</b>	<b>2098</b>	<b>1306</b>	<b>-4,21</b>	<b>3,27E-49</b>
<i>PXDN</i>	peroxidasin homolog (Drosophila)	745	299	150	2	1	1	8,31	7,68E-25
<i>AC010760.1</i>	Putative uncharacterized protein	0	215	1	0	0	2	6,73	1,56E-08
<b><i>SERPING1</i></b>	<b>serpin peptidase inhibitor, clade G (C1 inhibitor), member 1</b>	<b>232</b>	<b>12</b>	<b>17</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>6,58</b>	<b>9,14E-10</b>
<i>CYP2S1</i>	cytochrome P450, family 2, subfamily S, polypeptide 1	56	7	4	0	0	1	6,19	0
<i>AC096579.3</i>	Ig kappa chain C region	57	107712	48105	58	604	1617	6,1	2,38E-78
<i>IGHV3-21</i>	immunoglobulin heavy variable 3-21	116	0	0	2	0	0	6	0
<i>AC087749.2</i>	Putative myosin-XVB	16	50	23	0	0	2	5,51	0
<i>SLC7A3</i>	solute carrier family 7 (cationic amino acid transporter, y+ system), member 3	0	43	1	1	0	0	5,44	0,03
<i>LDHC</i>	lactate dehydrogenase C	224	241	6	5	1	6	5,35	5,00E-11
<i>TJP1</i>	tight junction protein 1 (zona occludens 1)	152	52	53	5	0	2	5,3	5,42E-09
<i>ZNF677</i>	zinc finger protein 677	37	40	195	0	0	9	5	7,62E-09
<b><i>LAD1</i></b>	<b>ladinin 1</b>	<b>191</b>	<b>237</b>	<b>3</b>	<b>0</b>	<b>3</b>	<b>11</b>	<b>4,99</b>	<b>5,66E-12</b>

Table 7. Deregulated genes in patients when compared to controls

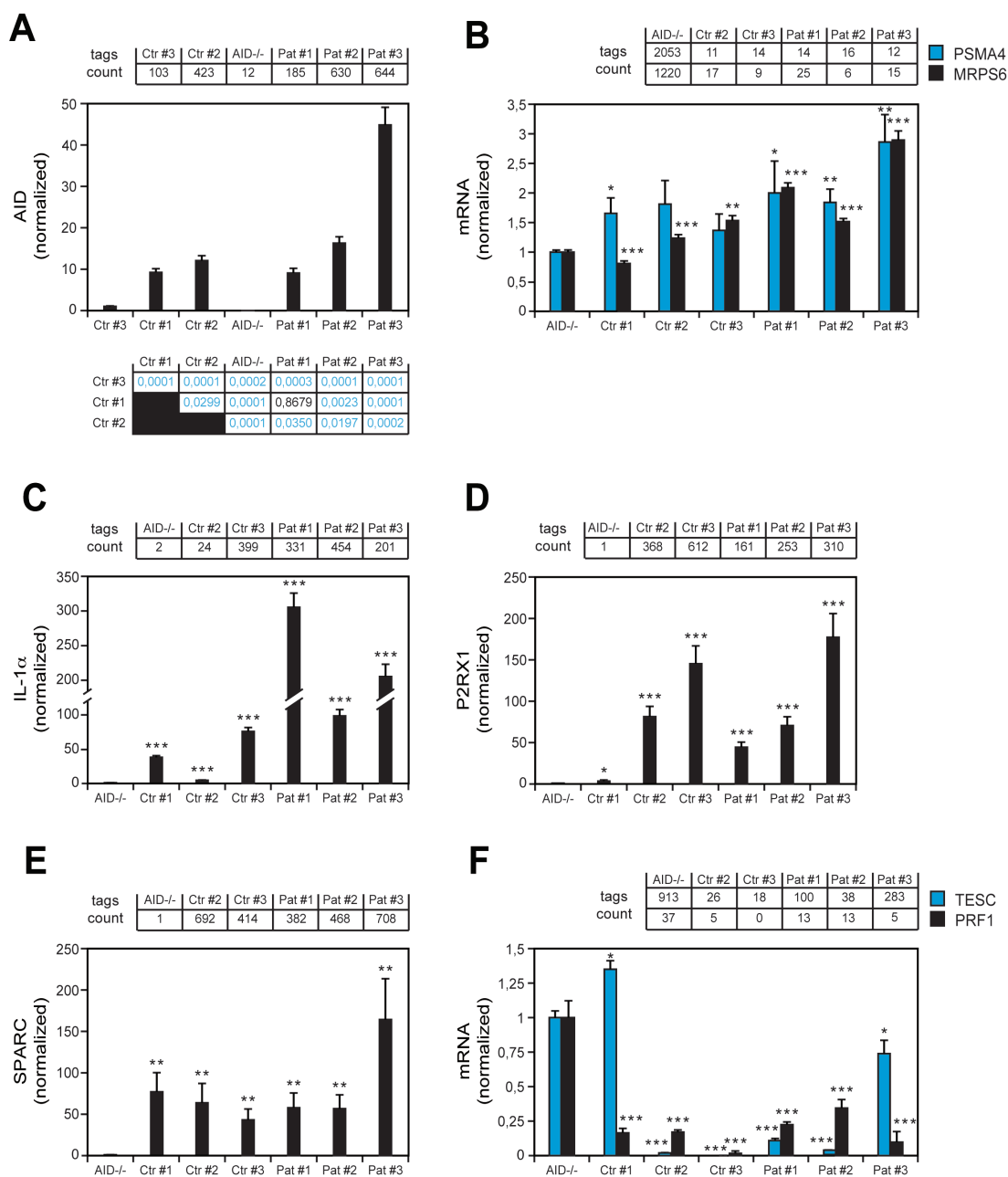
Example of deregulated genes in the CSR-ID patients (Pat) when compared to the healthy donor (Ctr) positive controls. Gene symbol, description, normalized tags number for each sample, log<sub>2</sub>FC and padj are shown. In red are indicated those genes whose expression has been verified by RT-qPCR; the line divides the set of downregulated genes (top) from the upregulated ones (bottom).

In order to validate the DGE data, we performed RT-qPCR on those genes highly upregulated or downregulated in each condition analyzed and indicated in Tables 5, 6 and 7 (Figures 20 and 21). As we disposed of limited amount of material from the Ctr #4, which was included in the transcriptome analysis, we were not able to include this sample in the validation tests; however, we included another control cell line (Ctr #1, Figures 20 and 21).

First of all, we assessed AID expression levels and we normalized our data to Ctr #3 (Figure 20A). By taking into account the expected variability between the human B lines in terms of expression levels for the genes tested and with the goal to avoid any bias in RT-qPCR data analysis, statistical significance has been calculated as relative to each of the control lines, as displayed in the table below the histogram (Figures 20A and 21). We observed a more than tenfold difference in AID expression levels in control lines, whereas the AID<sup>-/-</sup> patient displayed no detectable AID transcripts. Concerning the CSR-ID patients, AID expression was similar or significantly increased when compared to controls (Figure 20A), confirming that the CSR defect observed in patients' lines is AID-independent. Thus, we verified the expression levels of genes deregulated in patients when compared to AID<sup>-/-</sup> (Figures 20B-E). According to the transcriptome data, *PSMA4* and *MRPS6* were highly downregulated in patients; although, by RT-qPCR we detected a significant enrichment of *PSMA4* and *MRPS6* transcripts in the patient cell lines when compared to AID<sup>-/-</sup> (Figure 20B, blue and black bars respectively) and we were not able to confirm the DGE profile. On the other hand, the overexpression of *IL-1 $\alpha$* , *P2RX1* and *SPARC* shown by high-throughput analysis was confirmed by the significantly higher amount of transcripts detected by RT-qPCR in patients' lines when compared to the AID<sup>-/-</sup> one (Figures 20C-E). Moreover, we verified the expression of those genes deregulated in the controls when compared to AID<sup>-/-</sup> sample (Figures 20C-D and Figure 20F). We observed that *IL-1 $\alpha$*  and *P2RX1* were upregulated in controls (Figures 20C and 20D), consistent with the transcriptome analysis. When we quantified the relative expression of *TESC* and *PRF1*, two genes downregulated in controls when compared to AID<sup>-/-</sup>, we observed a similar profile: *PRF1* was downregulated in controls (Figure 20F, black bars) and *TESC* was barely detectable in Ctr #2 and Ctr #3 (Figure 20F, blue bars). We observed instead a higher expression of *TESC* in Ctr #1 when compared to AID<sup>-/-</sup>, and thus an opposite profile than expected (Figure 20F, blue bars). However, as mentioned above, the Ctr #1 cell line was not included in the DGE analysis and our results suggest that the difference observed for *TESC* expression levels might be due to a difference between the samples and not to a false negative identification of these transcripts upon transcriptome profiling. We conclude that most of the genes whose expression we assessed so far followed the trend displayed by the transcriptome analysis, despite we observed some incongruences that have to be taken into account in the further steps of data analysis.

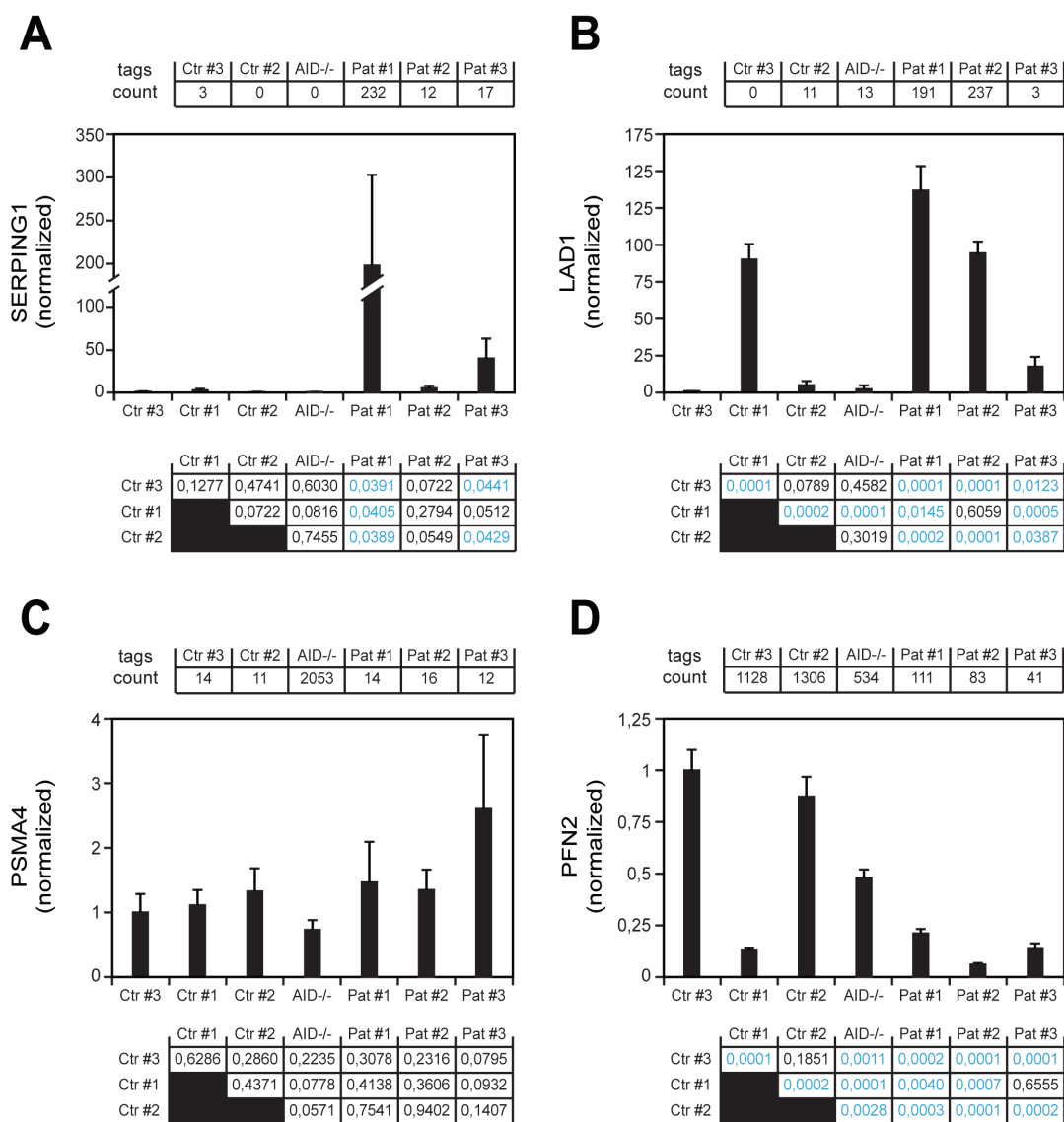
We then verified the expression profile obtained when comparing patients to controls, by assessing *SERPING1* and *LAD1* upregulation and *PSMA4* and *PFN2* downregulation by RT-qPCR (Figure 21). We observed a higher expression of *SERPING1* in the patient-derived cell lines relative to controls (Figure 21A) as well as for *LAD1*, relatively to the Ctr #3 (Figure 21B), whereas the Ctr #1 displayed a higher relative expression of *LAD1* when compared to the other two lines (Ctr #2 and #3, Figure 21B). Although we were not able to validate the downregulation of *PSMA4*, as the relative expression in

patients was enriched relative to the controls (Figure 21C), we did validate the profile detected for *PFN2*: we observed a significant reduction of *PFN2* expression in patients when compared to Ctr #3 and Ctr #2 (Figure 21D). Once again, Ctr #1 cell line displayed an opposite behavior compared to the other two, which were analyzed by DGE (Figure 21D). Thus, we conclude that most of the genes chosen for the validation of the DGE analysis display a similar expression profile when assessed by RT-qPCR and that, by taking into account the differences observed, we indeed would be able to further process our data.



**Figure 20. DGE data validation by RT-qPCR: genes upregulated and downregulated in patients and controls when compared to AID<sup>-/-</sup>**

(A) RT-qPCR for AID transcripts from EBV-immortalized human B cell lines is shown. Expression is normalized to GAPDH and is presented relative to the Ctrl #3, set as 1. Statistical significance vs. Ctrl #3, Ctrl #1 and Ctrl #2 (two-tailed Student's t-test) is indicated in the table below;  $p \leq 0.05$  is shown in blue. RT-qPCR for (B) PSMA and MRPS6, (C) IL-1 $\alpha$ , (D) P2RX1, (E) SPARC, (F) TESC and PRF1 transcripts is shown. Expression is normalized to GAPDH and is presented relative to AID<sup>-/-</sup>, set as 1. Statistical significance vs. AID<sup>-/-</sup> (two-tailed Student's t-test) is indicated: \*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ . The number of tags identified by DGE is indicated in the table above the histograms.



**Figure 21. DGE data validation by RT-qPCR: genes upregulated and downregulated in patients when compared to controls**

RT-qPCR for (A) SERPING1, (B) LAD1, (C) PSMA4 and (D) PFN2 transcripts is shown. Expression is normalized to GAPDH and is presented relative to Ctr #3, set as 1. Statistical significance vs. Ctr #3, Ctr #1 and Ctr #2 (two-tailed Student's t-test) is indicated in the table below;  $p \leq 0.05$  is shown in blue. The number of tags identified by DGE is indicated in the table above the histograms.

We analyzed the deregulated genes identified by DGE through pathway analysis tools, such as Ingenuity. By focusing on direct protein-protein interaction, we identified up to 24 networks for each of the conditions analyzed (patients vs. AID<sup>-/-</sup>; controls vs. AID<sup>-/-</sup> and patients vs. controls), whose top ten are shown in Table 8. This analysis did not allow the identification of obvious candidates, which could be responsible of the CSR defect observed in patients, thus we decided to cross the data we obtained by DGE in order to identify deregulated genes in CSR-ID patients which were AID-dependent.

#### Networks identified from deregulated genes in patients compared to AID<sup>-/-</sup>

Molecules	Top functions
88	Cellular Movement, Inflammatory Response, Immune Cell Trafficking
60	Dermatological Diseases and Conditions, Genetic Disorder, Lipid Metabolism
54	Cellular Development, Post-Translational Modification, Nervous System Development and Function
49	Psychological Disorders, Gastrointestinal Disease, Neurological Disease
51	Hematological Disease, Hematological System Development and Function, Organismal Functions
1	Carbohydrate Metabolism, Small Molecule Biochemistry, Antigen Presentation
1	Drug Metabolism, Protein Synthesis, Cell Death
1	Endocrine System Disorders, Gastrointestinal Disease, Genetic Disorder
1	Carbohydrate Metabolism
1	Infectious Disease, Carbohydrate Metabolism, Respiratory Disease

#### Networks identified from deregulated genes in controls compared to AID<sup>-/-</sup>

Molecules	Top functions
69	Cellular Movement, Cancer, Inflammatory Disease
50	Cancer, Gene Expression, Genetic Disorder
1	Carbohydrate Metabolism, Small Molecule Biochemistry, Antigen Presentation
1	Cellular Assembly and Organization, Nervous System Development and Function, Psychological Disorders
1	Cancer, Reproductive System Disease, Gene Expression
1	Carbohydrate Metabolism, Lipid Metabolism, Molecular Transport
1	Cellular Compromise, Cell Morphology, Cell-To-Cell Signaling and Interaction

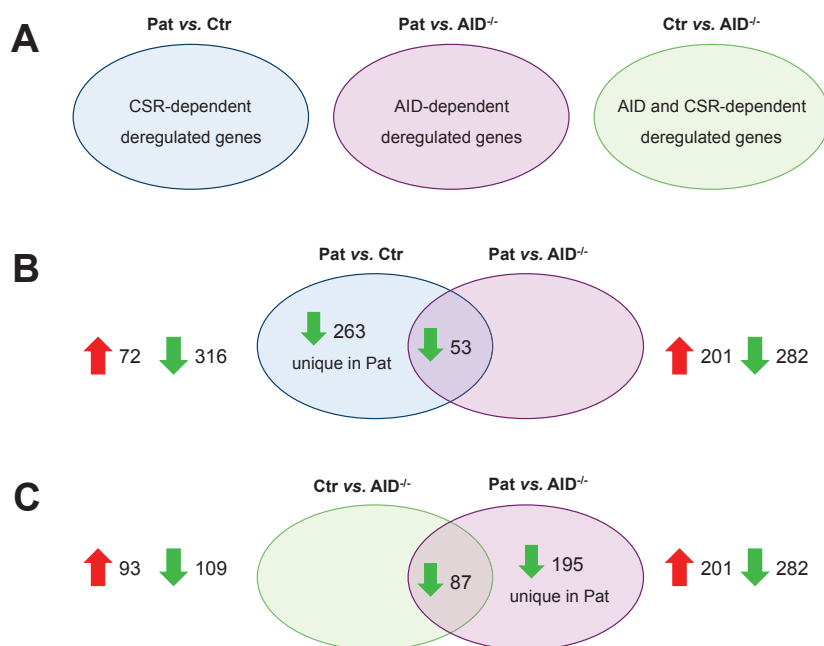
#### Networks identified from deregulated genes in patients compared to controls

Molecules	Top functions
80	Cellular Movement, Cellular Growth and Proliferation, Inflammatory Disease
56	Cellular Movement, Inflammatory Disease, Tissue Development
52	Skeletal and Muscular Disorders, Cancer, Genetic Disorder
47	Cell Morphology, Protein Synthesis, Cell Death
1	Cellular Movement, Embryonic Development, Cell Morphology
1	Cell-To-Cell Signaling and Interaction, Cellular Development, Nervous System Development and Function
1	Embryonic Development, Organ Development, Organismal Development
1	Cellular Movement, Tumor Morphology, Cancer
1	Cell-To-Cell Signaling and Interaction, Nervous System Development and Function, Organismal
1	Injury and Abnormalities
1	Cellular Assembly and Organization, Molecular Transport, Protein Trafficking

#### Table 8. List of top ten Ingenuity networks

List of top ten networks identified by the use of Ingenuity Pathway Analysis software while analyzing the genes deregulated in all the three conditions analyzed (patients vs. AID<sup>-/-</sup>; controls vs. AID<sup>-/-</sup> and patients vs. controls). The number of molecules belonging to the mentioned networks is indicated on the left. Networks indicated take into account direct protein-protein interaction and depend on the data set uploaded.

We pursued our analysis by comparing each condition to the other two: comparison between patients and controls provides the list of deregulated genes dependent on CSR, because controls are CSR proficient, and AID-independent because both patients and controls express AID; whereas the comparison between patients and AID<sup>-/-</sup> allows the identification of those genes whose deregulation depends on AID. On the other hand, the analysis of controls vs. AID<sup>-/-</sup> indicates those genes which are CSR- and AID-dependent (Figure 22A). By crossing the gene lists of differentially expressed genes in patients when compared to controls or AID<sup>-/-</sup>, we identified 263 genes downregulated exclusively in the patients which are CSR-dependent (Figure 22B). Whereas, by crossing the data obtained by comparing AID<sup>-/-</sup> to controls and patients, 195 genes appeared downregulated and were most likely related to AID expression in the patients (Figure 22C). These lists were submitted to our collaborators, who sequenced those genes corresponding to the major hits of our transcriptome profiling; however, no mutations were found in the candidates we proposed. So we conclude that, although our attempts to univocally identify the missing factor(s) in CSR-ID patients, this approach as well as the variability between human samples harbored some limitations that we decided to overcome by focusing on the B cells proteome.

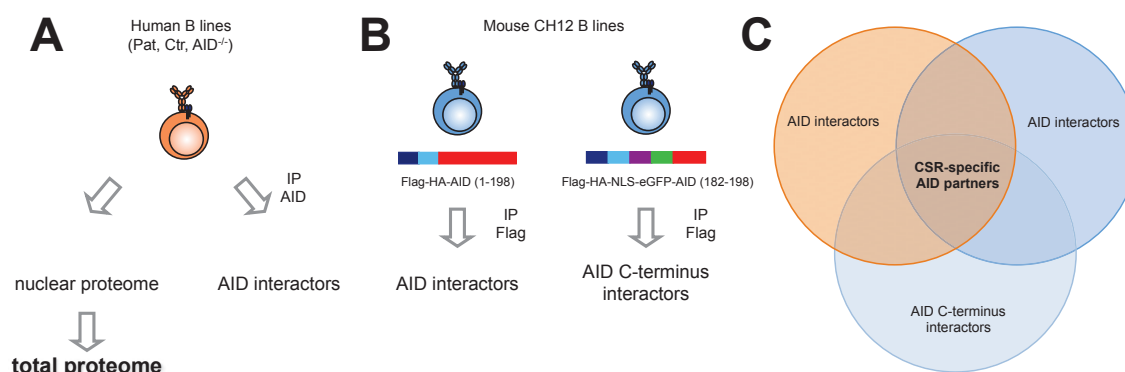


**Figure 22. Comparative analysis of deregulated genes in CSR-ID patients**

(A) Representation of the three data sets identified by DGE and their dependence on the conditions analyzed. (B) Venn diagram showing the comparison between deregulated genes identified when comparing patients vs. controls and patients vs. AID<sup>-/-</sup>. (C) Venn diagram showing the comparison between deregulated genes identified when comparing controls vs. AID<sup>-/-</sup> and patients vs. AID<sup>-/-</sup>. Red and green arrows represent upregulated and downregulated genes, respectively; the number of deregulated genes in each condition is indicated beside the diagrams.

## 2. Comparative proteomic analysis between human and mouse B cell lines to identify CSR-specific AID cofactors

The transcriptome profiling we undertook was based on the hypothesis that the mutation(s) in an annotated gene would affect the mRNA expression, allowing us to detect significant differences when comparing CSR-ID B cells with controls. However, we had to take into account that if the mutation did not affect the transcriptional expression of the gene we would not be able to detect it with this approach, leading to the hypothesis that the impaired targeting of AID to the switch regions could be dependent on a downstream defect, such as lower stability of the transcript, lower rate of translation or enhanced degradation. To identify proteins under-represented or missing in CSR-ID patients, we undertook a proteome screening on B cells isolated from CSR-ID patients, healthy donors and AID<sup>-/-</sup> patient. We extracted nuclear proteins and we analyzed the proteome composition by Multidimensional Protein Identification Technology (MudPIT) mass spectrometry (Fournier et al., 2007). Furthermore, we complemented our analysis on human B lymphocytes with the proteome identified in the mouse CH12 B cell line overexpressing either full length tagged-AID (Flag-HA-AID<sup>(1-198)</sup>) or the C-terminal domain of the protein (Flag-HA-NLS-eGFP-AID<sup>(182-198)</sup>), as depicted in Figure 23. AID C-terminus has been proposed to interact with factors specifically involved in CSR, as mutations located in this region lead to defective CSR despite normal levels of SHM (Barreto et al., 2003; Durandy et al., 2007; Geisberger et al., 2009; Imai et al., 2005; McBride et al., 2004; Ta et al., 2003). Therefore, by using CH12 B cells expressing this domain, we should be able to focus on CSR-specific AID partners. By crossing the different lists, in addition to factors previously described to play a role in CSR such as eEF1A, YY1, RPA, 14-3-3, CTNNBL1 and Spt5 (see Table S1, chapter VII), we identified the elongation factor Spt6.



**Figure 23. Strategy for proteome identification of CSR-specific AID partners on human and mouse B cell lines**

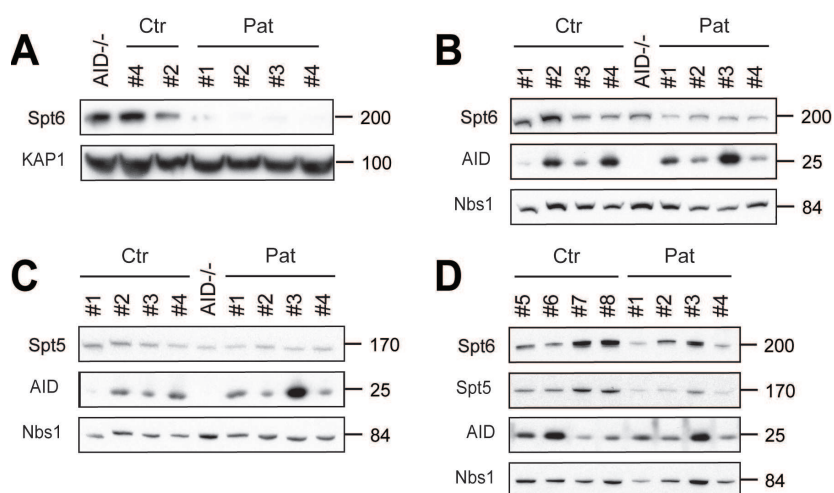
Scheme displaying the workflow of proteome analysis performed on human and mouse B cell lines. **(A)** Total nuclear proteins have been isolated from EBV-immortalized B cell lines obtained from CSR-ID patients (Patients), healthy donors (Ctrls) and AID<sup>-/-</sup> patient (AID<sup>-/-</sup>), upon – or not – IP of AID and followed by MS analysis. The data obtained have been crossed with those resulting from MS identification in **(B)** CH12 B cells overexpressing tagged AID (Flag-HA-AID<sup>(1-198)</sup>) or the last 17 C-terminal residues of the protein, in a construct harboring a nuclear localization signal (NLS) for proper nuclear translocation and an eGFP reporter gene (Flag-HA-NLS-eGFP-AID<sup>(182-198)</sup>) upon Flag IP. **(C)** Venn diagram displaying how the integration of data obtained from human and mouse B cell lines would allow the identification of those factors present in a complex with AID and specifically involved in CSR regulation.



### III. Spt6: the “missing factor” in CSR-ID patients?

One of the hits we identified in the proteome of human B cell lines was suppressor of Ty 6 homolog (Spt6), a histone chaperone and transcription elongation factor initially described in *S. cerevisiae* (Eitoku et al., 2008; Winston et al., 1984). We detected 5 and 6 peptides in the nuclear proteome of two healthy donor-derived cell lines and a single Spt6 peptide in one out of three healthy donor-derived B cell lines upon AID immunoprecipitation, while the protein was completely absent in the AID<sup>-/-</sup> B line and in the CSR-ID patient-derived lines. Spt6 has been described to co-localize with Spt5 and RNA polymerase II during transcription elongation (Kaplan et al., 2000) and to interact with the exosome complex (Andrulis et al., 2002). As both of these factors are implicated in CSR (Basu et al., 2011; Pavri et al., 2010), we decided to address the role of Spt6 in antibody diversification.

In order to confirm the results obtained with the MS analysis, we prepared nuclear extracts followed by western blots on controls and CSR-ID-derived B cell lines to evaluate Spt6 and AID expression (Figure 24).



**Figure 24. Spt6 and Spt5 expression in human B cell lines**

Nuclear extracts prepared from EBV-immortalized B cell lines obtained from healthy donors (Ctr), AID<sup>-/-</sup> patient (AID<sup>-/-</sup>) and CSR-ID patients (Pat). Extracts were blotted with antibodies specific for Spt6 (A, B and D), Spt5 (C and D) and AID (B-D); KAP1 (A) and Nbs1 (B-D) were used as loading controls.

On the first experiment we performed, we detected reduced levels of Spt6 in the patients compared to the positive controls and to the negative one (AID<sup>-/-</sup>), relative to KAP1 expression (Figure 24A). However, Spt6 reduction was less obvious in the following experiments we performed, as we detected a variable expression of the protein in healthy donors and AID<sup>-/-</sup>-derived extracts when compared to the loading control Nbs1 (Figures 24B and 24D). AID was expressed at different levels in healthy donor and patient-derived cell lines (Figures 24B-D) whereas it was undetectable in AID<sup>-/-</sup> B line (Figures 24B and 24C). Our data suggest that Spt6 expression is reduced in CSR-ID patients when

compared to healthy donor controls, thus we hypothesized that this partial decrease in Spt6 could be responsible for the AID-independent CSR defect in the CSR-ID patients.

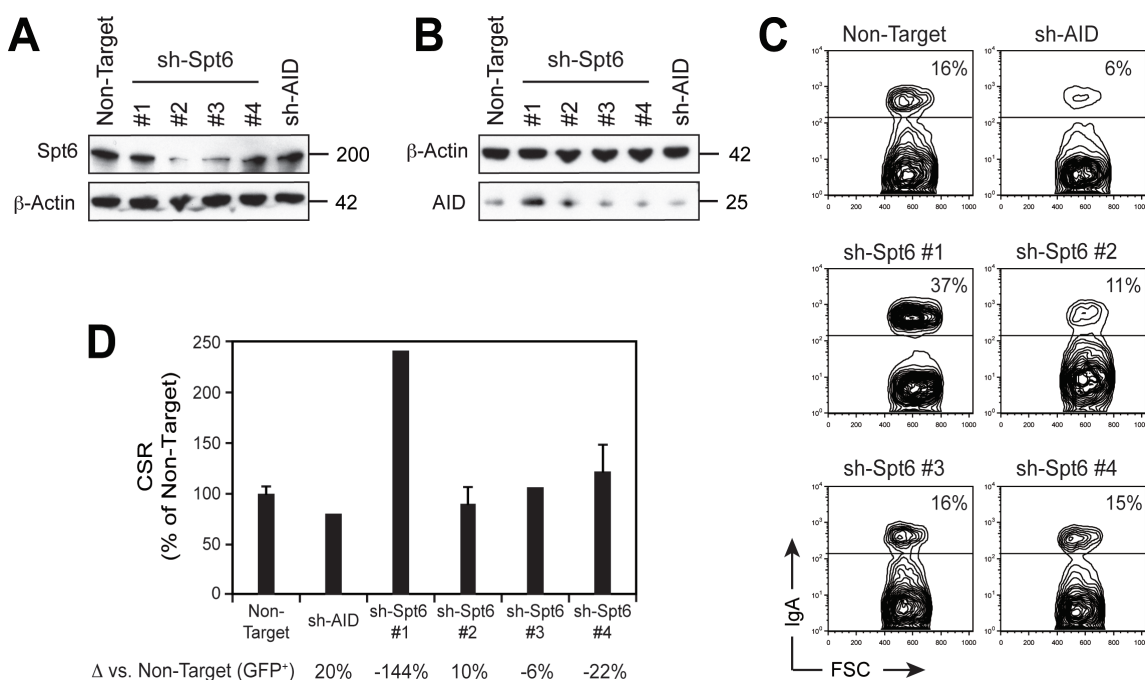
Additionally, taking into account the described interaction between Spt6 and Spt5 (Kaplan et al., 2000) and the role played by the latter in CSR (Pavri et al., 2010), we also assessed Spt5 expression in CSR-ID and control lines (Figure 24C and 24D). We observed no obvious difference in the patient cell lines when compared to controls (Figure 24C and 24D) and we concluded that Spt5 expression is not affected in CSR-ID patients, ruling out the possibility that Spt5 deficiency could contribute to the switching impairment detected in the patients. Therefore, our results suggest that Spt6 abundance does not seem to be related to Spt5.

According to the Spt6 expression levels detected in CSR-ID patients and to the hypothesis of a “missing factor” required for CSR, if Spt6 would have been the factor we were looking at, we would have expected to find it in a complex with AID and with Spt5, whose interaction with AID was previously described (Pavri et al., 2010). So we used the CH12 B line overexpressing tagged AID (Flag-HA-AID<sup>(1-198)</sup>) to perform immunoprecipitation experiments with an antibody anti-Flag followed by western blot analysis to identify Spt6 and Spt5 (see Figure 2B, chapter IV). We observed that AID, Spt5 and Spt6 are part of the same complex, as will be discussed in the next section of the results.

The identification of Spt6 and its presence in a complex with AID and Spt5 made us further investigate its role in class switch recombination (Figure 25). We took advantage of the mouse CH12 B lymphoma cell line that can be efficiently induced to undergo class switch recombination to the IgA isotype by stimulating cells with IL-4, anti-CD40 and TGF $\beta$  (Nakamura et al., 1996). We transduced cells with retroviral vectors expressing shRNAs targeting mouse *Spt6* sequence, AID as positive control or a non-target shRNA as negative control and we assessed Spt6 knockdown efficiency by western blot (Figure 25A). While cells transduced with Spt6 shRNA #1 and #4 as well as AID shRNA did not show any significant difference in Spt6 expression when compared to the non-target control, we detected reduced Spt6 levels upon silencing with shRNA #2 and #3 (Figure 25A) suggesting that in these two cell lines the gene silencing was efficient. Thus, we assessed the effect of Spt6 silencing on CSR by stimulating the transduced cells with IL-4, anti-CD40 and TGF $\beta$  for 72h and by measuring the percentage of cells expressing surface IgA (Figure 25C and 25D). We calculated the difference in CSR efficiency relative to the non-target transduced line ( $\Delta$ ) by subtracting from the non-target, set as 100%, the CSR value obtained for the Spt6 and AID shRNAs. AID knockdown resulted in 20% reduction CSR (Figure 25C and 25D), consistent with the partial reduction in AID expression we observed when compared to the non-target (Figure 25B). On the other hand, in those cell transduced with Spt6 shRNAs we detected a little reduction in CSR (10% for sh-Spt6 #2, Figure 25C and 25D) or a more efficient switching (-144%, -6% and -22% for sh-Spt6 #1, #2 and #3 respectively, Figure 25C and 25D).

As there is a direct correlation between the efficiency of recombination at the IgH locus and the amount of AID present within the cell, even little fluctuations in AID expression level might have an impact on switching. Thus we verified the expression of the deaminase in Spt6 knockdown cell lines (Figure 25B), to rule out that the effect we observed was dependent on AID. We found that AID was expressed at variable levels compared to cells transduced with the non-target shRNA (Figure 25B).

Interestingly, the cell lines which displayed efficient Spt6 silencing expressed either higher level (sh-Spt6 #2) or comparable level (sh-Spt6 #3) of AID to the non-target line, suggesting that the little reduction in CSR observed upon Spt6 knockdown with the shRNA #2 is not due to a lower AID expression (Figures 25B and 25D). Based on these experiments, we observed that under our experimental conditions the knockdown was not as striking as expected, so we modified the conditions according to the results published by Rushad Pavri and collaborators on Spt5 (Pavri et al., 2010), who successfully used lentiviruses expressing the shRNA with a GFP reporter gene.

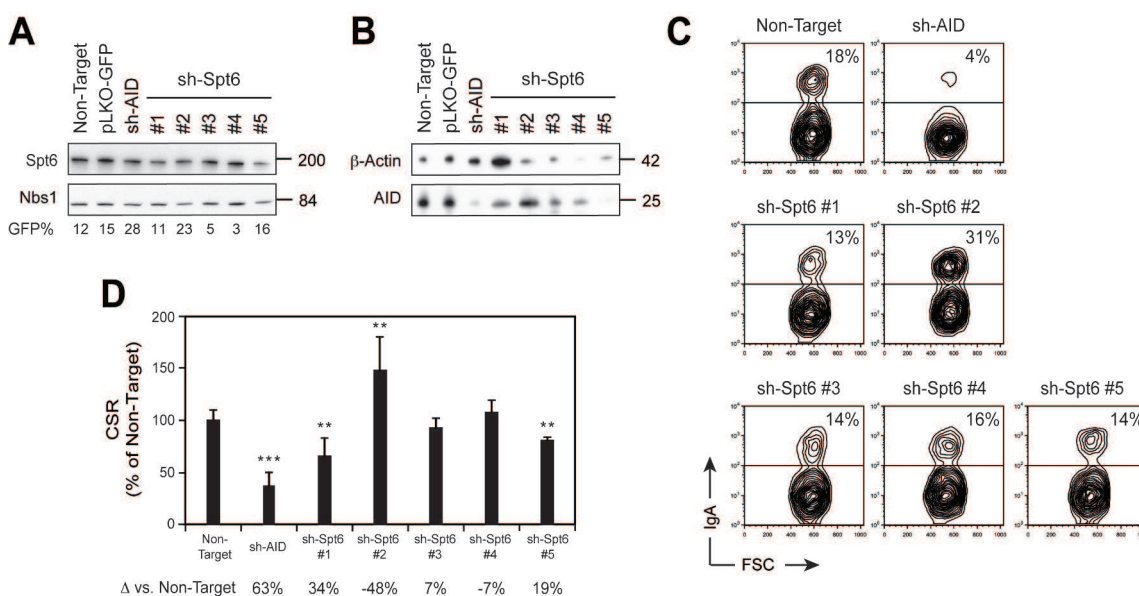


**Figure 25. Retrovirus-mediated Spt6 knockdown on CH12 B cells**

(A) Western blot for Spt6 and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced with a retrovirus expressing a GFP reporter and shRNA specific for Spt6, AID or a non-target shRNA control. (B) Western blot for AID and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced with retroviruses described in (A) and stimulated with IL-4, anti-CD40 and TGF $\beta$  for 72h. (C) IgA surface expression as determined by flow cytometry in stimulated CH12 cells transduced with retroviruses described in (A). Representative plots (gated on GFP<sup>+</sup>) from two independent experiments are shown. Dead cells have been excluded from the analysis by ToPro3 staining. (D) Percentage (+ s.d.) of CSR relative to the non-target shRNA control from two independent experiments by gating on cells expressing GFP. CSR in cells expressing the non-target shRNA control was set as 100%. The difference in CSR efficiency relative to the non-target control ( $\Delta$ ) is indicated below.

We thus tested five different shRNAs targeting Spt6, one targeting AID and a non-target shRNA control (Figure 26). We assessed knockdown efficiency by western blot in CH12 cells transduced with shRNAs targeting Spt6, AID, with the vector lacking the shRNA insert but harboring only the GFP reporter (pLKO-GFP) and with a non-target shRNA control prior to stimulation (Figure 26A). We observed slight reduction in Spt6 expression in the shRNA #5 when compared to the non-target and to the empty vector-transduced line and relative to the Nbs1 control, whereas the other shRNAs did not seem to reduce Spt6 levels (Figure 26A). AID silencing, instead, resulted very efficient as we

observed a significant reduction of the deaminase protein levels in cells transduced with the shRNA targeting AID when compared to the non-target (Figure 26B). Additionally, in order to test if Spt6 knockdown affects AID expression, we verified its expression levels in transduced lines stimulated for 72h and we observed a variable expression of AID in Spt6 knockdown lines when compared to the non-target line (Figure 26B). Although the protein loading relative to cells expressing the Spt6 shRNA #5 was not equal to the positive and negative control (non-target and sh-AID), we still detected lower AID expression (Figure 26B). From this experimental system, we were expecting to obtain high transduction efficiency, but we observed quite low percentage of GFP positive cells (Figure 26A). Although the shRNA and *GFP* gene transcription are dependent on different promoters – which could lead to transcription at different rates – we assumed the reporter expression as clear indication of the number of transduced cells which were actively expressing the indicated shRNA. However, in these experiments, we observed that the knockdown was difficult to assess by correlation with the GFP expression in transduced cells (Figure 26A).



**Figure 26. Lentivirus-mediated *Spt6* knockdown on CH12 B cells might have an effect on CSR**

(A) Western blot for *Spt6* and *Nbs1* as loading control on total extracts obtained from CH12 cells transduced with a lentivirus expressing a GFP reporter alone (pLKO-GFP) or a GFP reporter and shRNA specific for *Spt6*, AID or a non-target shRNA negative control. (B) Western blot for AID and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced with lentiviruses described in (A) and stimulated with IL-4, anti-CD40 and TGF $\beta$  for 72h. (C) IgA surface expression as determined by flow cytometry in stimulated CH12 cells transduced with lentiviruses described in (A). Representative plots (gated on GFP<sup>+</sup>) from two to four independent experiments are shown. Dead cells have been excluded from the analysis by ToPro3 staining. (D) Percentage (+ s.d.) of CSR relative to the non-target shRNA control from two to four independent experiments by gating on cells expressing GFP. CSR in cells expressing the non-target shRNA control was set as 100%. The difference in CSR efficiency relative to the non-target control ( $\Delta$ ) is indicated below. Statistical significance vs. the non-target control (two-tailed Student's t-test) is indicated: \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ .

As the *in vitro* functional assay can allow us to assess switching in cells expressing GFP, and presumably the shRNA, we decided to carry on CSR analysis. We stimulated cells with IL-4, TGF $\beta$

---

and anti-CD40 in order to assess recombination efficiency by measuring IgA surface expression by flow cytometry (Figures 26B and 26C). As expected, AID knockdown significantly impaired CSR in transduced and stimulated cells, leading to a 63% difference when compared to the non-target control (Figures 26B and 26C). Spt6 knockdown, instead, induced opposite effects on the ability of transduced cells to undergo CSR: we observed a significant reduction (19% and 34% for sh-Spt6 #5 and #1, respectively; Figures 26B and 26C) for two shRNAs tested as well as a more efficient switching, in cells transduced with the shRNA #2 (-48%; Figures 26B and 26C). Taking into account that a partial depletion of Spt6 was detected mainly for one out of five shRNAs used for the assay, we could not exclude that the results we obtained were reflecting the decrease in AID expression levels more than a potential Spt6 effect in stimulated cells (Figures 26A and 26C). Our results suggest that Spt6 may play a role in CSR, although cell sorting for GFP expression would help to better characterize knockdown and recombination efficiency upon Spt6 silencing.

While we were characterizing the role of Spt6 in switching regulation, a publication from the laboratory of Tasuku Honjo (Okazaki et al., 2011) described a CSR-specific role for Spt6, thus delineating the function we were trying to address. Nevertheless, as already mentioned, some results we obtained while we were investigating Spt6 have contributed to clarify the role of the PAF complex in antibody diversification, and the results obtained within this study will be discussed in the next section.

At this point of time, we focused further on the proteome screening in order to identify additional candidates whose activity could be pivotal in switching regulation.

## IV. Role of the PAF complex in CSR

When we identified Spt6 in our proteome screening, we undertook its functional characterization in CSR; however, the publication from the group of Tasuku Honjo (Okazaki et al., 2011) as well as the sequencing of CSR-ID patients and the fact that no mutations were found in *SUPT6H* gene obliged us to focus on other candidates, such as the Smc complexes. Nevertheless, some results we obtained while we were investigating the role of Spt6 in switching have supported a work performed in collaboration with the laboratory of Dr. Svend Petersen-Mahrt, focused on the RNA polymerase II-associated factor (PAF) complex (Willmann et al., 2012).

The PAF complex has been identified in the Petersen-Mahrt laboratory as co-immunoprecipitating with AID in the chicken DT40 B cell line, able to undergo gene conversion *in vitro*. By knocking-in the endogenous AID sequence with a tagged version of the gene, has been possible to identify those proteins that physiologically interact with endogenous AID. Interestingly, in addition to the PAF complex component (Paf1, Leo1 and Ctr9), we identified Spt5 and Spt6 as well as the RNA polymerase II and the FACT complex.

As the CH12 B cell line represents an efficient and relatively easy system to assess the functional involvement of a candidate protein in switching, we were able to apply our expertise to this study. So we confirmed the interaction between PAF and AID, previously detected in DT40 cells, in mouse B cells, by performing immunoprecipitation experiments in CH12 cells overexpressing tagged AID (Flag-HA-AID<sup>(1-198)</sup>), and we isolated Paf1, Leo1, Ctr9, Spt5, Spt6 and the RNA polymerase II in the same complex than nuclear AID. By confirming the interaction between AID and Spt6, I thus contributed to validate the AID interactome identified in chicken B cells. Moreover, further investigations showed that this interaction was direct and mediated by Paf1, and that it might occur through the N-terminal domain of AID.

To address the role of the PAF complex in CSR, we silenced Paf1, Leo1, Ctr9, Cdc73 as well as AID and Spt5 (as positive controls) gene expression in CH12 B cells by using recombinant retro- and lentiviruses. We observed impairment in CSR efficiency upon PAF complex knockdown, as well as reduced AID expression in cells where Paf1, Ctr9 and Cdc73 were depleted. We also detected a reduced rate of transcription at the acceptor S $\alpha$  region upon Paf1 and Ctr9 retrovirus-mediated knockdown. Interestingly, Leo1 depletion did not affect these early steps of CSR, suggesting that it might play a direct role in recombination.

We thus verified whether the PAF complex was actively recruited to the Ig loci in DT40 cells and we observed that its occupancy at the rearranged Ig $\lambda$  locus was not dependent on AID. Furthermore, to rule out a potential cause/effect relationship between the PAF complex localization at the Ig loci and the subsequent recruitment of AID, we performed ChIP experiments with an anti-AID antibody in AID and Leo1 knockdown CH12 B cells before or after stimulation, to verify AID presence at the S $\mu$  donor region. We found that AID recruitment to S $\mu$  region was impaired upon Leo1 silencing, suggesting that Leo1, as part of the PAF complex, regulates AID presence at the Ig loci and AID-dependent antibody diversification.

# A role for the RNA pol II-associated PAF complex in AID-induced immune diversification

Katharina L. Willmann,<sup>1,2</sup> Sara Milosevic,<sup>4</sup> Siim Pauklin,<sup>1</sup>  
Kerstin-Maike Schmitz,<sup>2</sup> Gopinath Rangam,<sup>1,2</sup> Maria T. Simon,<sup>1</sup>  
Sarah Maslen,<sup>3</sup> Mark Skehel,<sup>3</sup> Isabelle Robert,<sup>4</sup> Vincent Heyer,<sup>4</sup>  
Ebe Schiavo,<sup>4</sup> Bernardo Reina-San-Martin,<sup>4</sup> Svend K. Petersen-Mahrt<sup>1,2</sup>

<sup>1</sup>DNA Editing Laboratory, London Research Institute, South Mimms EN6 3LD, England, UK

<sup>2</sup>DNA Editing in Immunity and Epigenetics, IFOM-Fondazione Istituto FIRC di Oncologia Molecolare, Via Adamello 16, 20139 Milano, Italy

<sup>3</sup>Protein Analysis and Proteomics Laboratory, London Research Institute, South Mimms EN6 3LD, England, UK

<sup>4</sup>Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de la Santé et de la Recherche Médicale (INSERM) U964, Centre National de la Recherche Scientifique (CNRS) UMR7104, Université de Strasbourg, 67404 Illkirch, France

**Antibody diversification requires the DNA deaminase AID to induce DNA instability at immunoglobulin (Ig) loci upon B cell stimulation. For efficient cytosine deamination, AID requires single-stranded DNA and needs to gain access to Ig loci, with RNA pol II transcription possibly providing both aspects. To understand these mechanisms, we isolated and characterized endogenous AID-containing protein complexes from the chromatin of diversifying B cells. The majority of proteins associated with AID belonged to RNA polymerase II elongation and chromatin modification complexes. Besides the two core polymerase subunits, members of the PAF complex, SUPT5H, SUPT6H, and FACT complex associated with AID. We show that AID associates with RNA polymerase-associated factor 1 (PAF1) through its N-terminal domain, that depletion of PAF complex members inhibits AID-induced immune diversification, and that the PAF complex can serve as a binding platform for AID on chromatin. A model is emerging of how RNA polymerase II elongation and pausing induce and resolve AID lesions.**

## CORRESPONDENCE

Svend K. Petersen-Mahrt:  
svend.petersen-mahrt@ifom.eu

Abbreviations used: AID, activation-induced deaminase; ChIP, chromatin IP; CSR, class switch recombination; CTD, C-terminal domain (of RNA pol II); iGC, Ig gene conversion; IP, immunoprecipitation; PAF, RNA polymerase-associated factor; qRT-PCR, quantitative RT-PCR; SEC, size exclusion chromatography; SHM, somatic hypermutation; TSS, transcription start site.

In B cells, antibody diversity is created via two DNA instability mechanisms (Rajewsky, 1996). In the first, RAG1/2 mediate antigen-independent V(D)J recombination, and in the second, activation-induced deaminase (AID) drives antigen-dependent Ig diversification. The latter includes somatic hypermutation (SHM), Ig gene conversion (iGC), and class switch recombination (CSR). SHM and iGC induce variable (V) region diversification via templated and nontemplated DNA mutations (Di Noia and Neuberger, 2007), whereas CSR recombines DNA constant (C) switch regions, resulting in IgM to IgG, IgA, or IgE isotype switching (Stavnezer et al., 2008). Mechanistically, SHM,

iGC, and CSR are initiated by the DNA deaminase AID, which deaminates cytosine (dC) residues to uracil (dU) on single-stranded DNA (ssDNA; Petersen-Mahrt, 2002, 2005; Bransteitter et al., 2003; Chaudhuri et al., 2003). At the genetic level, deamination causes a change in base recognition, as uracil is read as thymine during replication. At the biochemical level, reformation of double-stranded DNA (dsDNA) causes an alteration of DNA structure, resulting in a dU:dG lesion, which in turn activates DNA repair pathways resulting in mutated or otherwise altered chromosomes.

Because of the high oncogenic potential of AID, understanding how DNA deaminases are

K.L. Willmann, S. Milosevic, and S. Pauklin contributed equally to this paper.

S. Pauklin's present address is Laboratory For Regenerative Medicine, University of Cambridge, Cambridge CB2 0SZ, England, UK.

© 2012 Willmann et al. This article is distributed under the terms of an Attribution-Noncommercial-Share Alike-No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms>). After six months it is available under a Creative Commons License (Attribution-Noncommercial-Share Alike 3.0 Unported license, as described at <http://creativecommons.org/licenses/by-nc-sa/3.0/>).

regulated at the target site is one of the most important aspects in the field of DNA editing and Ig diversification; however, little is known about the protein complexes and mechanisms involved. Mechanistically, AID requires ssDNA as a substrate, and although several chromatin alteration events could lead to ssDNA formation, transcription at the Ig locus is required for SHM and CSR. The rate of transcription correlates with the rate of SHM (Peters and Storb, 1996), and germline transcription through the switch and the constant region precedes CSR (Stavnezer-Nordgren and Sirlin, 1986). Interaction of AID with CTNBL1 (Conticello et al., 2008; Ganesh et al., 2011) demonstrated an association with RNA processing. More recently, though, direct links between AID and mRNA transcription were demonstrated. It was shown that CSR required the basal transcription factor SUPT5H (Pavri et al., 2010) and its associated factor SUPT4H (Stanlie et al., 2012), the transcription-associated chromatin modifier FACT complex (Stanlie et al., 2010), and histone chaperon SUPT6H (Okazaki et al., 2011), whereas AID activity during CSR was enhanced by components of the RNA-processing exosome (Basu et al., 2011).

To delineate the biochemical link of RNA pol II transcription to AID-induced Ig diversification, and to further characterize the AID interactome, we developed a novel biochemical approach: we C-terminally tagged the endogenous AID protein in Ig diversifying cells with a FLAG or a FLAG/Myc epitope (Pauklin et al., 2009), and we adapted a recently developed method for isolation of chromatin-bound protein complexes (Aygün et al., 2008). This method allowed, for the first time, the identification and characterization of proteins that are associated with AID on chromatin in their physiological environment. The majority of the identified proteins (FACT complex, SUPT5H, SUPT6H, RNA polymerase-associated factor (PAF) complex, RPB1A, RPB1B, and DNA topo I) are involved in RNA processing, chromatin remodeling, exosome processing, and RNA pol II transcription elongation/pausing. We identified a direct interaction of AID (the N-terminal domain) with PAF1, and by using knockdown experiments, we could demonstrate physiological importance of the PAF complex for Ig class switching and recruitment of AID to the Ig locus. A model of how this complex could influence AID efficacy at the Ig locus will be discussed.

## RESULTS

To determine the composition of the protein complexes that interact with AID on chromatin in B cells undergoing Ig receptor diversification, we developed cell lines in which endogenous AID was tagged with epitope-peptides at the C terminus (Pauklin et al., 2009). In the chicken B cell lymphoma DT40, which continuously undergoes AID-dependent diversification of the Ig locus, AID was tagged with either 3xFLAG peptides (3F) or the combination of 3xFLAG peptide, 2xTEV cleavage sites, and 3xMyc peptides (3FM). This yielded expression of tagged AID to levels that were comparable to endogenous amounts. Although it is known that the C terminus of AID plays an important

role in subcellular localization, we could not detect a significant change in AID relocation or immune diversification activity caused by the monoallelic C-terminal tags (unpublished data).

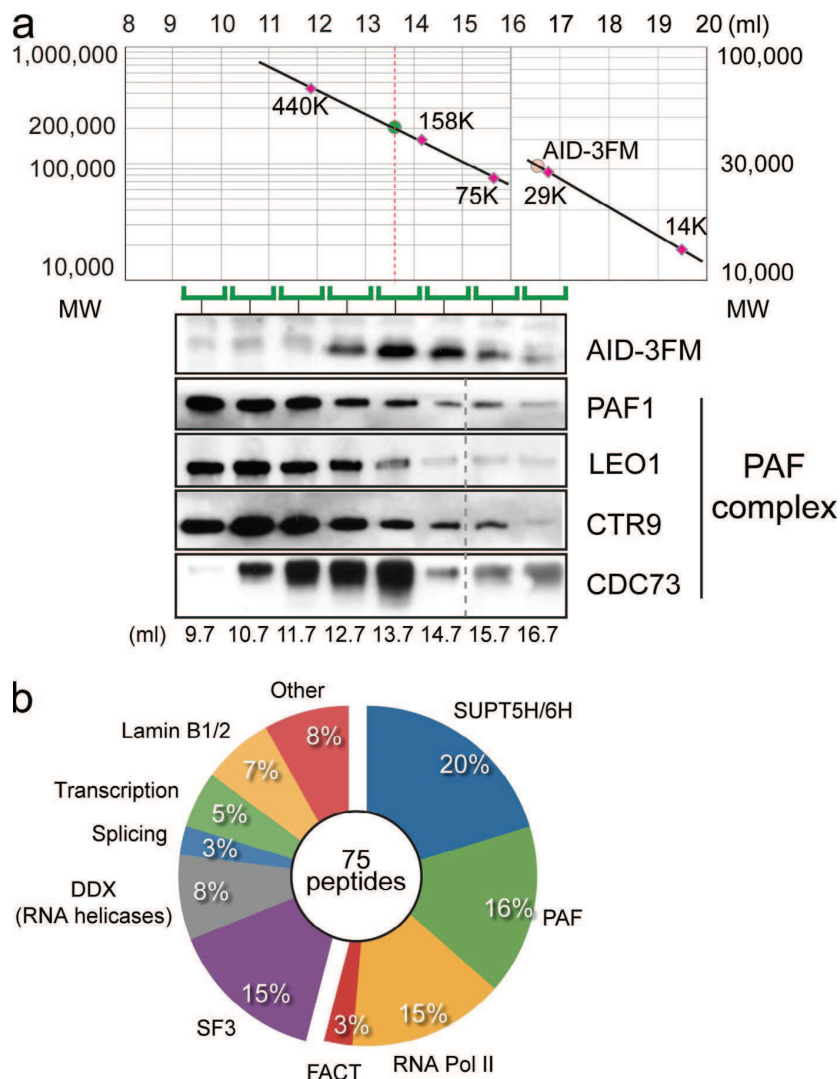
### Chromatin AID is part of a multimeric complex

Because AID is predominantly localized in the cytoplasm (Rada et al., 2002; Brar et al., 2004; Ito et al., 2004; McBride et al., 2004), and only limited amounts can be identified within the nucleus, we grew  $1-2 \times 10^{10}$  AID-3FM cells for biochemical analysis. Cell lysates were subfractionated into cytoplasm, nucleoplasm, and chromatin fractions. We focused on chromatin-bound AID, which we estimated to be <2% of total AID-3FM (unpublished data). The isolated chromatin fraction was further separated using a Superdex 200 column for size exclusion chromatography (SEC), thereby determining the size of the AID-associated protein complex bound to chromatin (Fig. 1 a). AID was identified as part of a 200-kD protein complex (120–300 kD based on standard proteins), whereas only a minor fraction of AID eluted at its theoretical monomeric size of 27 kD (Fig. 1 a). This demonstrated that AID isolated from chromatin under physiological conditions is part of a large heteromeric complex.

### The PAF complex associates with AID on chromatin in Ig diversifying cells

To identify proteins associated with chromatin-bound AID, we performed FLAG peptide immunoprecipitations (IPs), followed by one-dimension SDS-PAGE and mass spectrometry identification (Fig. S1). We obtained 1,319 peptide identities (Ids), corresponding to 391 proteins from AID-3FM cells. Mass spectrometric analysis of IPs from cytoplasmic, nucleoplasmic, and chromatin fractions of a control cell line (expressing AID without a tag) served as a control peptide Id database. Using this database, we eliminated 366 of the 391 proteins (>15-fold enrichment; all AID-interacting proteins are listed in Fig. S1). When we submitted the protein Ids into the Ingenuity Systems Pathway Analysis gene network software, we obtained a potential interacting network containing >80% of all isolated peptides (Fig. S2). The majority of the AID-associated proteins from the chromatin fraction were part of mRNA processing. Aside from the core RNA pol II subunits, we identified the core PAF complex (RNA polymerase-associated factor; PAF1, CTR9, LEO1), FACT complex (SSRP1, SUPT16H), SUPT5H, SUPT6H, and DNA topo I (Fig. 1 b). These factors play a direct role in RNA pol II pausing/restarting and elongation, as well as in chromatin modification and exosome processing. Furthermore, an additional 20 peptides comprised proteins involved in RNA metabolism (splicing-associated factors and RNA helicase). The high percentage (54%) of peptides that are part of the same biological process (early mRNA biogenesis), and which co-isolate with AID, indicated that our isolation and analysis procedure had identified key AID-interacting proteins at the chromatin level





**Figure 1. Chromatin-bound AID is in a multimeric complex.** (a) Elution profile of size exclusion chromatography. The chromatin fraction from DT40 AID-3FM cells was loaded onto a Superdex 200 column. (top) The elution profile of standard proteins is plotted in the graph. Red circle, theoretical elution volume of AID-3FM (29 kD). (bottom) Eluted fractions (elution volume indicated, milliliters) were analyzed by Western blot probing for the presence of AID-3FM, PAF1, LEO1, CTR9, and CDC73. (b) Analysis of filtered peptide hits. The most abundant peptides identified were grouped using gene network software analysis, and groups are displayed in a pie chart as percentage of total peptides (75) that could be assigned to each complex. A full table of chromatin peptides is listed in Fig. S1.

or parafibromin), as well as confirm SUPT5H and SUPT6H (Fig. 2 a). The AID-association with SUPT5H (Fig. 2 a), although technically difficult, was further confirmed by multiple large scale FLAG immunoprecipitation and mass spectrometry, in which SUPT5H association was identified in three out of five experiments (and SUPT6H and PAF1 were identified in each IP). In conclusion, our work has for the first time identified and verified AID-associated complexes on chromatin in diversifying B cells.

#### The PAF complex associates with AID in CSR-competent murine B cells

To determine whether the identified associations between AID and RNA pol II-associated factors observed in DT40 cells is also present in murine

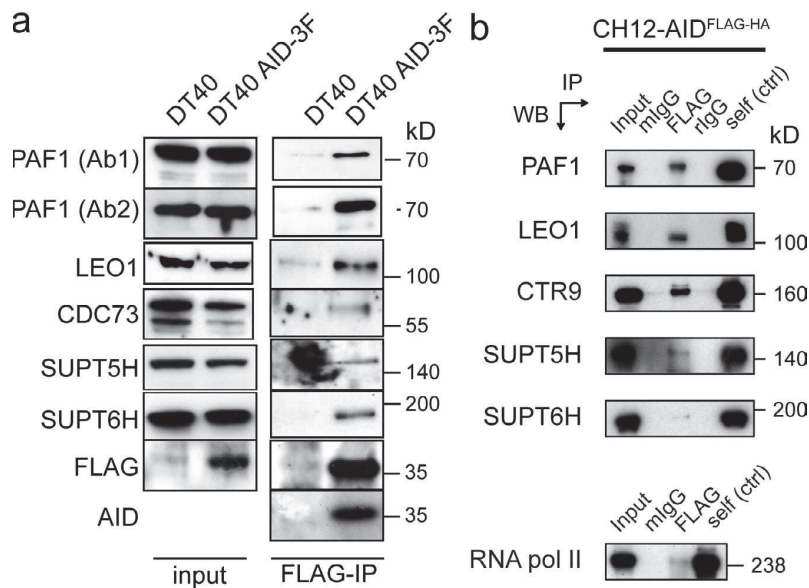
CSR-proficient cells, we performed a coIP experiment from nuclear extracts of CH12 B cells expressing tagged AID (AID<sup>Flag-HA</sup>; Jeevan-Raj et al., 2011). Consistent with the DT40 analysis (Fig. 2 a), PAF1, LEO1, CTR9, SUPT5H, and SUPT6H could be identified to associate with AID (Fig. 2 b). Moreover in a reciprocal experiment, in which PAF1, LEO1, CTR9, SUPT5H, SUPT6H, and RNA polymerase II were precipitated, we identified AID in all IPs performed (unpublished data). This indicated that the identified AID associations were present in both DT40 and CH12 cells, thereby establishing a potential biochemical link between V region diversification (DT40 cells) and CSR (CH12 cells).

#### AID associates with the PAF complex via PAF1

To further characterize the PAF complex association with AID, we used immunoblot analysis of the chromatin SEC

from immune-diversifying cells. Consistent with this, several of the proteins that we identified (RNA pol II, SUPT5H, SUPT6H, FACT complex, and DNA topo I) have been previously described to play a role in SHM and CSR. It is important to note that the chicken genome is not fully characterized and annotated, and thus the number of proteins we have identified may be underestimated.

Our mass spectrometry analysis of the AID chromatin interactome showed PAF1, CTR9, and LEO1 as AID-associated proteins on chromatin. They form part of the PAF complex, a RNA pol II-associated complex that promotes elongation (Kim et al., 2010) by recruiting enzymes for histone H2 monoubiquitination and other co-transcriptional chromatin marks (Jaehning, 2010). We could verify the associations of AID by analyzing the chromatin FLAG-IP for PAF1 (two different antibodies), LEO1, CDC73 (also known as HRPT2



**Figure 2. PAF complex and RNA pol II-associated proteins coIP with AID in Ig diversifying cells.** (a) Verification of DT40 AID-3F associations. DT40 chromatin fractions were immunoprecipitated via FLAG and eluted complexes were probed by Western blotting: anti-PAF1 (two antibodies), anti-LEO1, anti-CDC73, anti-SUPT5H, anti-SUPT6H, anti-FLAG, and anti-AID antibodies. Input lysates from both cell lines are shown on the left (input); parental DT40 cell line (DT40) served as negative control for FLAG-IPs from chromatin fractions. The anti-AID antiserum was unable to detect chicken AID at physiological levels (not shown). (b) Nuclear extracts from CH12 cells stably expressing AID<sup>FLAG-HA</sup> were immunoprecipitated via anti-FLAG and the following controls: murine IgGs (mlgG) and rabbit IgG (rlgG). The isolated complex was probed with the indicated antibodies. Input represents 1% of material used for IP. Self (ctrl) refers to an IP using the indicated antibody on the left. ND, not determined. Molecular weight in kilodaltons are indicated. Data are representative of three independent experiments.

fractions (Fig. 1 a), and demonstrated that PAF1, LEO1, and CTR9 co-migrate in a large (>400-kD) complex (Fig. 1 a, lanes 1–3), with the peak trailing fractions overlapping with the AID peak (Fig. 1 a, lane 5). Although AID did not fully co-migrate in the same peaks, the data indicated that the classical PAF complex was present in DT40 and partially associated with AID on chromatin. It was therefore likely that AID interacted with one of the components of the PAF complex rather than with each individual member.

We coexpressed AID with individual PAF members in *E. coli* and monitored binding by coIP and Western blot analysis (Fig. 3 a). This approach avoided possible eukaryotic bridging proteins being present in the assay and was likely to identify direct interaction. The cloned (human) cDNAs were FLAG tagged and coexpressed with untagged human AID from the same plasmid. FLAG-PAF1 was co-isolated in AID immunoprecipitates, whereas CDC73 (Fig. 3 a), SSRP1 (not depicted), and LEO1 (not depicted) did not show robust association. The PAF1–AID association was specific (Fig. 3 a, lanes 4–6) and did not occur in the absence of AID-specific antisera (Fig. 3 a, lanes 7–9). A reciprocal IP experiment was also performed (unpublished data), verifying the AID–PAF1 association. To confirm the possible direct interaction between AID and PAF1, we performed classical pull-down analysis with recombinant AID and in vitro-produced PAF1. As shown in Fig. 3 b, PAF1 associated with AID but not APOBEC2, a member of the AID/APOBEC deaminase family. We also attempted to identify AID and SUPT5H association in the *Escherichia coli* and in vitro translation assays, but unlike the robust PAF1 association, were unable to demonstrate significant co-isolations (unpublished data).

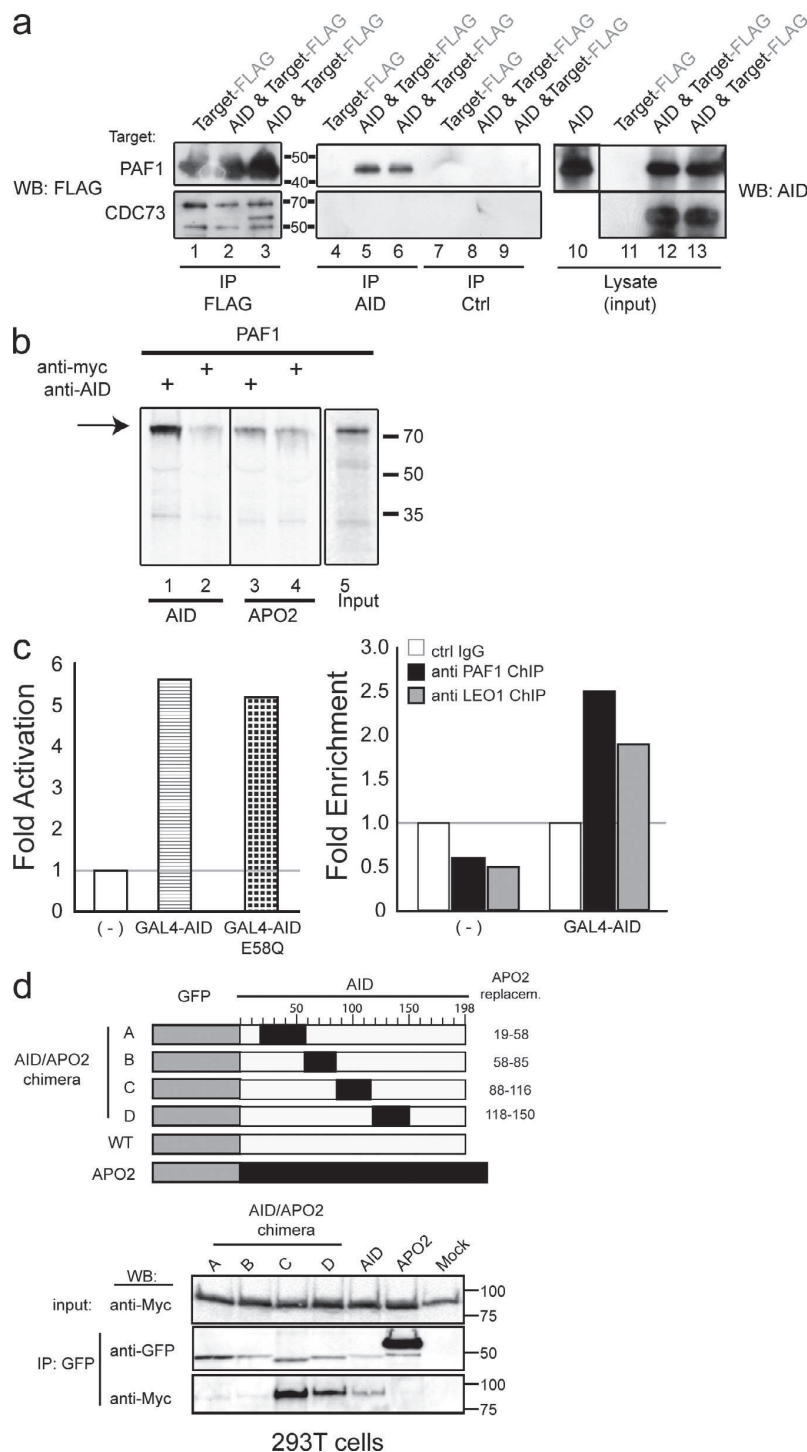
To demonstrate that the AID–PAF1 association can provide a functional consequence in mammalian cells, we used a transcription reporter assay. PathDetect HeLa luciferase

reporter (HLR) cells harbor a luciferase transgene in their genome that can be activated by the PKA-phosphorylated CREB transcription factor. The presence of GAL4-binding sites (UAS) within the promoter allows for monitoring the effect of GAL4-fusion proteins on transcription. When GAL4 fusions of AID or AID mutant (E58Q) protein were transiently transfected, luciferase activity was enhanced nearly sixfold (Fig. 3 c, left). PAF1 and LEO1 chromatin IP (ChIP) of the transfected cells demonstrated that endogenous PAF1 and LEO1 were recruited to the locus upon AID expression. (Fig. 3 c, right), further underlying a more direct association between AID and the PAF complex.

Mapping the domain of AID that fostered this association was demonstrated by the use of AID-APOBEC2 chimeras, which substitute corresponding APOBEC2 peptide regions in place of AID peptide regions (Conticello et al., 2008). GFP-tagged AID, APOBEC2, or AID/APOBEC2 chimera proteins were coexpressed with Myc-peptide tagged human PAF1 in HEK293T cells and subjected to coIP. While IPs of AID and chimeras C and D showed co-purification of PAF1, APOBEC2 and chimera A and B failed to isolate PAF1 (Fig. 3 d), suggesting that the N-terminus of AID is responsible for the PAF1 association.

#### The PAF complex is required for functional CSR

Our finding that RNA pol II elongation factors associate with AID on chromatin, along with the previously established link of transcription being essential for SHM and CSR, provides an insight into the mechanism of AID activity at Ig loci. To determine the biological relevance of the PAF complex in CSR, we undertook knockdown experiments in murine B cells. CH12 cells were transduced with retrovirus-expressing shRNAs specific for the different subunits of the PAF complex. Transduced cells were stimulated in vitro, and their capacity to



**Figure 3. AID-PAF complex association via PAF1.** (a) FLAG-PAF1 or FLAG-CDC73 were expressed alone (lanes 1, 4, 7, and 11) or from the same plasmid as AID (lanes 2, 3, 5, 6, 8, 9, 12, and 13) in *E. coli*. PAF1 is processed as a shortened (50 kD) fragment in bacteria. *E. coli* lysates were used for IP with anti-FLAG (lanes 1–3), anti-AID antibodies (lanes 4–6), or anti-MYC control antibody (lanes 7–9) and precipitates were analyzed in Western blots using anti-FLAG. AID expression in the lysates is shown in a separate blot on the right. Expressed protein and tag are indicated above lanes. (b) Pull-down assays were performed with <sup>35</sup>S-labeled, in vitro-translated PAF1 and recombinant AID (lanes 1 and 2) or APOBEC2 (APO2; lanes 3 and 4) purified from *E. coli*. Pull downs were performed using anti-AID (lanes 1 and 3) or control antibody anti-Myc (lane 2 and 4) and analyzed on SDS-PAGE, followed by autoradiography. Assays performed with APOBEC2 protein (lanes 3 and 4) served as controls. A 10% Input is shown in lane 5. (c, left) PathDetect HLR cells transfected with GAL4-AID and GAL4-AID E58Q were analyzed 30 h after transfection for transcription by luminometer (in triplicate). Data are representative of one of two independent experiments. (c, right) 30 h after transfection, chromatin was isolated and subjected to ChIP with anti-PAF1 or anti-LEO1 antibodies. Purified precipitated DNA was analyzed with qPCR (in triplicates), amplifying the 5' part of the luciferase gene. Data are representing one of two independent experiments. Mock transfected cells (-) served as references, and the control IgG was set to 1. (d, top) A scheme representing the AID/APOBEC2 chimera. The numbers indicate the amino acids of AID that are replaced in each chimera by corresponding APOBEC2 amino acids. (d, bottom) Myc-tagged PAF1 was coexpressed with GFP-AID, GFP-APOBEC2, or GFP-AID/APOBEC2 chimera A–D (Conticello et al., 2008) in HEK293T cells. GFP-proteins were immunopurified (anti-GFP) and IPs were probed with anti-MYC and anti-GFP antibodies (middle and bottom). For input, 5% of the lysate used for IP was analyzed for MYC-PAF1 expression (top).

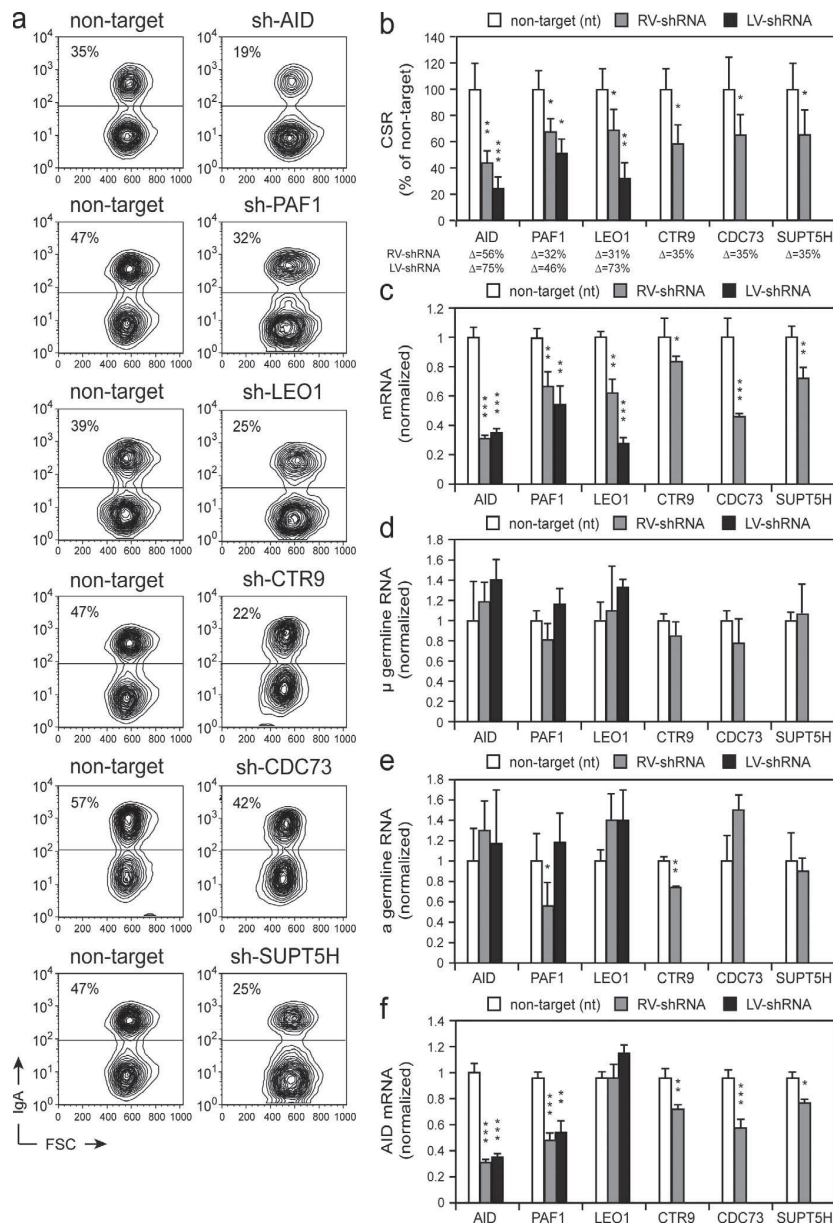
undergo CSR to IgA was determined by flow cytometry (Fig. 4, a and b). As controls we used shRNAs specific for AID and SUPT5H, together with a nontarget shRNA control.

Knockdown efficiencies were determined by qRT-PCR (Fig. 4 c). Consistent with previous results (Pavri et al., 2010), we found that knockdown of AID and SUPT5H resulted in a significant reduction of CSR efficiency (Fig. 4, a and b). Knockdown of PAF1, LEO1, and CTR9 resulted in a similar reduction in the efficiency of CSR, which ranged from 31 to 35% (Fig. 4 b, gray bars), thus indicating the involvement of the PAF complex in CSR. No effects on viability, as determined by Topro-3 staining, were observed (not depicted). CDC73 depletion showed a reduction in CSR, but the change was not as significant as that of the other PAF

complex members. To verify the retrovirus shRNA knock-down effects on the PAF complex and possibly enhance the efficacy, we developed a lentivirus-based system. Although the overall switching efficiency was reduced even in the control samples, the lentiviral caused effect was much more pronounced, with a LEO1 knockdown reducing switching by >70% (Fig. 4 b and not depicted). This enhanced CSR inhibition by LEO1, can be explained, in part, by the more pronounced reduction of the target mRNA (Fig. 4c). Importantly, although the knockdown did not lead to a complete loss of the target, biological changes in CSR were observed.

As the PAF complex is part of the RNA pol II transcription machinery, the knockdown of its individual subunits could have broader influences on the cell than just altering AID's function at the IgH locus during CSR. We thus monitored the effect of knockdown on switch region transcription and AID expression. Although transcription at the donor switch region was not affected by the knockdown of any of the PAF complex subunits (Fig. 4 d), we found that knockdown of PAF1 and CTR9 resulted in altered levels of germline transcription at the acceptor switch region (Fig. 4 e). Furthermore, knockdown of PAF1, CTR9, CDC73, and SUPT5H resulted in a significant reduction

in the level of AID mRNA (Fig. 4 f). Importantly, however, knockdown of LEO1 did not reduce AID mRNA expression (Fig. 4 f), nor reduce the levels of germline transcripts (Fig. 4, d and e), yet CSR was significantly reduced (Fig. 4, a and b); a finding that was confirmed with the lentivirus system. Because reduction in the expression of mismatch repair and base excision repair proteins, like UNG



**Figure 4. Knockdown of PAF complex members impairs CSR.** (a) CH12 cells were retrovirally transduced with shRNAs for AID, PAF1, LEO1, CTR9, CDC73, SUPT5H, or a respective nontarget. After stimulation, IgA surface expression was monitored by flow cytometry, representative plots are shown. Numbers within the FACS plots indicate the percentage of IgA-positive cells. (b) The mean percentage (+ SD) of CSR in stimulated cells that were RV-shRNA transduced (gray bars) or LV-shRNA transduced (black bars), relative to the nontarget shRNA control (white bars) set to 100% from three independent experiments, is shown. The difference ( $\Delta$ ) in CSR efficiency between nontarget and target shRNA knockdown is shown below. Significance of a decrease versus nontarget is indicated by the following p-values: \*, P < 0.05; \*\*, P < 0.01; \*\*\*, P < 0.001, based on Student's *t* test. Quantitative RT-PCR for AID, PAF1, LEO1, CTR9, CDC73, and SUPT5H transcripts (c),  $\mu$  germline transcript (d),  $\alpha$  germline transcript (e), and AID transcripts from cells transduced with RV-shRNA (gray bars) or LV-shRNA (black bars) shown relative to the corresponding nontarget shRNA control (f; white bars). Results are from three independent experiments. Transcript Ct values were normalized to CD79b or HPRT mRNA abundance and are presented relative to the levels in the nontarget shRNA negative control, set to 1. \*, P < 0.05; \*\*, P < 0.01; \*\*\*, P < 0.001, based on Student's *t* test.

and MSH2/MSH6, could also explain the observed reduction in CSR, we monitored their expression level (by qRT-PCR) after knockdown of AID, PAF1, and LEO1. We were unable to identify any significant changes in mRNA levels (unpublished data).

Reducing the expression of the PAF complex proteins induced a loss in CSR, thereby identifying the PAF complex as a key component during Ig diversification. The observation that the core PAF protein LEO1 knockdown reduced CSR threefold, whereas not altering the expression of key transcript units, indicated that the PAF complex (or at least LEO1) plays a direct role in regulating AID function at the chromatin target.

#### PAF is present on the functional Ig allele of DT40 independently of AID

As a complex associated with active transcription, the PAF complex is present on numerous genes. To determine whether the PAF complex is recruited to an active Ig locus, we performed ChIP from DT40 chromatin using antibodies specific for PAF1 and LEO1 (Fig. 5). As in most B cells, in DT40 there is a strong allelic exclusion bias with only one of the two Ig light chain (lambda) alleles being active. By designing specific primers for the active (R, rearranged) and inactive (UR, unrearranged) allele (Fig. 5, schematic), we could identify PAF1 and LEO1 to be specifically located at the active allele. The PAF1 and LEO1 occupancy near the C domain (which is present on both alleles) was analogous to that of the previously described SUPT5H, and indicated a presence of the PAF complex outside of AID-targeted regions. This also led us to investigate if AID presence was necessary for PAF complex presence at the Ig locus, and we performed the same ChIP in AID-deficient DT40 cells (Harris et al., 2002). We found that PAF1 and LEO1 occupancy at the rearranged allele was not disrupted, and was even increased, by AID-deficiency (Fig. 5), indicating an AID-independent function

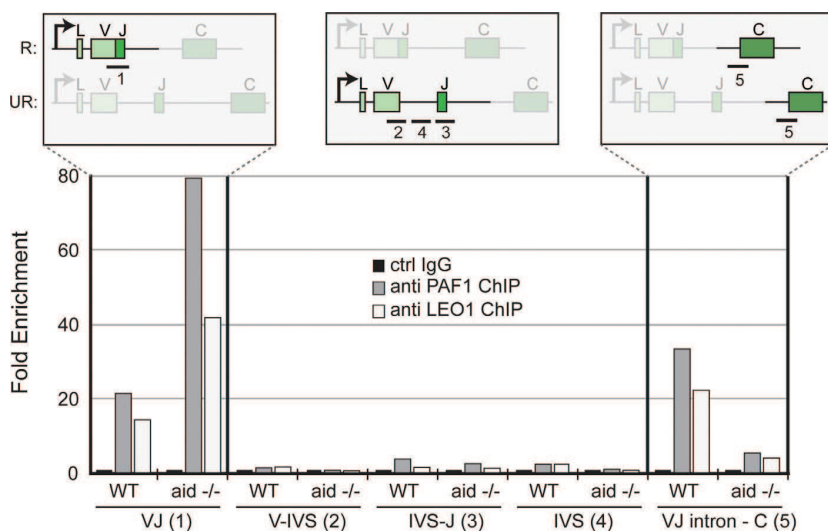
for the loading of the PAF complex proteins to Ig loci. We conclude that the PAF complex could serve as a binding platform for AID.

#### AID presence at S $\mu$ is impaired by LEO1 knockdown

If the PAF complex can serve as a site for AID association at Ig loci, then reducing PAF expression should alter AID's occupancy at an Ig locus. To determine whether AID recruitment to the S $\mu$  switch region is dependent on LEO1, ChIP experiments using an anti-AID antibody (Pavri et al., 2010) on unstimulated or stimulated transduced CH12 cells were performed (Jeevan-Raj et al., 2011). Before analysis, the cells had been transduced with a lentivirus expressing a shRNA specific for LEO1, AID, or a nontarget and were sorted for enhanced GFP expression. AID occupancy at the S $\mu$  switch region was significantly reduced in LEO1 knockdown cells when compared with the nontarget control shRNA (Fig. 6). AID-ChIP signal was specific, as there was no significant difference in AID occupancy between unstimulated CH12 cells (not expressing AID) and stimulated cells expressing an shRNA specific for AID. We conclude that AID binding to S $\mu$  is impaired by LEO1 knockdown. This result indicates that the functional mechanism of the PAF complex (at least LEO1) is to allow for AID to reside at an Ig locus during immune diversification.

#### DISCUSSION

Transcription has long been associated with AID-induced immune diversification. Early transgenic work demonstrated that the removal of the Ig promoter or enhancer elements abolished SHM (Betz et al., 1994). Mutation distribution across the V region of Ig genes indicated that AID-induced mutations are initiated 100–150 bp downstream of the transcription start site (TSS), and continue for ~1,500–2,000 bp (Rada et al., 2002). Recent work has identified a similar AID-induced mutation profile across non-Ig genes (Liu et al., 2008), although the extent and frequency of SHM on these non-Ig genes was much more restricted. This indicated that although transcription is



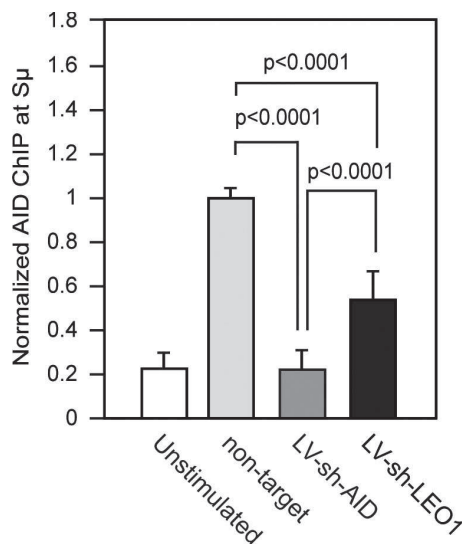
**Figure 5. PAF complex presence at the active Ig allele independent of AID.** DT40 ChIP was performed as described in the Materials and Methods section. (top) Schematics of the R (rearranged) and UR (unrearranged) lambda alleles in DT40 cells. PCR amplifications are shown as lines: 1, rearranged V/J region allele; 2, 3, and 4, unrearranged V/J region allele; 5, both C region alleles. WT, parental DT40; aid<sup>-/-</sup>, AID KO has previously been described (Harris et al., 2002). ChIP was performed using anti-PAF1 (gray bars) and anti-LEO1 (white bars), and qPCR results (in triplicate) were compared with IgG control ChIP (black bars set to 1). One of two representative experiments is shown.

crucial, location and chromatin configuration also play a significant role, whereas sequence alone does not.

Several AID-associated proteins have been identified, some of which are linked directly to RNA processing (Conticello et al., 2008; Pavri et al., 2010; Stanlie et al., 2010; Basu et al., 2011; Okazaki et al., 2011), whereas others are important for subcellular localization (Paternaude et al., 2009; Maeda et al., 2010) or substrate accessibility (Chaudhuri et al., 2003). After fractionating B cells undergoing Ig diversification, we focused on the chromatin-bound AID and its physiological interactome (Fig. 1), which consisted of RNA pol II core (RNA pol II sub unit 1A and 2A) and associated proteins (SUPT5H), splicing factors (SF3A and 3B, Prp6, PrP4), RNA helicases, chromatin modifiers (SUPT6H, SSRP1 and SUPT16H), and an RNA pol II elongation complex (PAF complex; PAF1, LEO1, CTR9, CDC73). We verified these associations in DT40 and CH12F3 cells (Fig. 2, a and b), and demonstrated that PAF1 was the likely AID-interacting subunit within the PAF complex (Fig. 3). The biological significance of the AID-PAF complex association was shown by LEO1 knock-down in induced CH12 cells, where we observed reduced CSR without reducing AID or Ig transcript levels (Fig. 4). Mechanistically, at the Ig locus, the presence of the PAF complex (Fig. 5) enhanced AID occupancy (Fig. 6).

#### Transcription-coupled AID function

Genetically, transcription has been linked to SHM and CSR (Stavnezer-Nordgren and Sirlin, 1986; Peters and Storb,



**Figure 6. AID binding to S $\mu$  is impaired after LEO1 knockdown.**

Unstimulated and/or stimulated CH12 cells were transduced with a lentivirus expressing shRNAs specific for AID, LEO1, or a nontarget shRNA control. AID occupancy at the S $\mu$  switch region was ChIP analyzed using anti-AID antibodies. Normalized AID-ChIP data from three independent experiments assayed with two different primer sets is shown. For each sample, AID-ChIP values were normalized to the input control. AID-ChIP signal in cells expressing a nontarget shRNA control was set to 1. P-values, two-tailed Student's *t* test.

1996), whereas an AID RNA pol II association has subsequently been implicated (Nambu et al., 2003). During SHM and CSR, mutations do not occur until after promoter escape (>100 bp downstream of the start site), and because of this the processing of RNA is likely a mechanistic link to AID activity. This was confirmed by the discoveries of an association between the following: AID and CTNNB1, a protein of the splicing machinery, which occurs concomitantly during RNA pol II elongation (Conticello et al., 2008); AID and PTBP2, a splicing protein (Nowak et al., 2011); AID and SUPT5H, a protein known to associate with paused and elongating RNA pol II (Pavri et al., 2010); AID and SUPT4H, a factor known to associate with SUPT5H (Stanlie et al., 2012); AID and SUPT6H, a histone chaperone (Okazaki et al., 2011); CSR and SET1, a methyl-transferase for H3K4me3 (Stanlie et al., 2010); CSR and the FACT complex, a chromatin-modifying complex during RNA processing (Stanlie et al., 2010). Because of the involvement of the various RNA biogenesis and chromatin modification proteins in AID-induced Ig diversification, one cannot exclude the possibility that some of these factors serve multiple roles in directly controlling AID at the Ig locus, in changing the chromatin state of the Ig locus through the regulation of key factors, and in influencing the pathway and resolution of AID-lesions based on altered chromatin states.

The RNA pol II C-terminal domain (CTD) tail, which is temporally and spatially modified, serves as a platform for co-transcriptional mRNA maturation and chromatin modification. The PAF complex helps to set the right co-transcriptional chromatin marks, itself serving as a docking platform for the H2B ubiquitination machinery, as well as for setting H3K4me3 marks (Jaehning, 2010). H3K4me3 serves as an important mark in CSR (Wang et al., 2009; Stanlie et al., 2010), but is generally restricted to the 5' end of a gene, and replaced by H3K36me3 toward the 3' end of the gene. Both of these marks are induced upon transcriptional activation of S-regions (Wang et al., 2009), but at these loci, the H3K4me3 domain is extended, whereas onset of H3K36me3 is pushed back toward the 3' end. This correlates roughly with the cease of mutational load/AID activity in C regions (Wang et al., 2009). Our ChIP data in DT40 confirm that the machineries required to set the various marks are also skewed along the transcription unit during Ig diversification (Fig. 5). This data also confirms that occupancy by AID-associated factors does not equate to AID occupancy, given that the gross SUPT5H and RNA pol II occupancy profile is not altered for several hundred base pairs, extending into the C region (Pavri et al., 2010), and not all stalled genes are target for AID binding or mutation (Yamane et al., 2011). Furthermore, AID has been associated with TSS of non-Ig genes (Yamane et al., 2011), yet no functional relevance (e.g., AID-induced mutations) has been identified at these locations. Therefore, the current data of linking the early transcriptional events to AID association provides further insight into the establishment of 5' boundary-marks of

SHM, whereas the understanding of molecular mechanism for the 3' boundary remains less clear.

Overall, our work now provides the biochemical (and physiological) foundation for the aforementioned AID associations, while at the same time providing the molecular link (PAF complex) between early transcription elongation, marked by SUPT5H/SUPT4H, and downstream extended chromatin modifications dependent on FACT (SSRP1 and SUPT16H), SET1, and SUPT6H (Pavri et al., 2006; Fleming et al., 2008; Chen et al., 2009; Jaehning, 2010; Selth et al., 2010). A possible order of events at the Ig locus (Fig. 7) would entail the following: RNA pol II pausing after promoter escape and phosphorylation of its CTD tail; binding of the SUPT4H–SUPT5H complex to RNA pol II; recruitment of PAF complex to the holocomplex and initiation of histone modifications near the pause site (H2B mono-ubiquitination by the BRE1/RAD6 complex serves as a platform for SET1 complex for H3K4 trimethylation) and phosphorylation of CTD and

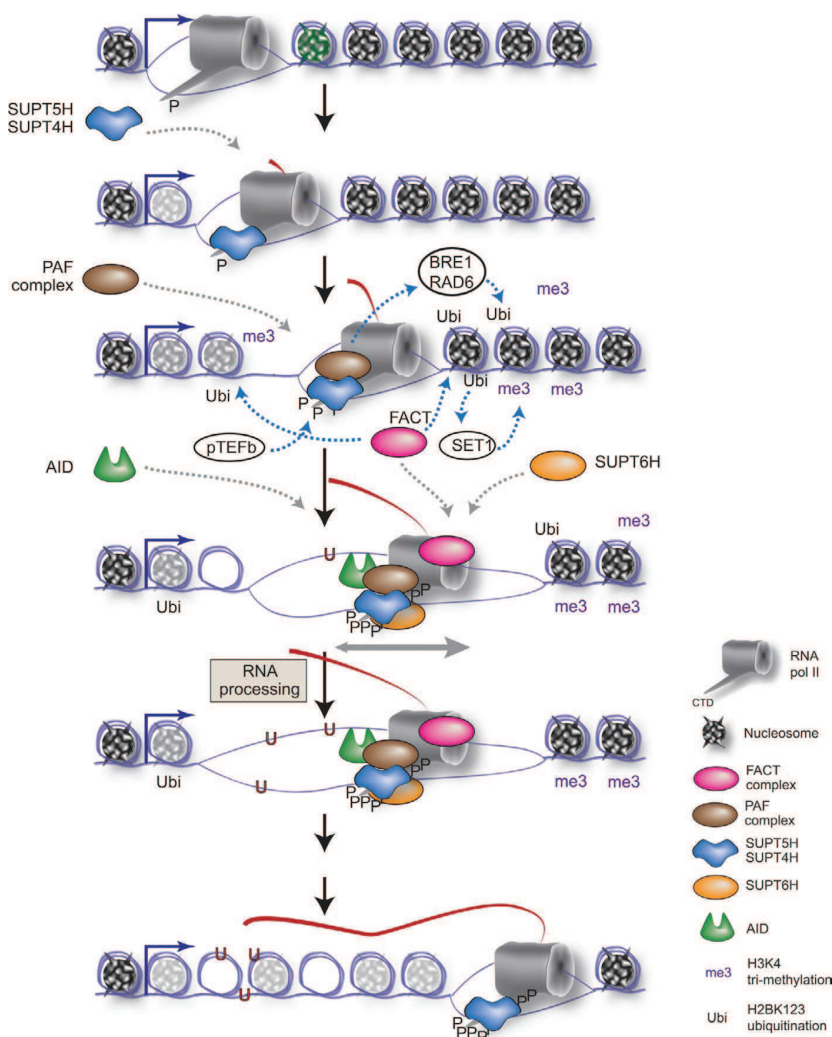
SUPT4H–SUPT5H complex by pTEFb; concomitant association of AID to the PAF–SUPT5H–RNA pol II complex, FACT complex recruitment and chromatin remodeling, SUPT6H association to the restarting polymerase; elongating/pausing transcription for enhanced AID resident time at Ig locus, RNA biogenesis, opening of chromatin and DNA for AID accessibility, recruitment of DNA repair factors to initiate SHM and CSR; hyperphosphorylation of the CTD, loss of AID association, and completion of RNA synthesis.

As mentioned above, several of the proposed proteins have been demonstrated to either associate with AID and/or play a role during Ig diversification. The identification of the nucleosome modifiers SUPT6H and FACT at the Ig locus, the demonstration that histone H3K4 trimethylation is necessary for CSR (Stanlie et al., 2010), and the correlation of H2Bser14 phosphorylation (Odegard and Schatz, 2006), H4K20 methylation (Schotta et al., 2008), H3 acetylation (Kuang et al., 2009; Wang et al., 2009), and H3K9 trimethylation (Chowdhury et al., 2008; Jeevan-Raj et al., 2011; Kuang et al., 2009; Wang et al., 2009) with Ig diversification indicates that the interplay of transcription and chromatin modification during AID-induced Ig diversification, although complex, is beginning to be unraveled. Although our data suggest that the predominant function of the PAF complex during SHM is to provide a site for AID association, we cannot exclude the possibility that reduced PAF activity also alters nucleosome marks needed for the resolution of AID-induced lesions, but more detailed future analysis may.

### SHM versus CSR

The AID–PAF complex and AID–SUPT5H interactions were isolated from DT40 cells, which undergo SHM as well as gene conversion, but do not undergo CSR. Past work has implicated histone modification during SHM, but detailed understanding is still lacking, whereas H3K4me3 seems to play an important role during CSR (Stanlie et al., 2010). Our isolation of most of the required components for setting this mark during transcription would imply a similar requirement during V region diversification.

**Figure 7. Model of the AID-linked transcriptional events at the Ig locus.** Factors not currently identified to associate with AID or SHM/CSR are in white circles; other factors are indicated in the legend.



Furthermore, we also identified PAF interactions from cells undergoing CSR. On the other hand, there have been indications that SUPT6H (Okazaki et al., 2011), SUPT4H (Stanlie et al., 2012), and the FACT components (Stanlie et al., 2010) have different functionality during SHM and CSR, but detailed analysis from knock-outs and the endogenous SHM or CSR loci need to confirm the exact mechanisms.

### Conclusion

Our work has provided biochemical and genetic insight into understanding the association of AID to the Ig locus. Our novel approach to isolate physiological AID-containing protein complexes only from chromatin has identified a new component, the PAF complex, as well as biochemically verified the significance of previously identified complexes (SUPT5H, SUPT6H, and FACT) in AID biology. Furthermore, our data extends the current model of AID gaining access to DNA by stalled RNA polymerase II to a more complex model, where AID is intimately and specifically linked with RNA pol II in the phase of pausing and elongation, surrounded by a specific chromatin environment defined by histone modification cascades.

The finding that AID interacts with the PAF and the RNA pol II elongation complexes is somewhat reminiscent of a model put forth by Peters and Storb (1996), where an unknown mutator (now known to be AID) would bind to initiating RNA pol II and travel along with the machinery during transcription elongation.

### MATERIALS AND METHODS

**Plasmids, cell lines, and antibodies.** Plasmids were constructed using standard PCR and molecular biology techniques; sequences are available upon request. Tagging AID exon 5 in DT40 has been previously described (Pauklin et al., 2009), with the following modification: instead of a 3xFLAG-2xTEV-3xMyc tagged AID construct (AID-3FM), we also generated AID-3F (3xFLAG). Expression plasmids for GFP-AID, GFP-Apobec2, and GFP-AID/Apobec2 chimera A-D were obtained from the Neuberger Laboratory (Conticello et al., 2008). A CMV promoter-driven MYC-PAF1 expression vector was obtained by cloning the human PAF1 cDNA into pcDNA<sup>TM</sup>3.1/myc-His (Invitrogen). For a complete list of antibodies used in this study please see Table S3.

**Chromatin AID-3FM and AID-3F isolation.** Isolation was based on a previously described method (Aygün et al., 2008), with modifications.  $1-2 \times 10^{10}$  DT40 cells (Pauklin et al., 2009) were collected by centrifugation at 1,200 rpm 4°C for 10 min, and cell pellets were washed twice with 50 ml cold 1xPBS. Cytoplasmic lysis: 5 times packed cell volume ( $\sim 1 \mu\text{l}$  PCV =  $10^6$  cells) of Hypotonic Lysis Buffer (HLB; 10 mM Tris HCl [pH 7.5], 2 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>, and 0.32 M sucrose, protease inhibitor cocktail [Roche], and phosphatase inhibitor cocktail [Roche]) was added to the cell pellet, resuspended gently, and incubated for 12 min on ice. To the swollen cells, 10% Triton X-100 was added to a final concentration of 0.3%. The suspension was mixed and incubated for 3 min on ice, centrifuged for 5 min at 1,000 g at 4°C, and the supernatants (cytoplasmic fraction) were collected. Nuclear pellets were washed once in HLB + 0.3% Triton X-100, resuspended in 2xPCV LB-T (LB - 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 50 mM KCl, 2 mM MgCl<sub>2</sub>, 1 mM EDTA, 10% Glycerol, protease inhibitor cocktail [Roche], phosphatase inhibitor cocktail [Roche], and 0.3% Triton X-100), and dounce homogenized with 30 strokes. The samples were incubated

with gentle agitation for 30 min at 4°C and ultracentrifuged at 33,000 g for 30 min at 4°C. The pellets were dounce homogenized until resistance was lost in 2xPCV LB-TB (LB-T + 150 U/ml Benzonase [VWR International]). The samples were incubated at room temperature for 30 min and ultracentrifuged as before. The supernatants (chromatin fraction) were subjected to a preclearing step with agarose beads before adding M2-affinity beads (Sigma-Aldrich) for IP. For PAF complex analysis, NaCl and KCl were doubled to give a final concentration of 300 mM, and the Triton X-100 concentration increased to 0.5%. For Western analysis of input, lysates between 0.5 and 3% of total lysate was loaded per lane.

**Size exclusion chromatography.** Chromatin extract of DT40 was prepared as described above. 1 ml of extract was loaded onto a Superdex 200 10/300 GL column, which had been equilibrated in LB and calibrated with standard proteins using Äkta Explorer (GE Healthcare). Fractions were collected at 1-ml volume steps using a 0.5 ml flow-rate, concentrated, and analyzed by Western blot.

**FLAG-IP.** Anti-FLAG M2 affinity beads (Sigma-Aldrich) were washed and equilibrated in LB-T. For chromatin fractions, 100  $\mu\text{l}$  of M2 beads per  $5 \times 10^9$  DT40 cells were incubated for 3-4 h with the chromatin at 4°C on a rotator and collected for 3 min at  $300 \times g$  at 4°C. Beads were washed 5 times in 25 $\times$  bead volume of LB-TF (LB-T supplemented with 0.5-1  $\mu\text{g}/\text{ml}$  1xFLAG peptide N-DYDDDDDK-C) and once with LB at 4°C for 10 min, followed by two elution steps in 4 $\times$  bead volume of EB (LB + 500  $\mu\text{g}/\text{ml}$  3xFLAG peptide N-MDYKDHGDYKDHHDIDYKDDDDK-C); first for 1 h at room temperature, and then over night at 4°C.

**Mass spectrometry.** Polyacrylamide gel slices (1-2 mm) containing IP-purified proteins were prepared for mass spectrometric analysis using the Janus liquid handling system (PerkinElmer). In brief, the excised protein gel pieces were placed in a well of a 96-well microtiter plate and de-stained with 50% vol/vol acetonitrile and 50 mM ammonium bicarbonate, and then reduced with 10 mM DTT and alkylated with 55 mM iodoacetamide. After alkylation, proteins were digested with 6  $\mu\text{g}/\mu\text{l}$  trypsin (Promega) overnight at 37°C. The resulting peptides were extracted in 2% vol/vol formic acid, 2% vol/vol acetonitrile. The digest was analyzed by nano-scale capillary LC-MS/MS using a nanoAcquity UPLC (Waters) to deliver a flow of  $\sim 300$  nL/min. A C18 Symmetry 5  $\mu\text{m}$ , 180  $\mu\text{m} \times 20$  mm  $\mu$ -Precolumn (Waters), trapped the peptides before separation on a C18 BEH130 1.7  $\mu\text{m}$ , 75  $\mu\text{m} \times 100$  mm analytical UPLC column (Waters). Peptides were eluted with a gradient of acetonitrile. The analytical column outlet was directly interfaced via a modified nano-flow electrospray ionization source, with a linear quadrupole ion trap mass spectrometer (LTQ XL/ETD, Thermo Fisher Scientific). LC-MS/MS information was collected using a data dependent analysis procedure. MS/MS scans were collected using an automatic gain control value of  $4 \times 10^4$  and a threshold energy of 35 for collision induced dissociation. LC-MS/MS data were then searched against a protein database (UniProt Knowledge Base) using the Mascot search engine program (Matrix Science; Perkins et al., 1999). Database search parameters were set with a precursor tolerance of 1.0 D and a fragment ion mass tolerance of 0.8 D. One missed enzyme cleavage was allowed and variable modifications for oxidized methionine, carbamidomethyl cysteine, phosphorylated serine, threonine and tyrosine were included. MS/MS data were validated using the Scaffold program (Proteome Software, Inc.; Keller et al., 2002). All data were additionally interrogated manually.

**Western blotting.** The antibodies used are shown in Table S3. Samples were prepared using standard procedures. Proteins were fractionated using NuPage Bis-Tris gels (Invitrogen) or homemade 10% PAA gels and transferred to Immobilon-P membranes (Millipore).

**Nuclear extracts and coIP in murine B cells.** Nuclear extracts were prepared using standard techniques from CH12F3 cells stably expressing AID<sup>FLAG-HA</sup> (Jeevan-Raj et al., 2011). coIPs and Western blot analyses were



performed as previously described (Jeevan-Raj et al., 2011). For antibodies used, please see Table S3.

**In vitro translation and coIP.** AID-His-tagged protein was expressed in *E. coli* and purified as previously described (Coker et al., 2006). <sup>35</sup>S-labeled PAF1 was expressed using the TnT T7 Coupled Reticulocyte Lysate System (IVT) according to the manufacturer's instructions (Promega). Labeled protein mixture was mixed with 100 ng of AID or 300 ng of APOBEC2 protein for 1 h at room temperature and for 30 min at 4°C. Proteins were isolated by anti-AID (hAnp52-1; Conticello et al., 2008) or anti-Myc (9E10) coupled to Sepharose beads for 1 h at 4°C, washed 5 times in 1× TBS-T (50 mM Tris, pH 8.0, 300 mM NaCl, 1% Triton X-100, 2.5 mM TCEP, 2% BSA, and protease inhibitor [Roche]), resuspended in SDS-PAGE loading buffer, and separated on 12% Bis-Tris polyacrylamide gels (Invitrogen). Gels were dried, exposed, and analyzed using a Fuji Imaging system.

**E. coli coIP.** cDNAs of PAF1 and CDC73 were fused to a C-terminal FLAG tag in a pET DUET derivative coexpressing untagged human AID. Plasmids were transformed into BL21-CODONPLUS (DE3)-RIL cells (Stratagene), and protein expression was induced at 16°C with 1 mM IPTG in the presence of 0.1 mM ZnCl<sub>2</sub> (3 h). Cells were sonicated in TBS-T, debris were pelleted at 19,000 g, and IPs were performed using Sepharose-coupled anti-AID hAnp52-1 or anti-FLAG M2 antibodies. After five washes with TBS-T, the immunoprecipitates were analyzed by Western blot, using polyclonal anti-AID (Abcam) and monoclonal anti-FLAG (M2-HRP; Sigma-Aldrich) antibodies.

**HeLa PathDetect analysis.** The Stratagene PathDetect HLR Cell Line and GAL4-CREB and PKA expression vectors were purchased from Agilent. This HeLa-based Luciferase Reporter cell line contains a single locus with integrated synthetic minimal promoter and five yeast GAL4-binding sites (UAS) driving expression of the luciferase gene. Plasmids expressing GAL4-CREB, PKA, and GAL4-AID were transfected using Lipofectamine2000 (Invitrogen), and luciferase activity (in triplicate) was monitored 24–48 h after transfection according to the Luciferase Assay System manual (Promega). ChIP analysis using anti-PAF1, anti-LEO1, and a control IgG were done as follows: cells were cross-linked with 1% formaldehyde, nuclei were isolated and lysed in sonication buffer (1% SDS, 50 mM Tris HCl, pH 8.0, and 10 mM EDTA). After sonication in a BioRuptor, fragmented chromatin was diluted and incubated with antibodies or control IgGs over night. Collected protein-DNA complexes were purified and analyzed by qPCR (in triplicate). ChIP data were normalized to the input signal for each chromatin sample, and control ChIPs were set to 1. For antibodies used, please see Table S3. For oligonucleotides used, please see Table S1. Two independent experiments were performed with one representative shown.

**coIP in HEK293T cells.** HEK293T cells were transfected with a plasmid expressing MYC-PAF1. 12 h after transfection cells were pooled to guarantee equal expression of MYC-PAF1, and split to allow a second transfection (12 h later) with expression plasmids for GFP-AID, GFP-APOBEC2, or GFP-AID/APO2 chimera. 24 h after second transfection, cells were lysed (lysis buffer: 50 mM Tris HCl, pH 8.0, 150 mM NaCl, 0.04% SDS, 1% NP-40), and GFP-protein expression in the lysates was estimated by scanning aliquots of a dilution series of the lysates with a Typhoon Scanner. Equal GFP and protein amounts were subjected to IP with anti-GFP at 4°C on, immunoprecipitates were collected with protein-A/G-Sepharose beads, and beads were washed and analyzed by Western blotting.

**shRNA knockdown.** Retroviral knockdown was done as follows: vectors containing shRNAs specific for SUPT5H, PAF1, LEO1, CTR9, CDC73, and the nontarget shRNA control were purchased from OriGene (Table S2). The hairpin sequence for AID (5'-ACCAGTCGCCATTATAATGCAA-3') was cloned into the LMP retroviral vector (Open Biosystems). CH12 cells were transfected as previously described (Barreto et al., 2003). Transduced cells were selected with 0.5 µg/ml puromycin for 1–5 d before induction

and sorting. Lentiviral knockdown was done as follows: The hairpin sequences for AID, PAF1, LEO1, and the nontarget shRNA control were cloned into the pLKO.1-puro-CMV-TurboGFP lentiviral vector (Sigma-Aldrich). Lenti-X 293T cells (Takara Bio Inc.) were transfected with vectors to produce the virus. 2 d later, CH12F3 cells were spin-infected with viral supernatants supplemented with 10 µg/ml polybrene (Santa Cruz Biotechnology). Cells were selected for 5 d with 1 µg/ml puromycin before induction. Hairpin sequences used are listed in Table S2.

**Cell culture and flow cytometry.** Retrovirally transduced CH12 cells were cultured with 5 ng/ml IL-4 (Sigma-Aldrich), 0.3 ng/ml TGF-β (R&D Systems), monoclonal 200 ng/ml anti-CD40 antibody (eBioscience), and 0.5 µg/ml puromycin and analyzed after 48–72 h for CSR (IgM to IgA) by flow cytometry, as previously described (Robert et al., 2009).

**Real-time quantitative RT-PCR.** RNA and cDNA were prepared using standard techniques. qPCR was performed in triplicates using SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich) and a LightCycler 480 (Roche). Transcript quantities were calculated relative to standard curves and normalized to CD79b or HPRT mRNA. For primers see Table S1.

**ChIP from DT40 and CH12 cells.** In brief, DT40 cells were treated and analyzed as for the ChIP in the HeLa PathDetect analysis section. For antibodies used please see Table S3. For oligonucleotides used please see Table S1. Two independent experiments were performed, with one representative shown. For quantitative AID-ChIP from shRNA knockdowns: CH12 cells were transduced with a lentivirus expressing shRNAs specific for AID, LEO1 and a nontarget control. Cells were stimulated for 48 h and sorted for enhanced GFP expression using a FACS Aria II (BD) and/or FACSVantage SE (BD) cell sorters before ChIP analysis. Cells were cross-linked with 1% formaldehyde for 10 min. Chromatin was prepared and immunoprecipitated with an anti-AID antibody (Pavri et al., 2010) and analyzed by quantitative PCR as previously described (Jeevan-Raj et al., 2011). Raw data were normalized to the input signal for each sample. AID-ChIP signal in cells expressing a nontarget shRNA control was assigned an arbitrary value of 1. Statistical analysis was performed using a two-tailed Student's *t* test.

**Online supplemental material.** Fig. S1 is a schematic of how the isolation and analysis of the AID-associated complex was undertaken and a table of peptide IDs. Fig. S2 shows the AID interactome. Table S1 shows primer sequences. Table S2 lists shRNA sequences. Table S3 lists antibodies used in this study. Online supplemental material is available at <http://www.jem.org/cgi/content/full/jem.20112145/DC1>.

We would like to thank the members of the Petersen-Mahrt and Reina-San-Martin laboratories for discussions; Cancer Research UK (CRUK) cell services for performing DT40 growth; Michel Nussenzweig for the anti-AID antibody and Anna Gazumyan for advice on AID-ChIP; Gudrun Bachmann and Dafne Solera for help in the generation of cell lines; Claudine Ebel for cell sorting; and Jesper Svejstrup for discussion and critical reading of the manuscript.

K.L. Willman, S. Pauklin, G. Rangam, M.T. Simon, S. Maslen, and M. Skehel were supported by CRUK. S. Pauklin was supported in part by SA Archimedes-Estonian Foundation of European Union Education and Research. K.-M. Schmitz is supported by a Marie Curie FP 7 fellowship. S. Milosevic was supported by la Ligue Contre le Cancer, France. B. Reina-San-Martin is an AVENIR-INSERM young investigator. This work was supported by grants to B. Reina-San-Martin from the Agence Nationale pour la Recherche (ANR-07-MIME-004-01) and the Institut National de la Santé et de la Recherche Médicale (INSERM), and to S.K. Petersen-Mahrt from Istituto FIRC di Oncologia Molecolare, Italy (IFOM) and CRUK.

The authors have no conflicting financial interests.

Author contributions: K.L. Willman, S. Milosevic, S. Pauklin, K.-M. Schmitz, G. Rangam, M.T. Simon, I. Robert, V. Heyer, and E. Schiavo performed experiments. S. Pauklin performed the initial AID on chromatin fractionation and isolation. S. Maslen and M. Skehel performed mass spectrometry analysis. K.L. Willman,

S. Milosevic, K.-M. Schmitz, B. Reina-San-Martin, and SKPM analyzed the data. K.-M. Schmitz, K.L. Willman, S. Milosevic, B. Reina-San-Martin, and S.K. Petersen-Mahrt wrote the paper. K.L. Willman, SM, S. Pauklin, K.-M. Schmitz, B. Reina-San-Martin, and S.K. Petersen-Mahrt designed the experiments. S.K. Petersen-Mahrt conceived the approach.

Submitted: 10 October 2011

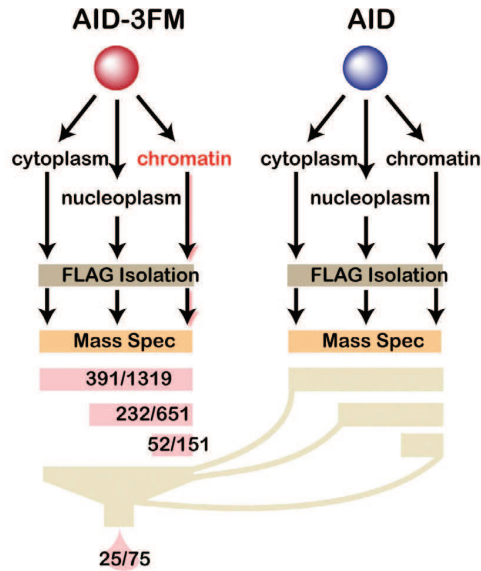
Accepted: 22 August 2012

## REFERENCES

- Ayguin, O., J. Svejstrup, and Y. Liu. 2008. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proc. Natl. Acad. Sci. USA.* 105:8580–8584. <http://dx.doi.org/10.1073/pnas.0804424105>
- Barreto, V., B. Reina-San-Martin, A.R. Ramiro, K.M. McBride, and M.C. Nussenzweig. 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol. Cell.* 12:501–508. [http://dx.doi.org/10.1016/S1097-2765\(03\)00309-5](http://dx.doi.org/10.1016/S1097-2765(03)00309-5)
- Basu, U., F.L. Meng, C. Keim, V. Grinstein, E. Pefanis, J. Eccleston, T. Zhang, D. Myers, C.R. Wasserman, D.R. Wesemann, et al. 2011. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell.* 144:353–363. <http://dx.doi.org/10.1016/j.cell.2011.01.001>
- Betz, A.G., C. Milstein, A. González-Fernández, R. Pannell, T. Larson, and M.S. Neuberger. 1994. Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell.* 77:239–248. [http://dx.doi.org/10.1016/0092-8674\(94\)90316-6](http://dx.doi.org/10.1016/0092-8674(94)90316-6)
- Bransteitter, R., P. Pham, M.D. Scharff, and M.F. Goodman. 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA.* 100:4102–4107. <http://dx.doi.org/10.1073/pnas.0730835100>
- Brar, S.S., M. Watson, and M. Diaz. 2004. Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *J. Biol. Chem.* 279:26395–26401. <http://dx.doi.org/10.1074/jbc.M403503200>
- Chaudhuri, J., M. Tian, C. Khuong, K. Chua, E. Pinaud, and F.W. Alt. 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature.* 422:726–730. <http://dx.doi.org/10.1038/nature01574>
- Chen, Y., Y. Yamaguchi, Y. Tsugeno, J. Yamamoto, T. Yamada, M. Nakamura, K. Hisatake, and H. Handa. 2009. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev.* 23:2765–2777. <http://dx.doi.org/10.1101/gad.1834709>
- Chowdhury, M., O. Forouhi, S. Dayal, N. McCloskey, H.J. Gould, G. Felsenfeld, and D.J. Fear. 2008. Analysis of intergenic transcription and histone modification across the human immunoglobulin heavy-chain locus. *Proc. Natl. Acad. Sci. USA.* 105:15872–15877. <http://dx.doi.org/10.1073/pnas.0808462105>
- Coker, H.A., H.D. Morgan, and S.K. Petersen-Mahrt. 2006. Genetic and in vitro assays of DNA deamination. *Methods Enzymol.* 408:156–170. [http://dx.doi.org/10.1016/S0076-6879\(06\)08010-4](http://dx.doi.org/10.1016/S0076-6879(06)08010-4)
- Conticello, S.G., K. Ganesh, K. Xue, M. Lu, C. Rada, and M.S. Neuberger. 2008. Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNB1. *Mol. Cell.* 31:474–484. <http://dx.doi.org/10.1016/j.molcel.2008.07.009>
- Di Noia, J.M., and M.S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* 76:1–22. <http://dx.doi.org/10.1146/annurev.biochem.76.061705.090740>
- Fleming, A.B., C.-F. Kao, C. Hillyer, M. Pikaart, and M.A. Osley. 2008. H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol. Cell.* 31:57–66. <http://dx.doi.org/10.1016/j.molcel.2008.04.025>
- Ganesh, K., S. Adam, B. Taylor, P. Simpson, C. Rada, and M. Neuberger. 2011. CTNNB1 is a novel nuclear localization sequence-binding protein that recognizes RNA-splicing factors CDC5L and Prp31. *J. Biol. Chem.* 286:17091–17102. <http://dx.doi.org/10.1074/jbc.M110.208769>
- Harris, R.S., J.E. Sale, S.K. Petersen-Mahrt, and M.S. Neuberger. 2002. AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr. Biol.* 12:435–438. [http://dx.doi.org/10.1016/S0960-9822\(02\)00717-0](http://dx.doi.org/10.1016/S0960-9822(02)00717-0)
- Ito, S., H. Nagaoka, R. Shinkura, N. Begum, M. Muramatsu, M. Nakata, and T. Honjo. 2004. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc. Natl. Acad. Sci. USA.* 101:1975–1980. <http://dx.doi.org/10.1073/pnas.0307335101>
- Jaehning, J.A. 2010. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim. Biophys. Acta.* 1799:379–388. <http://dx.doi.org/10.1016/j.bbagr.2010.01.001>
- Jeevan-Raj, B.P., I. Robert, V. Heyer, A. Page, J.H. Wang, F. Cammas, F.W. Alt, R. Losson, and B. Reina-San-Martin. 2011. Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *J. Exp. Med.* 208:1649–1660. <http://dx.doi.org/10.1084/jem.20110118>
- Keller, A., A.I. Nesvizhskii, E. Kolker, and R. Aebersold. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74:5383–5392. <http://dx.doi.org/10.1021/ac025747h>
- Kim, J., M. Guermah, and R.G. Roeder. 2010. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell.* 140:491–503. <http://dx.doi.org/10.1016/j.cell.2009.12.050>
- Kuang, F.L., Z. Luo, and M.D. Scharff. 2009. H3 trimethyl K9 and H3 acetyl K9 chromatin modifications are associated with class switch recombination. *Proc. Natl. Acad. Sci. USA.* 106:5288–5293. <http://dx.doi.org/10.1073/pnas.0901368106>
- Liu, M., J.L. Duke, D.J. Richter, C.G. Vinuesa, C.C. Goodnow, S.H. Kleinstein, and D.G. Schatz. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature.* 451:841–845. <http://dx.doi.org/10.1038/nature06547>
- Maeda, K., S.K. Singh, K. Eda, M. Kitabatake, P. Pham, M.F. Goodman, and N. Sakaguchi. 2010. GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *J. Biol. Chem.* 285:23945–23953. <http://dx.doi.org/10.1074/jbc.M110.131441>
- McBride, K.M., V. Barreto, A.R. Ramiro, P. Stavropoulos, and M.C. Nussenzweig. 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J. Exp. Med.* 199:1235–1244. <http://dx.doi.org/10.1084/jem.20040373>
- Nambu, Y., M. Sugai, H. Gonda, C.-G. Lee, T. Kataikai, Y. Agata, Y. Yokota, and A. Shimizu. 2003. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science.* 302:2137–2140. <http://dx.doi.org/10.1126/science.1092481>
- Novak, U., A.J. Matthews, S. Zheng, and J. Chaudhuri. 2011. The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nat. Immunol.* 12:160–166. <http://dx.doi.org/10.1038/ni.1977>
- Odegard, V.H., and D.G. Schatz. 2006. Targeting of somatic hypermutation. *Nat. Rev. Immunol.* 6:573–583. <http://dx.doi.org/10.1038/nri1896>
- Okazaki, I.-M., K. Okawa, M. Kobayashi, K. Yoshikawa, S. Kawamoto, H. Nagaoka, R. Shinkura, Y. Kitawaki, H. Taniguchi, T. Natsume, et al. 2011. Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proc. Natl. Acad. Sci. USA.* 108:7920–7925. <http://dx.doi.org/10.1073/pnas.1104423108>
- Patenaude, A.M., A. Orthwein, Y. Hu, V.A. Campo, B. Kavli, A. Buschiazzo, and J.M. Di Noia. 2009. Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat. Struct. Mol. Biol.* 16:517–527. <http://dx.doi.org/10.1038/nsmb.1598>
- Pauklin, S., I.V. Sernández, G. Bachmann, A.R. Ramiro, and S.K. Petersen-Mahrt. 2009. Estrogen directly activates AID transcription and function. *J. Exp. Med.* 206:99–111. <http://dx.doi.org/10.1084/jem.20080521>
- Pavri, R., B. Zhu, G. Li, P. Trojer, S. Mandal, A. Shilatifard, and D. Reinberg. 2006. Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell.* 125:703–717. <http://dx.doi.org/10.1016/j.cell.2006.04.029>

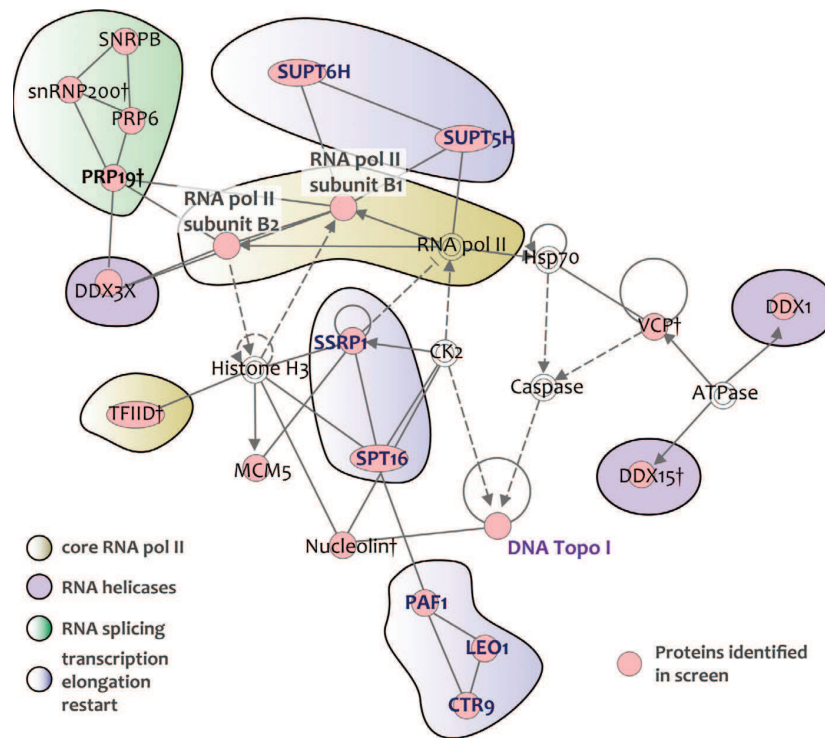
- Pavri, R., A. Gazumyan, M. Jankovic, M. Di Virgilio, I. Klein, C. Ansarah-Sobrinho, W. Resch, A. Yamane, B. Reina San-Martin, V. Barreto, et al. 2010. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*. 143:122–133. <http://dx.doi.org/10.1016/j.cell.2010.09.017>
- Perkins, D.N., D.J. Pappin, D.M. Creasy, and J.S. Cottrell. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 20:3551–3567.
- Peters, A., and U. Storb. 1996. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity*. 4:57–65. [http://dx.doi.org/10.1016/S1074-7613\(00\)80298-8](http://dx.doi.org/10.1016/S1074-7613(00)80298-8)
- Petersen-Mahrt, S. 2005. DNA deamination in immunity. *Immunol. Rev.* 203:80–97. <http://dx.doi.org/10.1111/j.0105-2896.2005.00232.x>
- Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger. 2002. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*. 418:99–103. <http://dx.doi.org/10.1038/nature00862>
- Rada, C., J.M. Jarvis, and C. Milstein. 2002. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc. Natl. Acad. Sci. USA*. 99:7003–7008. <http://dx.doi.org/10.1073/pnas.092160999>
- Rajewsky, K. 1996. Clonal selection and learning in the antibody system. *Nature*. 381:751–758. <http://dx.doi.org/10.1038/381751a0>
- Robert, I., F. Dantzer, and B. Reina-San-Martin. 2009. Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *J. Exp. Med.* 206:1047–1056. <http://dx.doi.org/10.1084/jem.20082468>
- Schotta, G., R. Sengupta, S. Kubicek, S. Malin, M. Kauer, E. Callén, A. Celeste, M. Pagani, S. Opravil, I.A. De La Rosa-Velazquez, et al. 2008. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* 22:2048–2061. <http://dx.doi.org/10.1101/gad.476008>
- Selth, L.A., S. Sigurdsson, and J.Q. Svejstrup. 2010. Transcript Elongation by RNA Polymerase II. *Annu Rev Biochem*.
- Stanlie, A., M. Aida, M. Muramatsu, T. Honjo, and N.A. Begum. 2010. Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proc. Natl. Acad. Sci. USA*. 107:22190–22195. <http://dx.doi.org/10.1073/pnas.1016923108>
- Stanlie, A., N.A. Begum, H. Akiyama, and T. Honjo. 2012. The DSIF subunits Spt4 and Spt5 have distinct roles at various phases of immunoglobulin class switch recombination. *PLoS Genet.* 8:e1002675. <http://dx.doi.org/10.1371/journal.pgen.1002675>
- Stavnezer, J., J.E. Guikema, and C.E. Schrader. 2008. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* 26:261–292. <http://dx.doi.org/10.1146/annurev.immunol.26.021607.090248>
- Stavnezer-Nordgren, J., and S. Sirlin. 1986. Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *EMBO J.* 5:95–102.
- Wang, L., R. Wuerffel, S. Feldman, A.A. Khamlichi, and A.L. Kenter. 2009. S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.* 206:1817–1830. <http://dx.doi.org/10.1084/jem.20081678>
- Yamane, A., W. Resch, N. Kuo, S. Kuchen, Z. Li, H.-W. Sun, D.F. Robbiani, K. McBride, M.C. Nussenzweig, and R. Casellas. 2011. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol.* 12:62–69. <http://dx.doi.org/10.1038/ni.1964>

## SUPPLEMENTAL MATERIAL

Willmann et al., <http://www.jem.org/cgi/content/full/jem.20112145/DC1>

Filtered Proteins	Name	Complex	kD	NC	NA	CC	CA
Transcription elongation factor SPT5	SUPT5H	DSIF	120 kD	0	0	0	11
Splicing factor 3A	SF3A		89 kD	0	4	0	8
RNA polymerase-associated protein CTR9 homolog	CTR9	PAF	134 kD	0	0	0	6
DNA-directed RNA polymerase II subunit RPB2	RNA pol II B2	core RNA pol II	134 kD	0	0	0	6
DNA-directed RNA polymerase II subunit RPB1	RNA pol II B1	core RNA pol II	217 kD	1	0	0	5
RNA polymerase II-associated factor 1 homolog	PAF1	PAF	60 kD	0	0	0	4
Transcription elongation factor SPT6	SUPT6H	chromatin/ elongation	199 kD	0	0	0	4
Lamin-B2	Lamin B2		68 kD	0	0	0	4
DNA topoisomerase I	DNA Topo I		91 kD	0	0	0	2
ATP-dependent RNA helicase DDX3Y	DDX3Y		72 kD	0	1	0	2
Interleukin enhancer binding factor 3	IL3IF		77 kD	0	1	0	2
RNA polymerase-associated protein LEO1	LEO1	PAF	75 kD	0	0	0	2
DNA replication licensing factor MCM5	MCM5		82 kD	0	0	0	2
WD40 repeat-containing protein SMU1	SMU1		57 kD	0	0	0	2
Nuclease-sensitive element-binding protein 1	YBOX1		36 kD	0	1	0	1
Pre-mRNA-processing factor 6	PRP6		107 kD	0	1	0	1
Putative uncharacterized protein	Unknown		72 kD	0	0	0	1
U4/U6 small nuclear ribonucleoprotein Prp4	PRP4		58 kD	0	0	0	1
FACT complex subunit SPT16	SPT16	FACT	120 kD	0	0	0	1
ATP-dependent RNA helicase DDX1	DDX1		82 kD	0	0	0	1
FACT complex subunit SSRP1	SSRP1	FACT	80 kD	0	0	0	1
Uncharacterized protein C17orf85 homolog	C17orf85		69 kD	0	0	0	1
Lamin-B1	Lamin B1		67 kD	0	0	0	1
Splicing factor 3B subunit 1	SF3B		146 kD	0	4	1	3
ATP-dependent RNA helicase DDX15	DDX15		88 kD	1	2	0	3
<b>Total</b>				<b>2</b>	<b>14</b>	<b>1</b>	<b>75</b>

**Figure S1. Immunoprecipitation of AID from chromatin of DT40 and complex analysis.** (a) Schematic of complex isolation and analysis. DT40 cells expressing tagged AID-3FM (red) and untagged AID (blue, control) were fractionated into cytoplasm, nucleoplasm, and chromatin. Each fraction was then subjected to FLAG bead IP and Mass Spec analysis. A total of 1,319 peptides (391 proteins) were identified from all AID-3FM fractions, with 151 peptides (52 proteins) from the chromatin fraction only. The equivalent Mass Spec dataset from control cells was used as a filter to remove false positives (yellow lines/boxes), leaving 75 peptides (25 proteins). (b) Table of the peptide ids from Mass Spec analysis of AID-3FM - chromatin fraction after filtering. NC, nucleoplasm control; NA, nucleoplasm AID-3FM; CC, chromatin control; CA, chromatin AID-3FM. RNA pol II B1 was included because it usually associates with RNA pol II B2, and this complex has 11 peptides with AID and 1 without. SF3B was included because it associates with SF3A, and this complex has 19 peptides with AID and 1 without. DDX15 was included because it has five peptides with AID and one without.



**Figure S2.** Filtered AID chromatin mass spec ID data were submitted to the Ingenuity Systems Pathway Analysis gene network software and allowed to generate a Network tree, receiving a tree score of 75. A second tree was also generated, and it obtained a score of 22. Those proteins that were also detected in the control dataset were removed from the tree. Proteins marked with a † were specific (0 peptides in control) or highly enriched (<15% in control) in the chromatin AID fraction, but less or nonspecific in the nucleoplasm. Identified proteins are in light red, solid lines represent direct interactions, dashed lines represent functional interactions. Color schemes as indicated.

**Table S1.** Primer sequences

Gene	Primer 1 – sequence	Primer 2 – sequence	Reference / Name	Purpose
<i>CD79b</i>	5'-CCACACTGGTGTCTTCC-3'	5'-GGGCTTCCTTGGAAATTCAG-3'		qRT-PCR
<i>HPRT</i>	5'-GTTGGATACAGCCAGACTTGTG-3'	5'-GATTCAACTTGCCTCATCTTAGGC-3'		qRT-PCR
<i>GLT<math>\alpha</math></i>	5'-CAAGAAGGAGAAGGTGATTAG-3'	5'-GAGCTGGTGGGAGTGTCAAGT-3'		qRT-PCR
<i>GLT<math>\mu</math></i>	5'-ACCTGGGAATGTATGGTTGTGGCTT-3'	5'-TCTGAACCTCAAGGATGCTCTTG-3'		qRT-PCR
<i>SPT5</i>	5'-TGCACTGCAAGAAGCTGGTGA-3'	5'-GCTCATAGGAGTGAAGCCACCA-3'	OriGene (MP216410)	qRT-PCR
<i>PAF1</i>	5'-GGAGGAAGAGATGGAGGCTGAA-3'	5'-CACTTGCCTCATCTCTGTACC-3'	OriGene (MP210531)	qRT-PCR
<i>LEO1</i>	5'-GAGGAGCAAGACCAGAAGTCA-3'	5'-TGTCGCTGTCTGCTCGGAATC-3'	OriGene (MP207353)	qRT-PCR
<i>CTR9</i>	5'-GTGACACCTACTCTATGCTGGC-3'	5'-TGGCAGCATACAGGTTCTTGGC-3'	OriGene (MP203022)	qRT-PCR
<i>CDC73</i>	5'-GAGAGAGTGTGGAGGACAAGAAC-3'	5'-GCACGACCTTCTCTCTGGCTT-3'	OriGene (MP202272)	qRT-PCR
<i>AID</i>	5'-GAAAGTCACGCTGGAGCCG-3'	5'-TCTCATGCCGTCCCTTGG-3'		qRT-PCR
<i>MSH2</i>	5'-GGGATGTGACGAAGCCGAGCC-3'	5'-TGCTCTCCTCCGACATGGCAGT-3'		qRT-PCR
<i>MSH6</i>	5'-CTCGTCGCCGAGGCAAAGG-3'	5'-TAGGCAAGGCCACCAGGGT-3'		qRT-PCR
<i>UNG</i>	5'-GTCTATCCGCCCCCGGAGCA-3'	5'-AACTGGGCGGGGGTGAAGT-3'		qRT-PCR
<i>HeLa A</i>	5'-CTCCGAGCGGAGACTCTAGAG-3'	5'-CGTACGTGATGTTACCTCG-3'	KMS6703 / KMS6716	ChIP-qPCR
<i>HeLa B</i>	5'-GAAGCGAAGGTTGTGGATCTG-3'	5'-CTTGTAATCAAGGCGTTGGTC-3'	KMS6705 / KMS6706	ChIP-qPCR
<i>DT40 1</i>	5'-CCTTCACGATTCTCCGTTT-3'	5'-CACCTAGGACGGTCAGGGT-3'	DS4506 / DS4507	ChIP-qPCR
<i>DT40 2</i>	5'-CCTTCACGATTCTCCGTTT-3'	5'-TTCCCATGCTTGTGTGTCAC-3'	DS4506 / DS4516	ChIP-qPCR
<i>DT40 3</i>	5'-CATCCATCACTCTGACCC-3'	5'-CACCTAGGACGGTCAGGGT-3'	DS4517 / DS4507	ChIP-qPCR
<i>DT40 4</i>	5'-CCTTGAAGAGGTGAGGAGG-3'	5'-GCAGAAACAGCCCAAGCAGC-3'	KMS6911 / KMS6912	ChIP-qPCR
<i>DT40 5</i>	5'-CAGAGGTGCATGTGTCTG-3'	5'-GTTGAGCTCCTCTTGTGACG-3'	KMS6915 / KMS6916	ChIP-qPCR

**Table S2.** shRNA (single and sets)

Target	Vector	Sequence	Reference	Source
PAF1	RV-1	5'-GGTGACGGAGTTACTACAATGAGCTGGA-3'	GI518753	OriGene
LEO1	RV-1	5'-GTGGCAGTGACAATCACTCTGAACGGTCA-3'	GI529047	OriGene
CTR9	RV-1	5'-GATGAGGATCCGACAGTGACCAGCCGTC-3'	GI528588	OriGene
CDC73	RV-1	5'-GACGTGCTCAGCGCTCCTGCGACAGTACAA-3'	GI561719	OriGene
SPT5	RV-1	5'-GCTTGGCTACTGGAACCAGCAGATGGTGC-3'	GI336016	OriGene
non-target	RV-1	5'-GCACTACCAGAGCTAACTCAGATAGTACT-3'	TR30007	OriGene
AID	RV-2	5'-ACCAGTCGCCATTATAATGCAA-3'		Open BioSystems
PAF1	LV	5'-GAACCAAGTTGTGGCTTATTT-3'	TRCN0000197886	Sigma-Aldrich
LEO1	LV	5'-GACTTGGGCAATGACTTATAT-3'	TRCN0000243542	Sigma-Aldrich
AID	LV	5'-GCGAGATGCATTTCTGTATGTT-3'	TRCN0000112031	Sigma-Aldrich
non-target	LV	5'-CAACAAGATGAAGAGCACCAA-3'		Sigma-Aldrich

Vectors: RV-1, pGFP-V-RS (retrovirus); RV-2, pLMP (retrovirus); LV, pLKO.1-puro-CMV-TurboGFP+ (lentivirus).

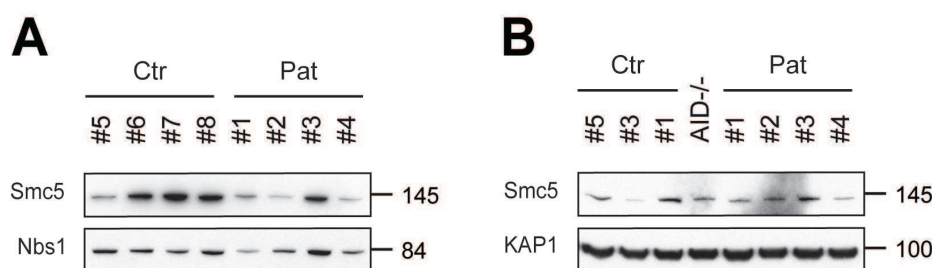
**Table S3.** Antibodies

Antibody	Clone	Source	Usage
FLAG	M2	Sigma-Aldrich	IP, Western
PAF1	ab20662	Abcam	Western
PAF1	A300-172A	Bethyl Laboratories	Western, ChIP
PAF1	A300-173A	Bethyl Laboratories	ChIP
LEO1	A300-175A	Bethyl Laboratories	Western
LEO1	ab70630	Abcam	Western, ChIP
CTR9	A301-395A	Bethyl Laboratories	Western
CDC73	ab43256	Abcam	Western
SUPT5H	ab26259	Abcam	Western
SUPT5H	sc-28678	Santa Cruz	Western, IP
SUPT6H	NB100-2582	Novus Biologicals	IP
SUPT6H	A300-801A	Bethyl Laboratories	Western
RNA pol II	ab5408	Abcam	Western, IP
AID	Strasbg 9	(Jeevan-Raj et al., 2011)	Western
AID		(Pavri et al., 2010)	ChIP
AID	Ab59361	Abcam	Western
AID	h52-1	(Conticello et al., 2008)	IP
Myc peptide	9E10	CRUK	Western
CD40	HM40-3	eBioscience	CSR stimulation
GFP	Ab290	Abcam	IP
GFP	11814460001	Roche	Western
rabbit IgGs	Ab37415-5	Abcam	ChIP
rabbit IgGs	Ab27478	Santa Cruz	ChIP

## V. Smc5: a new potential regulator of CSR

The multi-approach strategy previously described allowed us to integrate data obtained from the transcriptome analysis of CSR-ID patients and control-derived B cell lines with the proteome of human and mouse B cells. However, our data had to be complemented by genetic analysis on CSR-ID patients in order to restrict – within our list of AID CSR-specific interactors/factors expressed at lower levels in these patients – our interest to those protein(s) whose coding gene(s) was mutated.

We thus focused on Smc5, a member of the structural maintenance of chromosome (Smc) protein family, on the base of a mutation found in one patient analyzed by our collaborators. Additionally, Smc5 was present in our proteome screening, as detected in one of the healthy donors and in the list of proteins interacting with full-length AID expressed in CH12 B cells (Flag-HA-AID<sup>(1-198)</sup>). Smc5 forms an heterodimer with Smc6 and this complex has been described to regulate cell cycle and to mediate DSBs repair through homologous recombination (Kegel and Sjogren, 2010). Based on this evidence, we decided to delineate the role of Smc5 in antibody diversification.



**Figure 27. Smc5 expression is variable in CSR-ID patients**

Nuclear extracts prepared from EBV-immortalized B cell lines obtained from healthy donors (Ctr), AID<sup>-/-</sup> patient (AID<sup>-/-</sup>) and CSR-ID patients (Pat). Extracts were blotted with antibodies specific for Smc5 (A and B), AID (A), and as loading controls Nbs1 (A) and KAP1 (B). Notice that the western blot shown in Figure 27A is performed on the same samples analyzed in Figure 24, thus AID and Nbs1 results correspond to the ones displayed in Figure 24D.

Thus, we assessed Smc5 expression levels in CSR-ID and control-derived cell lines (Figure 27). We observed that Smc5 was expressed at high levels in three out of four healthy donor controls (Ctr #6-8, Figure 27A) whereas variable levels were detected in CSR-ID patients (Pat #1-4, Figure 27A). This result was confirmed by analyzing additional controls (Figure 27B). As for Spt6, the variability in Smc5 protein expression level in CSR-ID-derived B cell lines was puzzling: being unable to predict whether the mutation identified in the patient would affect the stability of the protein and lead to a loss of expression, we silenced Smc5 expression in mouse CH12 B cells in order to investigate its role in CSR (Figure 28). For these experiments switching has been assessed 48h and 72h post-stimulation, as the low stability of the system – monitored through the GFP reporter gene expression by the transduced cell lines – and the partial selection efficiency could impair us from detecting a phenotype at the latest time point applied (72h). We transduced CH12 B cells with shRNAs targeting Smc5, AID and with a non-target control. We assessed Smc5 expression in transduced cell lines prior (Figure

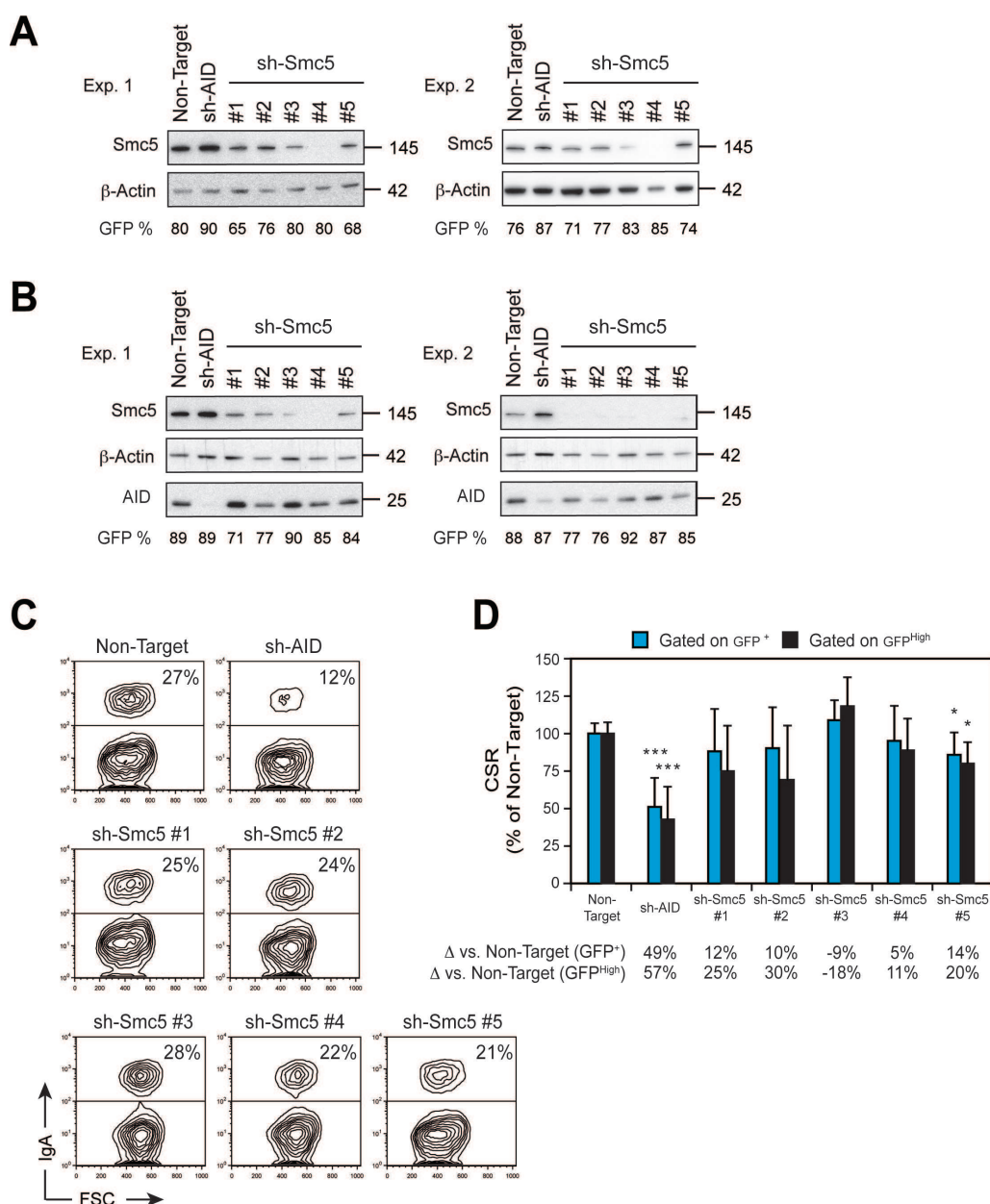


28A, Exp. 1 and 2) and after 48h stimulation (Figure 28B, Exp. 1 and 2). We detected efficient knockdown for one of the shRNAs tested when compared to the non-target control (sh-Smc5 #4, Figures 28A and 28B) and, at a lower extent, reduced Smc5 protein levels for the other four shRNAs used in our experiments. We thus stimulated transduced cells with IL-4, anti-CD40 and TGF $\beta$  for 48h and 72h and measured IgA surface expression by flow cytometry (Figures 28C and 28D).

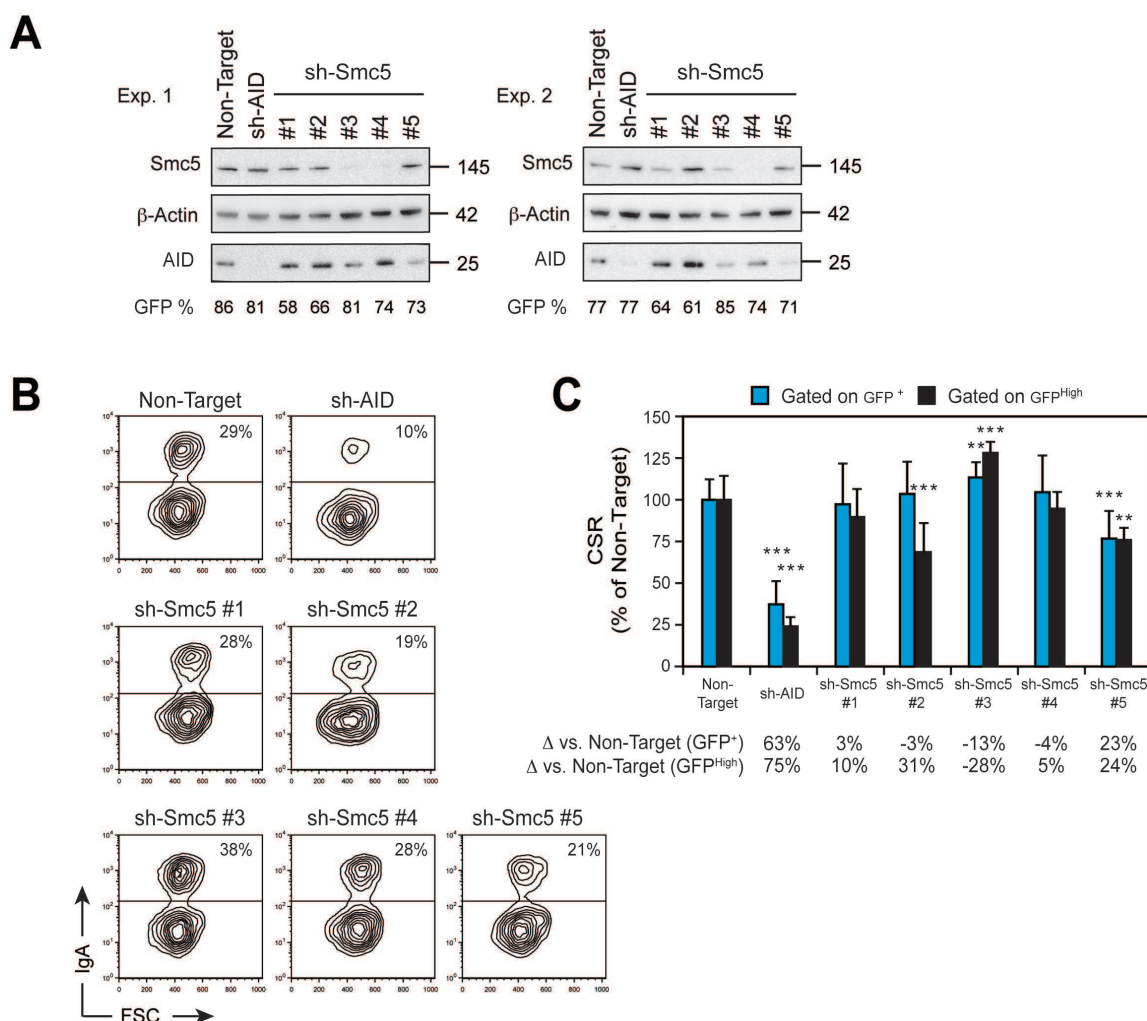
As expected, AID knockdown significantly impaired CSR in cells stimulated for 48h (Figures 28C and 28D), consistent with the barely detectable protein levels (Figure 28B). Interestingly, by looking at CSR efficiency in the total population of cells expressing GFP, we observed that Smc5 knockdown had opposite effects by inducing a reduction ranging from 5 to 14% (Figures 28C and 28D, blue bars) or an increase in switching, which was however not significant (-9%; Figures 28C and 28D, blue bars). Additionally, the optimized conditions for CH12 cells transduction with lentiviruses (reflected by a higher percentage of cells expressing GFP, Figures 28 and 29) led us to enrich the population of cells expressing the GFP reporter gene with cells expressing the reporter at higher levels. By focusing on this population (named GFP<sup>High</sup>; Figure 28D, black bars) we observed a more prominent effect exerted by the different shRNAs targeting either AID or Smc5. In this latter case, the reduction in IgA surface expression, when compared to the non-target control, was ranging from 11% to 30% (Figure 28D, black bars) and, on the other hand, a more efficient switching in cell lines transduced with shRNA #3 (-18%; Figure 28D, black bars) was detected. The effect in CSR was not due to a lower expression of AID, as the transduced cell lines expressed normal or higher protein levels when compared to the non-target control (Figure 28B). It has to be taken into account, instead, that slight fluctuations in AID expression within these cells strongly reflect on their ability to express IgA, as higher switching rates were detected for those cells transduced with Smc5 shRNA #3, which seem to overexpress AID (Figures 28B Exp. 1 and 28D). These results suggest that Smc5 might play a role in CSR as we detected a significant reduction in switching efficiency upon sh-Smc5 #5-mediated knockdown (Figures 28C and 28D).

When we assessed Smc5 expression in transduced cells stimulated for 72h, we still detected an efficient knockdown in cells transduced with shRNA #4 when compared to the non-target control (Figure 29A), whereas a tiny Smc5 re-expression started to be detected in cells transduced with the other shRNAs (Figure 29A, Exp. 2). Moreover, AID silencing was still efficient (Figures 29A and 29B), and we observed a significant reduction in CSR in cells depleted for AID within the GFP<sup>+</sup> population (63% when compared to the non-target; Figures 29B and 29C, blue bars). On the other hand, consistent with the profile described at the earlier time point, Smc5 shRNAs-transduced cells displayed either a significant reduction in switching (23% for sh-Smc5 #5; Figures 29B and 29C, blue bars), or a more efficient recombination which significantly increased the percentage of cells expressing IgA (-13% for sh-Smc5 #3; Figures 29B and 29C, blue bars). This profile was even more pronounced in cells expressing the GFP reporter gene at high levels (GFP<sup>High</sup>; Figure 29C, black bars): the trend observed in cell lines expressing sh-Smc5 #3 and #5 was the same and, additionally, we found a significant impairment in CSR upon sh-Smc5 #2 knockdown, which was not detected in the total population of GFP-expressing cells (Figures 29C and 29D, black bars).

The results we obtained made difficult to conclude about a potential involvement of Smc5 in CSR: if, on one hand, fluctuation in AID expression might explain the significant reduction observed in sh-Smc5 #5 lines, on the other hand this does not seem to be the case in cells transduced with sh-Smc5 #3 which display a more efficient switching (Figure 29A and 29C). Furthermore, the CSR impairment observed upon sh-Smc5 #2 knockdown was not detectable in the total population of cells expressing the GFP reporter gene (Figures 29B and 29C). However, further investigations on cells sorted for GFP expression would clarify the relationship between Smc5 shRNA expression, AID availability and recombination efficiency.



**Figure 28. Lentivirus-mediated *Smc5* knockdown induces a partial CSR reduction upon 48h stimulation**  
**(A)** Western blot for *Smc5* and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced with a lentivirus expressing a GFP reporter and shRNA specific for *Smc5*, *AID* or a non-target shRNA negative control. Two independent experiments are shown. **(B)** Western blot for *Smc5*, *AID* and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced cell lines described in (A) and stimulated with IL-4, anti-CD40 and TGF $\beta$  for 48h. Two independent experiments are shown. **(C)** IgA surface expression as determined by flow cytometry in CH12 cells transduced with lentiviruses described in (A) and stimulated for 48h. Representative plots (gated on GFP<sup>+</sup>) from six independent experiments are shown. Dead cells have been excluded from the analysis by ToPro3 staining. **(D)** Percentage (+ s.d.) of CSR relative to the non-target shRNA control from six independent experiments by gating on cells expressing GFP (GFP<sup>+</sup>; blue bars) or high levels of GFP (GFP<sup>High</sup>; black bars). CSR in cells expressing the non-target shRNA control was set as 100%. The difference in CSR efficiency relative to the non-target control ( $\Delta$ ) is indicated below. Statistical significance vs. the non-target control (two-tailed Student's t-test) is indicated: \*:  $p \leq 0.05$ ; \*\*\*:  $p \leq 0.001$ .



**Figure 29. Lentivirus-mediated *Smc5* knockdown induces opposite effects on CSR upon 72h stimulation**  
**(A)** Western blot for *Smc5*, *AID* and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced with a lentivirus expressing a GFP reporter and shRNA specific for *Smc5*, *AID* or a non-target shRNA negative control and stimulated with IL-4, anti-CD40 and TGF $\beta$  for 72h. **(B)** IgA surface expression as determined by flow cytometry in CH12 cells transduced with lentiviruses described in (A) and stimulated for 72h. Representative plots (gated on GFP<sup>+</sup>) from seven to eleven independent experiments are shown. Dead cells have been excluded from the analysis by ToPro3 staining. **(C)** Percentage (+ s.d.) of CSR relative to the non-target shRNA control from seven to eleven independent experiments by gating on cells expressing GFP (GFP<sup>+</sup>; blue bars) or high levels of GFP (GFP<sup>High</sup>; black bars). CSR in cells expressing the non-target shRNA control was set as 100%. The difference in CSR efficiency relative to the non-target control ( $\Delta$ ) is indicated below. Statistical significance vs. the non-target control (two-tailed Student's t-test) is indicated: \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ .

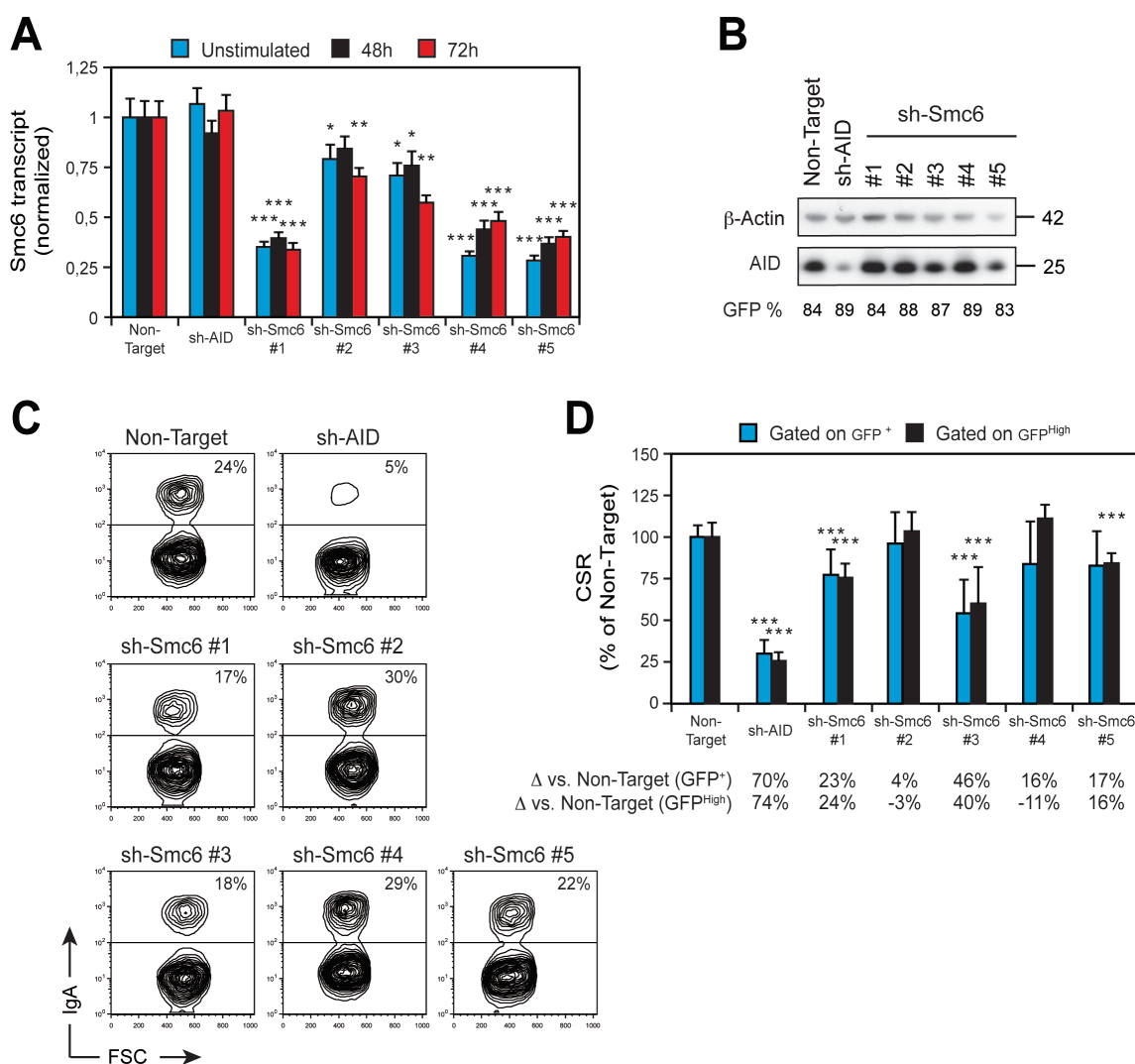
## VI. Smc5/6 complex: is Smc6 required for CSR?

The attempt to investigate a potential role for Smc5 in switching regulation in light of the mutation found in one of the CSR-ID patients, and the identification of Smc6, which is in a complex with Smc5, as co-immunoprecipitating with AID in CH12 B cells overexpressing the deaminase (Flag-HA-AID<sup>(1-198)</sup>) let us wonder whether Smc6 depletion could have any impact on class switch recombination. We thus transduced CH12 cells with lentiviruses expressing shRNAs for Smc6, AID and a non-target control and assessed Smc6 knockdown by RT-qPCR prior and after 48h or 72h stimulation (Figure 30A). While AID depletion did not affect Smc6 expression, Smc6 knockdown significantly reduced Smc6 transcripts levels in cells before stimulation (Figure 30A, blue bars) or after stimulation (Figure 30A, black and red bars).

When we assessed switching efficiency in cells stimulated for 48h, we observed that AID knockdown significantly reduced CSR, ranging from 70% to 74% in cells GFP<sup>+</sup> or GFP<sup>High</sup> respectively (Figures 30C and 30D, blue and black bars), consistent with the efficient depletion of AID shRNA-mediated in transduced cells as detected by western blot (Figure 30B). Smc6 silencing, instead, gave rise to a defect ranging from 23% to 46% in the total population of transduced cells expressing the GFP reporter gene (GFP<sup>+</sup>; Figures 30C and 30D, blue bars). Interestingly, we observed a higher reduction in CSR than what was observed after Smc5 knockdown while comparing cells expressing GFP at high levels to the total GFP<sup>+</sup> population. As AID expression in Smc6 depleted lines is comparable to the one of the non-target control (Figure 30B), we conclude that, 48h post-stimulation, Smc6 depletion has an impact on class switch recombination.

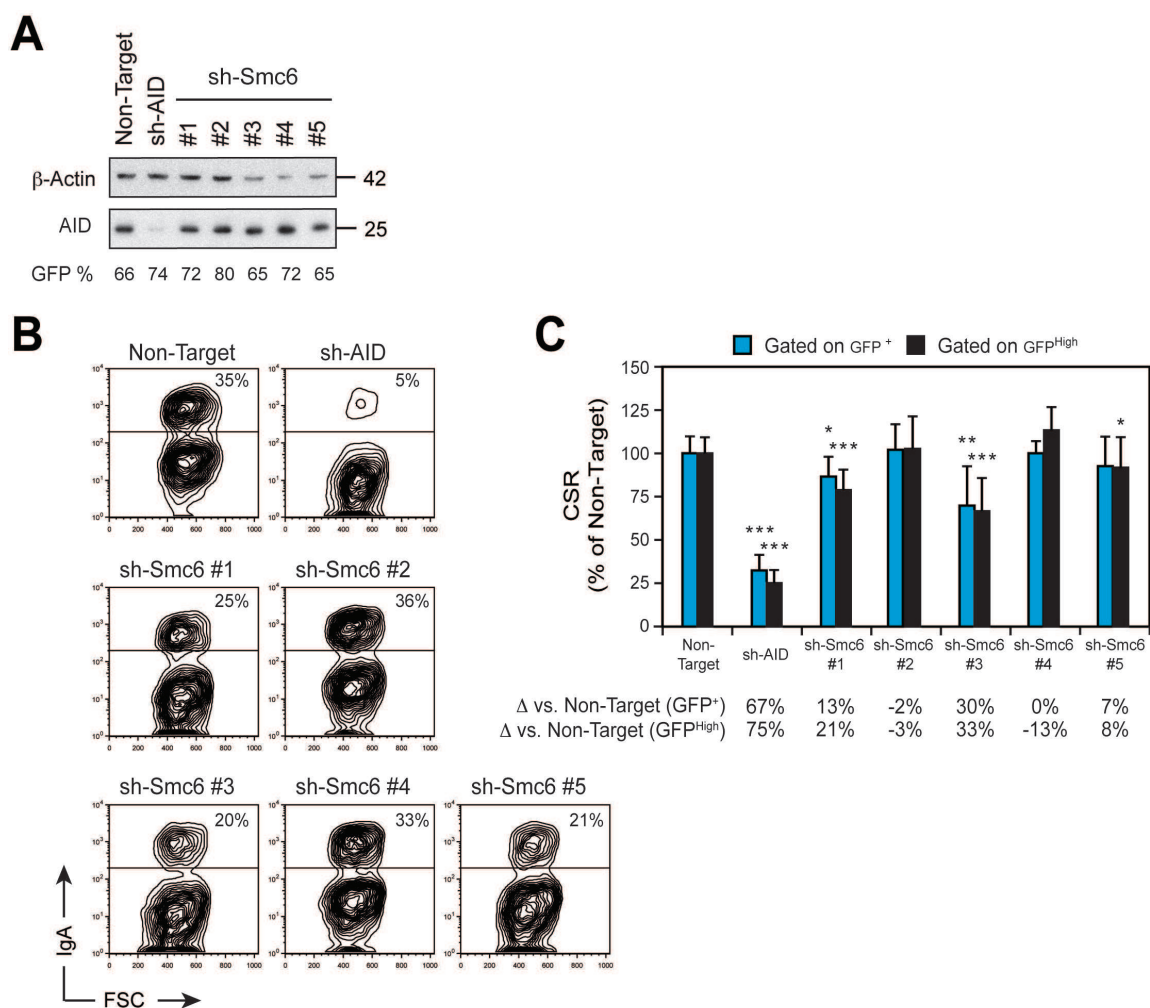
When cells stimulated for 72h were analyzed, AID expression was still robust in cells transduced with shRNAs targeting Smc6 when compared to the non-target control and to the AID knockdown line, where the protein was barely detectable (Figure 31A). AID depletion induced a persistent CSR impairment (67% and 75% in cells GFP<sup>+</sup> and GFP<sup>High</sup>; Figure 31B and 31C, blue and black bars respectively) while Smc6 knockdown lines displayed a less pronounced phenotype when compared to the earlier time point (Figures 31B and 31C and Figure 30D). The difference with the non-target transduced lines was ranging from 13% to 30% in the GFP<sup>+</sup> population (Figure 31C, blue bars) and from 8% to 33% in cells expressing high levels of GFP (Figures 31C, black bars). Thus our results show that, according to the RNA levels detected for Smc6 in transduced cell lines, low levels of Smc6 lead to an impairment in class switch recombination, which was more pronounced 48h post-stimulation compared to 72h, suggesting that Smc6 might play a role in regulating CSR.

Whereas Smc5/6 characterization in class switching regulation will need further investigations, the proteome screening performed on mouse B cells allowed the identification of the cohesin complex, and in the next section I will present my contribution to its characterization in Ig diversifications and the results we obtained with our investigations.



**Figure 30. Lentivirus-mediated *Smc6* knockdown has an impact on CSR upon 48h stimulation**

(A) RT-qPCR for *Smc6* transcripts from CH12 cells transduced with a lentivirus expressing a GFP reporter and shRNA specific for *Smc6*, AID or a non-target shRNA negative control and stimulated with IL-4, anti-CD40 and TGF $\beta$  for the indicated time points are shown. Expression is normalized to HPRT and is presented relative to the non-target control, set as 1. Statistical significance vs. the non-target control (two-tailed Student's *t*-test) is indicated: \*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ . Data are representative from five to seven independent experiments. (B) Western blot for AID and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced with a lentivirus expressing a GFP reporter and shRNA specific for *Smc6*, AID or a non-target shRNA negative control and stimulated with IL-4, anti-CD40 and TGF $\beta$  for 48h. (C) IgA surface expression as determined by flow cytometry in CH12 cells transduced with lentiviruses described in (B) and stimulated for 48h. Representative plots (gated on GFP<sup>+</sup>) from five to nine independent experiments are shown. Dead cells have been excluded from the analysis by ToPro3 staining. (D) Percentage (+ s.d.) of CSR relative to the non-target shRNA control from five to nine independent experiments by gating on cells expressing GFP (GFP<sup>+</sup>; blue bars) or high levels of GFP (GFP<sup>High</sup>; black bars). CSR in cells expressing the non-target shRNA control was set as 100%. The difference in CSR efficiency relative to the non-target control ( $\Delta$ ) is indicated below. Statistical significance vs. the non-target control (two-tailed Student's *t*-test) is indicated: \*\*\*:  $p \leq 0.001$ .



**Figure 31. Lentivirus-mediated *Smc6* knockdown affects CSR upon 72h stimulation**

(A) Western blot for AID and  $\beta$ -actin as loading control on total extracts obtained from CH12 cells transduced with a lentivirus expressing a GFP reporter and shRNA specific for *Smc6*, AID or a non-target shRNA negative control and stimulated with IL-4, anti-CD40 and TGF $\beta$  for 72h. (B) IgA surface expression as determined by flow cytometry in CH12 cells transduced with lentiviruses described in (A) and stimulated for 72h. Representative plots (gated on GFP<sup>+</sup>) from five to nine independent experiments are shown. Dead cells have been excluded from the analysis by ToPro3 staining. (C) Percentage (+ s.d.) of CSR relative to the non-target shRNA control from five to nine independent experiments by gating on cells expressing GFP (GFP<sup>+</sup>; blue bars) or high levels of GFP (GFP<sup>High</sup>; black bars). CSR in cells expressing the non-target shRNA control was set as 100%. The difference in CSR efficiency relative to the non-target control ( $\Delta$ ) is indicated below. Statistical significance vs. the non-target control (two-tailed Student's t-test) is indicated: \*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ .

## VII. Role of the cohesin complex in CSR

The proteome analysis we performed on CH12 B cells overexpressing tagged AID has proven to be a powerful tool in the identification of AID interactors and, in addition to the Smc5/6 complex, we found two additional Smc complexes: the condensins (Smc2/4) and the cohesins (Smc1/3). The latter has been the object of extensive investigations in our lab, and my work contributed to the results achieved and is included in the following manuscript.

The cohesin complex has been initially described as regulator of sister chromatids cohesion during cell division, as involved in homologous recombination and has been proposed to control gene expression (Feeney et al., 2010). Strikingly, cohesins have been shown to mediate the long-range interactions occurring at the Ig and TCR loci in early B and T cell development, respectively (Degner et al., 2011; Seitan et al., 2011). It has been thus proposed that cohesins might play a role in favoring the interaction between enhancers and promoters, as they bind to sites occupied by CTCF as well as colocalize with the Mediator complex (Kagey et al., 2010). In light of this evidence, and considering the structure of the IgH locus and the basic requirements for an efficient switching reaction - the juxtaposition of the donor and acceptor S regions which promotes recombination of AID-mediated DSBs previously generated - we wondered whether the cohesin complex could be involved in CSR regulation.

By focusing on Smc1 and Smc3, the core components of the cohesin complex, and on the loading and unloading factors into the DNA, Nipbl and Wapal, we confirmed that they indeed exist in a complex with AID, in the nucleus and bind to chromatin. We thus investigated the role of cohesins in the IgH locus 3D rearrangement by performing ChIP-Seq experiments on resting and stimulated mouse B cells. We found that Smc1 and Smc3 colocalize with CTCF at the 3'RR and at the C $\alpha$  region in resting B cells, whereas, upon activation, they are actively recruited to the S $\mu$ -C $\mu$  region, in a CTCF-independent manner. These results show for the first time the active recruitment of cohesin in mature B cells undergoing antibody diversification.

The recruitment of Smc1 and Smc3 to the donor S region suggested that they might be required for efficient CSR, thus we silenced *Smc1*, *Smc3*, *Nipbl* and *Wapal* gene expression in CH12 B cells by using recombinant lentiviruses. We observed an impairment in CSR in the transduced cell lines, which was not due to a slightly lower proliferation rate observed in Smc1, Smc3 and Nipbl shRNA-transduced lines nor to an altered cell cycle. As the knockdown did not affect AID expression nor transcription at the donor and acceptor S regions, S $\mu$  and S $\alpha$ , our data suggest that cohesins play a role in CSR, which is independent on AID availability for DNA deamination or germline transcription at the IgH locus. I have contributed to optimize the experimental conditions which allowed us to silence cohesins gene expression in CH12 cells, I evaluated the consequences of their loss in CSR as well as the impact of the knockdown on cell proliferation and progression through the cell cycle. Furthermore, I have been able to apply the improvements we made in terms of transduction, culture and analysis of the cells lines generated to address the role of Smc5/6 complex in antibody diversification.



Finally, the last step of CSR is represented by the joining of the DSBs generated at the donor and acceptor S regions through efficient recombination. As the cohesin complex has been involved in homologous recombination during meiosis, we assessed whether cohesin depletion had an impact on DNA repair. We cloned and sequenced S junctions in stimulated CH12 B cells previously transduced with shRNAs for Smc1, Smc3, Nipbl and Wapal and we observed, upon Smc1, Smc3 and Nipbl knockdown, a significant bias in the microhomology usage with the preference toward longer microhomologies (hallmark of the A-NHEJ pathway) instead of the short ones, usually mediated by proteins involved in the C-NHEJ pathway.

Our findings suggest that the cohesin complex plays an active role in regulating switching efficiency, which could be exerted by mediating the interaction between the donor S region and the regulatory region at the 3' of the IgH locus. Furthermore, cohesin recruitment might affect the choice of the pathway involved in the resolution of DSBs, and we speculate that the balance between the IgH locus structural reorganization as well as the dynamic interaction of the cohesin complex with repair factors allows an efficient recombination.

# Manuscript: The cohesin complex regulates immunoglobulin class switch recombination

Anne-Sophie Thomas-Claudepierre<sup>1\*</sup>, Ebe Schiavo<sup>1\*</sup>, Vincent Heyer<sup>1</sup>, Marjorie Fournier<sup>1</sup>, Adeline Page<sup>1</sup>, Isabelle Robert<sup>1</sup> and Bernardo Reina-San-Martin<sup>1</sup>.

<sup>1</sup> *Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Institut National de la Santé et de la Recherche Médicale (INSERM) U964 / Centre National de la Recherche Scientifique (CNRS) UMR 7104 / Université de Strasbourg, 67404 Illkirch, France.*

*\* Equal contribution*

**Contact:** Bernardo Reina-San-Martin: reinab@igbmc.fr

**Running Title:** Cohesin is required for efficient CSR

*Accepted at the Journal of Experimental Medicine*

## Abstract

Immunoglobulin class switch recombination (CSR) is initiated by the transcription-coupled recruitment of activation induced cytidine deaminase (AID) to switch regions and by the subsequent generation of dsDNA breaks (DSBs). These DNA breaks are ultimately resolved through the non-homologous end joining (NHEJ) pathway. Here we show that during CSR AID associates with subunits of cohesin, a complex previously implicated in sister chromatid cohesion, DNA repair and in the formation of DNA loops between enhancers and promoters. Furthermore, we implicate the cohesin complex in the mechanism of CSR by showing that cohesin is dynamically recruited to the S<sub>μ</sub>-C<sub>μ</sub> region of the IgH locus during CSR and that knockdown of cohesin or its regulatory subunits results in impaired CSR and increased usage of microhomology-based end joining.

## Introduction

During immune responses, B cells diversify their receptors through somatic hypermutation (SHM) and class switch recombination (CSR). SHM introduces mutations in immunoglobulin variable regions that modify the affinity of the receptor for its cognate antigen (Di Noia and Neuberger, 2007). CSR replaces the antibody isotype expressed (from IgM to IgG, IgE or IgA), providing novel antibody effector functions (Chaudhuri et al., 2007). Mechanistically, SHM and CSR are initiated by activation induced cytidine deaminase (AID), an enzyme that deaminates cytosines in both strands of transcribed DNA substrates (Basu et al., 2011; Petersen-Mahrt et al., 2002). AID-induced DNA deamination is then processed to trigger mutations in variable regions during SHM or to generate double stranded DNA break (DSB) intermediates in switch (S) regions during CSR (Chaudhuri et al., 2007; Di Noia and Neuberger, 2007). These breaks activate the DNA damage response (Ramiro et al., 2007) and are resolved through classical and alternative non-homologous end joining (NHEJ) (Stavnezer et al., 2010).

CSR is a transcription-dependent, long-range recombination that occurs at the immunoglobulin heavy chain (IgH) locus and that involves the joining of two S regions, which may be separated by several hundreds of kilobasepairs. For CSR to succeed, donor and acceptor S regions must be brought into close proximity. This is believed to occur through three-dimensional conformational changes involving the generation of transcription-coupled DNA loops (Kenter et al., 2012). Nevertheless, the precise mechanisms controlling these conformational changes remain to be elucidated.

The cohesin complex has been described to play a prominent role in sister chromatid cohesion during cell division, in favoring DNA repair by homologous recombination (Nasmyth and Haering, 2009), in modulating gene expression (Dorsett, 2009) and in promoting the transcription-coupled formation of long-range DNA loop structures (Kagey et al., 2010). In addition, cohesin and the transcriptional insulator CTCF (Dorsett, 2009; Nasmyth and Haering, 2009), have been shown to control the RAG1/2-dependent rearrangement of antigen receptor genes during early B and T lymphocyte development, by mechanisms involving the regulation of transcription and formation of long-range *in cis* DNA interactions (Degner et al., 2011; Guo et al., 2011; Seitan et al., 2011). Here, we have examined the role of cohesin in mature B cells undergoing CSR.

## Results and Discussion

### Nuclear and chromatin-bound AID associate with cohesin

We have previously shown that nuclear AID exists in a large molecular weight complex containing proteins that are required for CSR (Jeevan-Raj et al., 2011). To further characterize this complex and investigate the functional role of novel AID partners in CSR we have carried out additional co-immunoprecipitation experiments coupled to identification by mass spectrometry. Nuclear and chromatin extracts prepared from CH12 cells expressing a full-length N-terminally tagged AID protein (AID<sup>Flag-HA</sup>) or the epitope tags alone (Flag-HA) as negative controls were immunoprecipitated using an anti-Flag antibody. Eluted proteins were submitted for identification by mass spectrometry. Among the proteins identified, we found multiple AID partners previously implicated in CSR and/or SHM (Table S1). In addition, we found several proteins with no known function in CSR (Table 1), including subunits of the cohesin, condensin, Smc5/6 complex and Ino80 complexes. Given the described role for cohesin in mediating long-range recombination during B cell and T cell differentiation, we focused on the potential role of cohesin in CSR. The association between AID and the cohesin complex subunits (Smc1, Smc3, Nipbl and Wapal) was confirmed by reciprocal co-immunoprecipitations and western blotting in the nuclear (Figure 1A) and chromatin fractions (Figure 1B) and was specific, as they did not co-precipitate with an irrelevant tagged protein (EGFP<sup>Flag-HA</sup>; Figure 1C). Importantly, these interactions were not mediated by non-specific nucleic acid binding, as extracts and immunoprecipitations were done in the presence of the benzonase nuclease. We conclude that endogenous subunits of the cohesin complex associate with a fraction of nuclear and chromatin-bound tagged AID through interactions that do not involve non-specific nucleic acid binding.

### Smc1 and Smc3 are dynamically recruited to the IgH locus during CSR

To determine whether cohesin is recruited to the IgH locus in B cells undergoing CSR, we performed ChIP-Seq experiments on chromatin prepared from resting or activated splenic B cells isolated from wild-type mice and using antibodies specific for Smc1, Smc3 and CTCF (Figure 2). In resting B cells, we found that Smc1, Smc3 and CTCF are co-recruited to the 3' regulatory region (3'RR; Figure 2A). This is consistent with published ChIP data on CTCF (Chatterjee et al., 2011) in mature B cells and ChIP-Seq results for CTCF and cohesin (Rad21) in Rag1-deficient pro-B cells (Degner et al., 2011). A sharp peak of CTCF, Smc1 and Smc3 binding was observed at C $\alpha$ . This peak occurred over a region containing a predicted DNaseI hypersensitive site and a CTCF consensus motif (Nakahashi et al., 2013). No significant enrichment was observed at the E $\mu$  enhancer, S $\mu$  or S $\gamma$ 1 (Figure 2A). After stimulation, under conditions that induce CSR to IgG1, we found that Smc1 and Smc3 are significantly co-recruited, independently of CTCF, to a region spanning from the 5' end of the donor switch region (S $\mu$ ) to the 3' end of the C $\mu$  constant region that did not comprise the E $\mu$  enhancer (Figure 2B). Surprisingly, we failed to detect a reproducible recruitment of Smc1 or Smc3 over the S $\gamma$ 1 switch region (Figure 2B), suggesting that Smc1 and Smc3 are not recruited to the acceptor switch region upon activation. It is possible however, that our cell culture conditions (in which approximately 15-20%

of the cells switch to IgG1) are not robust enough to detect a specific enrichment. Consistent with this, we were unable to reproducibly detect a specific enrichment of AID at  $S_{\gamma 1}$  by ChIP-qPCR (Figure 2E).

The ChIP-Seq signal obtained in resting and activated B cells for Smc1 and Smc3 (Figure 2A and 2B) is consistent with the fact that they are known to exist as a heterodimer, was reproducible and specific, as we did not observe any significant enrichment at the IgH locus when using an IgG antibody as a negative control (Figure 2A and 2B). The recruitment of Smc1 and Smc3 at the IgH locus only partially correlated with that reported for AID (Yamane et al., 2010) and is consistent with the fact that only a fraction of chromatin-bound AID associates with the cohesin complex (Figure 1B). This suggests that cohesin is not a targeting factor for AID. The recruitment of Smc1, Smc3 and CTCF in resting and activated B cells observed by ChIP-Seq (Figure 2A and B) was confirmed by additional independent analytical-scale ChIP-qPCR experiments, using primer pairs at individual locations across the IgH locus (Figure 2C, 2D). We conclude that Smc1 and Smc3 complex are dynamically recruited, independently of CTCF, to the IgH locus (at the  $S_{\mu}$ - $C_{\mu}$  region) during CSR. As  $E_{\mu}$  is not bound by cohesin in resting B cells, the constitutive long-range interactions between  $E_{\mu}$  and the 3'RR that take place in resting B cells (Wuerffel et al., 2007), are most likely cohesin-independent. Nevertheless, given the dynamic recruitment of Smc1 and Smc3 at  $S_{\mu}$ - $C_{\mu}$  (and possibly  $S_{\gamma 1}$ ) in activated B cells, we speculate that cohesin may play a role in supporting the structural changes occurring at the IgH locus upon B cell activation.

## Cohesin is required for efficient CSR

To determine the functional relevance of the cohesin complex in CSR we undertook knockdown experiments in CH12 cells, a B cell line which can be induced to undergo CSR from IgM to IgA *in vitro* and which allows to study the role of specific factors in CSR (Pavri et al., 2010; Willmann et al., 2012). CH12 cells were transduced with lentiviruses expressing a GFP reporter together with shRNAs specific for AID (as a positive control), the core subunits of the cohesin complex (Smc1 and Smc3), the cohesin loader/unloader subunits (Nipbl and Wapal) and a Non-Target shRNA as a negative control. Knockdown efficiencies were determined by western blot or by quantitative RT-PCR (qRT-PCR) on GFP<sup>+</sup> sorted cells (Figure 3A). Transduced cells were stimulated for 48h, and their ability to undergo CSR to IgA was determined by flow cytometry (Figure 3B and 3C). As expected, knockdown of AID resulted in a robust reduction in the efficiency of CSR relative to the Non-Target shRNA control (Figure 3B and 3C). Interestingly, we found that knockdown of Smc1, Smc3, Nipbl and Wapal resulted in a significant reduction in the efficiency of CSR (18%-41%) in GFP<sup>+</sup> cells (Figure 3B and 3C). This reduction was more pronounced (30%-52%), when the analysis was performed by gating on cells expressing high levels of GFP (Figure 3C).

The effect on CSR after cohesin knockdown was not due to decreased survival (as determined by ToPro-3 staining; unpublished data), strong defects in proliferation (CFSE dilution; Figure S1), significant activation of the DNA damage response and cell cycle checkpoints (western blot for  $\gamma$ -H2AX and p-Chk1, Figure S2A), nor to defective cell cycle progression (flow cytometry, Figure S2B and S2C).

To determine whether switch region transcription is affected by the knockdown of cohesin subunits, we measured the level of donor ( $S_{\mu}$ ) and acceptor ( $S_{\alpha}$ ) switch region transcripts by qRT-PCR in activated CH12

cells. We found that the level of  $S_{\mu}$  and  $S_{\alpha}$  transcripts was increased after knockdown of AID and cohesin (relative to the Non-Target control), with the exception of  $S_{\alpha}$  transcripts after knockdown of Wapal (Figure 3D and 3E), as expected from cells in which CSR is compromised and that continue to transcribe the switch regions. As no significant reduction in the level of these transcripts after Smc1, Smc3 and Nipbl knockdown was observed, we conclude that switch regions continue to be efficiently transcribed and that they are accessible for DNA deamination by AID. Therefore, cohesin appears not to be involved in the transcriptional regulation of switch regions during CSR. Importantly, we excluded a potential reduction in AID expression levels by western blot (Figure 3F). We conclude that the cohesin complex is required for efficient CSR in CH12 B cells. The role of cohesin in CSR appears to be independent of regulating switch region transcription and/or AID accessibility. Concerning a potential more global effect on transcription, we cannot exclude the possibility that the expression of additional genes required for CSR (other than AID) is affected by the knockdown of cohesin.

### **Knockdown of cohesin affects non-homologous end joining**

DSBs triggered by AID in switch regions during CSR are resolved through the NHEJ pathway and the resulting switch junctions display small insertions and short stretches of microhomology (Stavnezer et al., 2010). In the absence of core NHEJ components, an increase in the usage of microhomology is observed concomitantly with a complete loss of direct joining (Yan et al., 2007). To determine whether cohesin knockdown affects the resolution of DSBs generated during CSR we cloned and sequenced  $S_{\mu}/S_{\alpha}$  switch junctions from stimulated CH12 transduced with lentiviruses expressing shRNAs for Smc1, Smc3, Nipbl, Wapal and a Non-Target negative control (Figure 4) and sorted for GFP expression. Sequence analysis (Figure S3) revealed that knockdown of cohesin subunits resulted in a significant increase in the usage of microhomology when compared to the Non-Target control (Figure 4). While the average length of overlap (excluding insertions) was of 1.58 bp for the Non-Target control, it was increased to 3.22 bp for Smc1 ( $p=0.0001$ ), 2.60 bp for Smc3 ( $p=0.0139$ ) and 2.90 bp for Nipbl ( $p=0.0066$ ). The switch junctions obtained after Wapal knockdown displayed an overlap of 2.04 bp that was not statistically different from the Non-Target control ( $p=0.6125$ ). The increase in microhomology was due to sequences bearing more than 7 bp of microhomology at the junction and a reduction in those bearing short insertions (Figure 4), similar to what has been described in human patients with deficiency in DNA ligase IV (Du et al., 2008), Artemis (Du et al., 2008) or ATM (Pan-Hammarstrom et al., 2006). In contrast to deficiency in core NHEJ components (Stavnezer et al., 2010), we did not find a reduction in the frequency of direct joining events (Figure 4). We conclude that switch recombination junctions generated after Smc1, Smc3 and Nipbl knockdown (but not Wapal) are biased towards the usage of longer microhomologies. Given the role of Wapal in releasing cohesin from chromatin (Kueng et al., 2006), this suggests that cohesin is recruited but not released from the IgH locus and that NHEJ proceeds unaffected. Therefore, it appears that the loading of cohesin is sufficient to determine the outcome of DSB repair and that cohesin participates in the resolution of AID-induced DNA breaks.

Increased usage of microhomology at the junctions is reminiscent of what is observed in B cells defective for core components of the NHEJ pathway (Yan et al., 2007). Nevertheless, deficiency in XRCC4 or DNA ligase

IV also results in a complete loss of sequences repaired through a direct joining (Yan et al., 2007). Therefore, it is unlikely that the cohesin complex is *per se* part of the NHEJ machinery. As cohesin has been implicated in the recruitment of 53BP1 to  $\gamma$ -irradiation-induced foci (Watrin and Peters, 2009), and that 53BP1-deficiency leads to defective CSR, increased DNA end resection and preferential usage of microhomology (Bothmer et al., 2010), we speculate that cohesin could participate in the recruitment of 53BP1 to AID-induced DSBs and that defective 53BP1 recruitment could account for the increased usage of microhomology observed.

Overall, our results implicate the cohesin complex in the mechanism of CSR and provide evidence for the involvement of cohesin in regulating the repair of programmed double-stranded DNA breaks.



## Materials and Methods

**Nuclear extracts and co-immunoprecipitation.** Nuclear extracts and chromatin fractions were prepared using standard techniques (in the presence of 100 U/ml of benzonase; Novagen) from CH12F3 cells stably expressing AID<sup>Flag-HA</sup>, EGFP<sup>Flag-HA</sup> or the tags alone (Jeevan-Raj et al., 2011). Co-immunoprecipitations (in the presence of 100 U/ml of benzonase; Novagen) and western blot analysis were performed as described (Jeevan-Raj et al., 2011). See Table S2 for antibodies used.

**Mass spectrometry analysis.** 20 mg of nuclear extract were immunoprecipitated with Flag M2-agarose beads, washed and eluted with Flag peptide as described (Jeevan-Raj et al., 2011). Flag eluates were fractionated by one-dimension electrophoresis and processed as described (Jeevan-Raj et al., 2011) for identification by nanoLC-MS/MS or directly submitted to Multidimensional Protein Identification Technology (MudPIT). MudPIT analyses were performed as previously described (Florens et al., 2006; Washburn et al., 2001). Briefly, protein mixtures were TCA-precipitated, urea-denatured, reduced, alkylated and digested with endoproteinase Lys-C (Roche) followed by modified trypsin digestion (Promega). Peptide mixtures were loaded onto a triphasic 100  $\mu$ m inner diameter fused silica microcapillary column. Loaded columns were placed in-line with a Thermo Scientific Dionex Ultimate 3000 nanoLC and a LTQ Velos linear ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (Thermo Scientific). A fully automated 12-steps MudPIT run was performed as described (Florens et al., 2006), during which each full MS scan (from 300 to 1700 m/z range) was followed by 20 MS/MS events using data-dependent acquisition. Proteins were identified by database searching using SEQUEST (Thermo Scientific) with Proteome Discoverer 1.3 software (Thermo Scientific) against the mouse Swissprot database (2011-02 release). Peptides were filtered with Xcorr versus charge state 1.5-1, 2.5-2, 3-3, 3.2-4 and peptides of at least 7 amino acids in length.

**shRNA-mediated knockdown.** The lentiviral vectors (pLKO.1 and pLKO.1-puro-CMV-TurboGFP) expressing shRNAs specific for AID (TRCN0000112031), Smc1 (TRCN0000109034), Smc3 (TRCN0000109007), Nipbl (TRCN0000124037) and Wapal (TRCN0000177268) or a Non-Target control (SHC002) were obtained from Sigma-Aldrich. The lentiviral vectors were transiently transfected into Lenti-X 293T cells (Clontech) to produce infectious viral particles as described (Willmann et al., 2012). Two days later CH12 cells were spin-infected with viral supernatants supplemented with polybrene (10  $\mu$ g/ml; Sigma-Aldrich). Cells were selected for 5 days with puromycin (1  $\mu$ g/ml) before CSR induction.

**Real time quantitative RT-PCR.** RNA and cDNA were prepared using standard techniques. qPCR was performed in triplicates using the Universal Probe Library (UPL) system (Roche) or SyberGreen (Qiagen) and a LightCycler 480 (Roche). Transcript quantities were calculated relative to standard curves and normalized to  $\beta$ -Actin, CD79b or HPRT mRNA. See Table S3 for primers and probes.

**Cell culture and flow cytometry.** Lentivirally transduced CH12 cells were cultured with IL-4 (5 ng/ml; Sigma-Aldrich), TGF- $\beta$  (1 ng/ml; R&D System), monoclonal anti-CD40 antibody (200 ng/ml; eBioscience)

and puromycin (1  $\mu$ g/ml) and analyzed after 48h - 72h for cell surface expression of IgA by flow cytometry as described (Robert et al., 2009).

For proliferation analysis, transduced cells (lacking a GFP reporter) were labeled with 5  $\mu$ M CFSE (Invitrogen) prior stimulation and analyzed by flow cytometry. For cell cycle analysis, unstimulated or stimulated transduced cells were sorted for GFP expression, fixed in 70% ethanol, incubated with 50  $\mu$ g/ml RNase A (Sigma), stained with 25  $\mu$ g/ml propidium iodide (Sigma) and analyzed for DNA content by flow cytometry. As positive controls cells were treated with 2-10 mM hydroxyurea (HU) or 1-5 ng/ml neocarzinostatin (NCS) for 6h. Resting splenic B cells were isolated from 8-12 week C57BL/6 mice using CD43-microbeads (Miltenyi Biotec) and cultured for 60h with LPS (50  $\mu$ g/ml; Sigma-Aldrich) and IL-4 (5 ng/ml; Sigma-Aldrich) as described (Jeevan-Raj et al., 2011). All animal work was performed under protocols approved by the Direction des Services Vétérinaires du Bas-Rhin, France (Authorization N° 67-343).

**Switch junction analysis.** S $\mu$ -S $\alpha$  switch junctions were amplified using previously described primers (Ehrenstein and Neuberger, 1999; Schrader et al., 2002) and conditions (Robert et al., 2009) from genomic DNA prepared from lentivirally transduced CH12 cells stimulated for 72h and sorted for GFP expression. PCR products were cloned using TOPO-TA cloning kit (Invitrogen) and sequenced using T7 universal primers. Sequence analysis was performed using the CSRTTool software (manuscript in preparation).

**ChIP-Seq.** Resting or activated B cells were crosslinked for 10 min at 37°C with 1% (vol/vol) formaldehyde, followed by quenching with glycine (0.125 M final concentration). Crosslinked samples were then sonicated to obtain DNA fragments 200–500 bp in length using a Covaris sonicator (Covaris). Chromatin (from  $10 \times 10^7$  cells) was precleared with protein A magnetic beads pre-washed with PBS 0.05% Tween, 5% BSA and immunoprecipitated in ChIP dilution buffer (0.06% SDS, 20 mM Tris (pH 8.1), 2 mM EDTA, 160 mM NaCl, 1.045% Triton X-100, 0.05 X PIC) overnight at 4°C with protein A magnetic beads (Invitrogen) coupled to 100  $\mu$ g of Smc1 or Smc3 antibodies and processed according to the Millipore protocol. Crosslinks were reversed for 4h at 65°C in Tris-EDTA buffer with 0.3% (wt/vol) SDS and proteinase K (1 mg/ml). ChIP DNA was extracted with IPure Kit (Diagenode). Libraries were prepared for sequencing following the manufacturer's protocol (Illumina) and sequenced on the Illumina Genome Analyzer IIx as single-end 50 base reads following Illumina's instructions. Image analysis and base calling were performed using the Illumina Pipeline and sequence reads were mapped to reference genome mm9/NCBI37 using Bowtie v0.12.7. Peak calling was performed using MACS (Zhang et al., 2008) with default parameters. Global comparison of samples and clustering analysis were performed using seqMINER (Ye et al., 2011).

**ChIP-qPCR.** Analytical-scale ChIP was performed on chromatin prepared from  $10^7$  (resting or activated) splenic B cells isolated from a pool of 5 mice. qPCR was performed at several locations across the IgH locus using primers listed in Table S3. Results are expressed as percent input and represent the mean of three qPCR technical replicates. Error bars represent the corresponding standard deviation.

**Accession codes.** GEO: ChIP-Seq data for CTCF, Smc1 and Smc3 on resting and activated B cells (GSE43594).

**Acknowledgments.** We thank members of the Reina-San-Martin laboratory for discussions; E. Soutoglou, I. Sumara and M. Nussenzweig for comments on the manuscript; S. Pattabhiraman, M. Mendoza Parra, S. Legras and B. Jost for assistance with ChIP-Seq analyses; V. Chavant and F. Ruffenach for assistance in mass spectrometry analysis; C. Ebel for assistance with cell sorting and A. Gazumyan for advice on lentiviral infections. A.S.T.C. was supported by the Ministère de l'Enseignement Supérieur et de la Recherche, France. E.S. was supported by the IGBMC International PhD Program and the Fondation ARC pour la Recherche sur le Cancer. This work was supported by grants to B.R.S.M. from the Agence Nationale pour la Recherche (ANR-Blanc), the Fondation ARC pour la Recherche sur le Cancer (Programme ARC) and the Institut National de la Santé et de la Recherche Médicale (Avenir-INSERM). The authors have no conflicting financial interests.

**Online supplemental material.** Figure S1: CSR and proliferation analysis by CFSE dilution. Figure S2: Western blot analysis of the DNA damage response and cell cycle checkpoint activation and cell cycle analysis. Figure S3: Examples of switch junction sequence alignments. Table S1: Known AID partner proteins found. Table S2: Antibodies. Table S3: Primers. Table S4: ChIP-qPCR statistical analysis.

## References

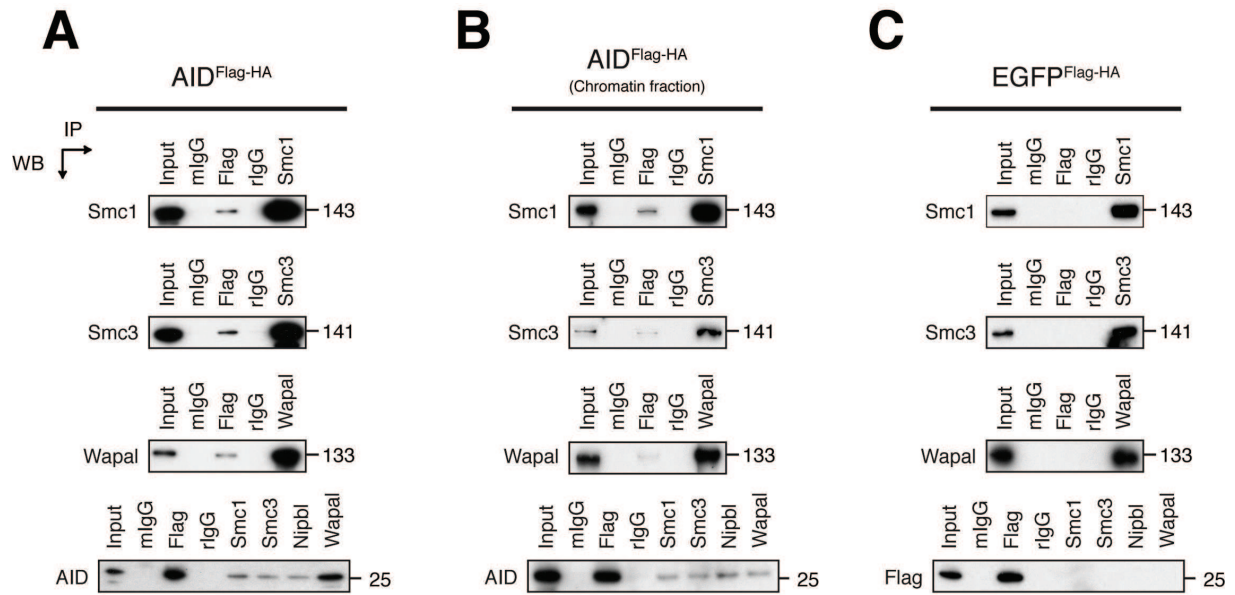
- Basu, U., F.L. Meng, C. Keim, V. Grinstein, E. Pefanis, J. Eccleston, T. Zhang, D. Myers, C.R. Wasserman, D.R. Wesemann, K. Januszyk, R.I. Gregory, H. Deng, C.D. Lima, and F.W. Alt. 2011. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*. 144:353-363.
- Bothmer, A., D.F. Robbiani, N. Feldhahn, A. Gazumyan, A. Nussenzweig, and M.C. Nussenzweig. 2010. 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J Exp Med*. 207:855-865.
- Chatterjee, S., Z. Ju, R. Hassan, S.A. Volpi, A.V. Emelyanov, and B.K. Birshtein. 2011. Dynamic changes in binding of immunoglobulin heavy chain 3' regulatory region to protein factors during class switching. *J Biol Chem*. 286:29303-29312.
- Chaudhuri, J., U. Basu, A. Zarrin, C. Yan, S. Franco, T. Perlot, B. Vuong, J. Wang, R.T. Phan, A. Datta, J. Manis, and F.W. Alt. 2007. Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv Immunol*. 94:157-214.
- Degner, S.C., J. Verma-Gaur, T.P. Wong, C. Bossen, G.M. Iverson, A. Torkamani, C. Vettermann, Y.C. Lin, Z. Ju, D. Schulz, C.S. Murre, B.K. Birshtein, N.J. Schork, M.S. Schlissel, R. Riblet, C. Murre, and A.J. Feeney. 2011. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci U S A*. 108:9566-9571.
- Di Noia, J.M., and M.S. Neuberger. 2007. Molecular Mechanisms of Antibody Somatic Hypermutation. *Annu Rev Biochem*. 76:1-22.
- Dorsett, D. 2009. Cohesin, gene expression and development: lessons from Drosophila. *Chromosome Res*. 17:185-200.
- Du, L., M. van der Burg, S.W. Popov, A. Kotnis, J.J. van Dongen, A.R. Gennery, and Q. Pan-Hammarstrom. 2008. Involvement of Artemis in nonhomologous end-joining during immunoglobulin class switch recombination. *J Exp Med*. 205:3031-3040.
- Ehrenstein, M.R., and M.S. Neuberger. 1999. Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *Embo J*. 18:3484-3490.
- Florens, L., M.J. Carozza, S.K. Swanson, M. Fournier, M.K. Coleman, J.L. Workman, and M.P. Washburn. 2006. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods*. 40:303-311.
- Guo, C., H.S. Yoon, A. Franklin, S. Jain, A. Ebert, H.L. Cheng, E. Hansen, O. Despo, C. Bossen, C. Vettermann, J.G. Bates, N. Richards, D. Myers, H. Patel, M. Gallagher, M.S. Schlissel, C. Murre, M. Busslinger, C.C. Giallourakis, and F.W. Alt. 2011. CTCF-binding elements mediate control of V(D)J recombination. *Nature*. 477:424-430.
- Jeevan-Raj, B.P., I. Robert, V. Heyer, A. Page, J.H. Wang, F. Cammas, F.W. Alt, R. Losson, and B. Reina-San-Martin. 2011. Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *J Exp Med*. 208:1649-1660.
- Kagey, M.H., J.J. Newman, S. Bilodeau, Y. Zhan, D.A. Orlando, N.L. van Berkum, C.C. Ebmeier, J. Goossens, P.B. Rahl, S.S. Levine, D.J. Taatjes, J. Dekker, and R.A. Young. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 467:430-435.
- Kenter, A.L., S. Feldman, R. Wuerffel, I. Achour, L. Wang, and S. Kumar. 2012. Three-dimensional architecture of the IgH locus facilitates class switch recombination. *Ann N Y Acad Sci*. 1267:86-94.
- Kueng, S., B. Hegemann, B.H. Peters, J.J. Lipp, A. Schleiffer, K. Mechtler, and J.M. Peters. 2006. Wapl controls the dynamic association of cohesin with chromatin. *Cell*. 127:955-967.
- Nakahashi, H., K.R. Kwon, W. Resch, L. Vian, M. Dose, D. Stavreva, O. Hakim, N. Pruett, S. Nelson, A. Yamane, J. Qian, W. Dubois, S. Welsh, R.D. Phair, B.F. Pugh, V. Lobanenkov, G.L. Hager, and R. Casellas. 2013. A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Rep*. 3:1678-1689.
- Nasmyth, K., and C.H. Haering. 2009. Cohesin: its roles and mechanisms. *Annu Rev Genet*. 43:525-558.
- Pan-Hammarstrom, Q., A. Lahdesmaki, Y. Zhao, L. Du, Z. Zhao, S. Wen, V.L. Ruiz-Perez, D.K. Dunn-Walters, J.A. Goodship, and L. Hammarstrom. 2006. Disparate roles of ATR and ATM in immunoglobulin class switch recombination and somatic hypermutation. *The Journal of experimental medicine*. 203:99-110.
- Pavri, R., A. Gazumyan, M. Jankovic, M. Di Virgilio, I. Klein, C. Ansarah-Sobrinho, W. Resch, A. Yamane, B. Reina-San-Martin, V. Barreto, T.J. Nieland, D.E. Root, R. Casellas, and M.C. Nussenzweig. 2010. Activation-Induced Cytidine Deaminase Targets DNA at Sites of RNA Polymerase II Stalling by Interaction with Spt5. *Cell*. 143:122-133.
- Petersen-Mahrt, S.K., R.S. Harris, and M.S. Neuberger. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature*. 418:99-103.

- Ramiro, A., B. Reina-San-Martin, K. McBride, M. Jankovic, V. Barreto, A. Nussenzweig, and M.C. Nussenzweig. 2007. The role of activation-induced deaminase in antibody diversification and chromosome translocations. *Adv Immunol.* 94:75-107.
- Robert, I., F. Dantzer, and B. Reina-San-Martin. 2009. Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *J Exp Med.* 206:1047-1056.
- Schrader, C.E., J. Vardo, and J. Stavnezer. 2002. Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class switching shown by sequence analysis of recombination junctions. *J Exp Med.* 195:367-373.
- Seitan, V.C., B. Hao, K. Tachibana-Konwalski, T. Lavagnoli, H. Mira-Bontenbal, K.E. Brown, G. Teng, T. Carroll, A. Terry, K. Horan, H. Marks, D.J. Adams, D.G. Schatz, L. Aragon, A.G. Fisher, M.S. Krangel, K. Nasmyth, and M. Merckenschlager. 2011. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature.* 476:467-471.
- Stavnezer, J., A. Bjorkman, L. Du, A. Cagigi, and Q. Pan-Hammarstrom. 2010. Mapping of switch recombination junctions, a tool for studying DNA repair pathways during immunoglobulin class switching. *Adv Immunol.* 108:45-109.
- Washburn, M.P., D. Wolters, and J.R. Yates, 3rd. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* 19:242-247.
- Watrin, E., and J.M. Peters. 2009. The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *Embo J.* 28:2625-2635.
- Willmann, K.L., S. Milosevic, S. Pauklin, K.M. Schmitz, G. Rangam, M.T. Simon, S. Maslen, M. Skehel, I. Robert, V. Heyer, E. Schiavo, B. Reina-San-Martin, and S.K. Petersen-Mahrt. 2012. A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *The Journal of experimental medicine.* 209:2099-2111.
- Wuerffel, R., L. Wang, F. Grigera, J. Manis, E. Selsing, T. Perlot, F.W. Alt, M. Cogne, E. Pinaud, and A.L. Kenter. 2007. S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity.* 27:711-722.
- Yamane, A., W. Resch, N. Kuo, S. Kuchen, Z. Li, H.W. Sun, D.F. Robbiani, K. McBride, M.C. Nussenzweig, and R. Casellas. 2010. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol.* 12:62-69.
- Yan, C.T., C. Boboila, E.K. Souza, S. Franco, T.R. Hickernell, M. Murphy, S. Gumaste, M. Geyer, A.A. Zarrin, J.P. Manis, K. Rajewsky, and F.W. Alt. 2007. IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature.* 449:478-482.
- Ye, T., A.R. Krebs, M.A. Choukrallah, C. Keime, F. Plewniak, I. Davidson, and L. Tora. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* 39:e35.
- Zhang, Y., T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9:R137.

**Table 1. Novel AID partner proteins identified by Flag co-immunoprecipitation coupled to nanoLC-MS/MS and MudPIT analysis**

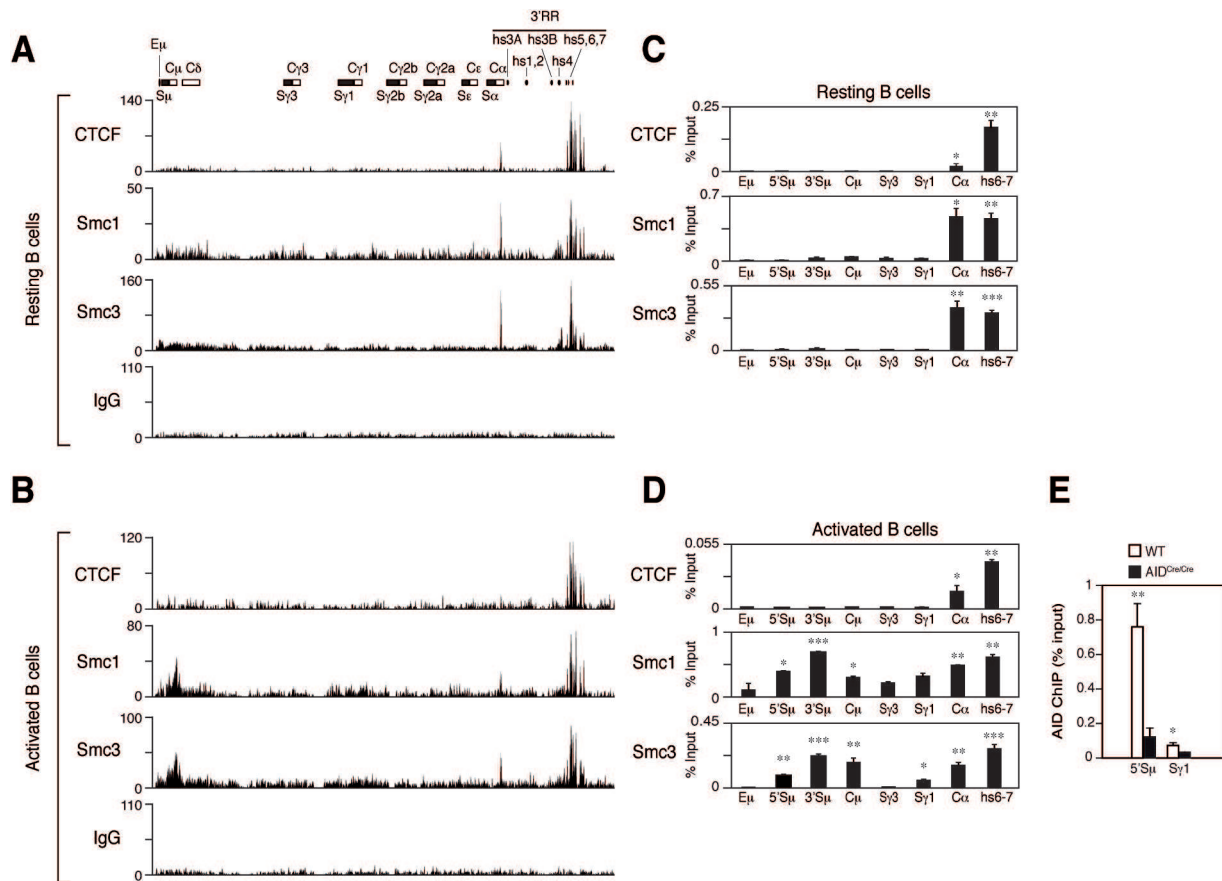
Gene Name (GN), protein description, molecular weight in kilodaltons (MW) and corresponding number of peptides (Pep) and spectral counts (SCs) found are shown on Flag immunoprecipitations conducted on extracts prepared from CH12 cells expressing AID<sup>Flag-HA</sup> or the epitope tags alone (Flag-HA), as a negative control.

GN	Description	MW	Flag-HA		AID <sup>Flag-HA</sup>	
			Pep	SCs	Pep	SCs
<i>Aicda</i>	Activation-induced cytidine deaminase	24	0	0	17	981
<b>Cohesin complex</b>						
<i>Wapal</i>	Wings apart-like protein homolog	134	3	5	37	254
<i>Nipbl</i>	Nipped-B-like protein	315.3	4	8	11	92
<i>Pds5a</i>	Sister chromatid cohesion protein PDS5 homolog A	150.2	0	0	8	12
<i>Smc1a</i>	Structural maintenance of chromosomes protein 1A	143.1	1	1	9	11
<i>Pds5b</i>	Sister chromatid cohesion protein PDS5 homolog B	164.3	2	2	6	8
<i>Smc3</i>	Structural maintenance of chromosomes protein 3	141.5	1	1	5	6
<i>Stag2</i>	Cohesin subunit SA-2	141.3	2	2	3	6
<i>Stag3</i>	Cohesin subunit SA-3	141.1	1	1	2	4
<b>Condensin complex</b>						
<i>Smc2</i>	Structural maintenance of chromosomes protein 2	134.2	2	2	13	19
<i>Smc4</i>	Structural maintenance of chromosomes protein 4	146.8	2	2	8	14
<i>Ncapd2</i>	Condensin complex subunit 1	155.6	2	2	5	9
<i>Ncapd3</i>	Condensin-2 complex subunit D3	169.3	2	3	5	6
<i>Ncapg2</i>	Condensin-2 complex subunit G2	130.8	0	0	3	4
<i>Ncaph</i>	Condensin complex subunit 2	82.3	1	1	2	2
<b>Smc5/6 complex</b>						
<i>Smc5</i>	Structural maintenance of chromosomes protein 5	128.7	3	3	2	5
<i>Smc6</i>	Structural maintenance of chromosomes protein 6	127.1	1	1	3	3
<b>Ino80 complex</b>						
<i>Ruvbl1</i>	RuvB-like 1	50.2	5	5	14	172
<i>Ruvbl2</i>	RuvB-like 2	51.1	4	9	16	64
<i>Yy1</i>	Transcriptional repressor protein YY1	44.7	2	3	8	15
<i>Ino80</i>	Putative DNA helicase INO80 complex homolog 1	176.4	1	1	4	5
<i>Ino80b</i>	INO80 complex subunit B	40.5	1	1	1	1
<i>Ino80c</i>	INO80 complex subunit C	20.4	0	0	1	1



**Figure 1. Nuclear AID associates with cohesin subunits**

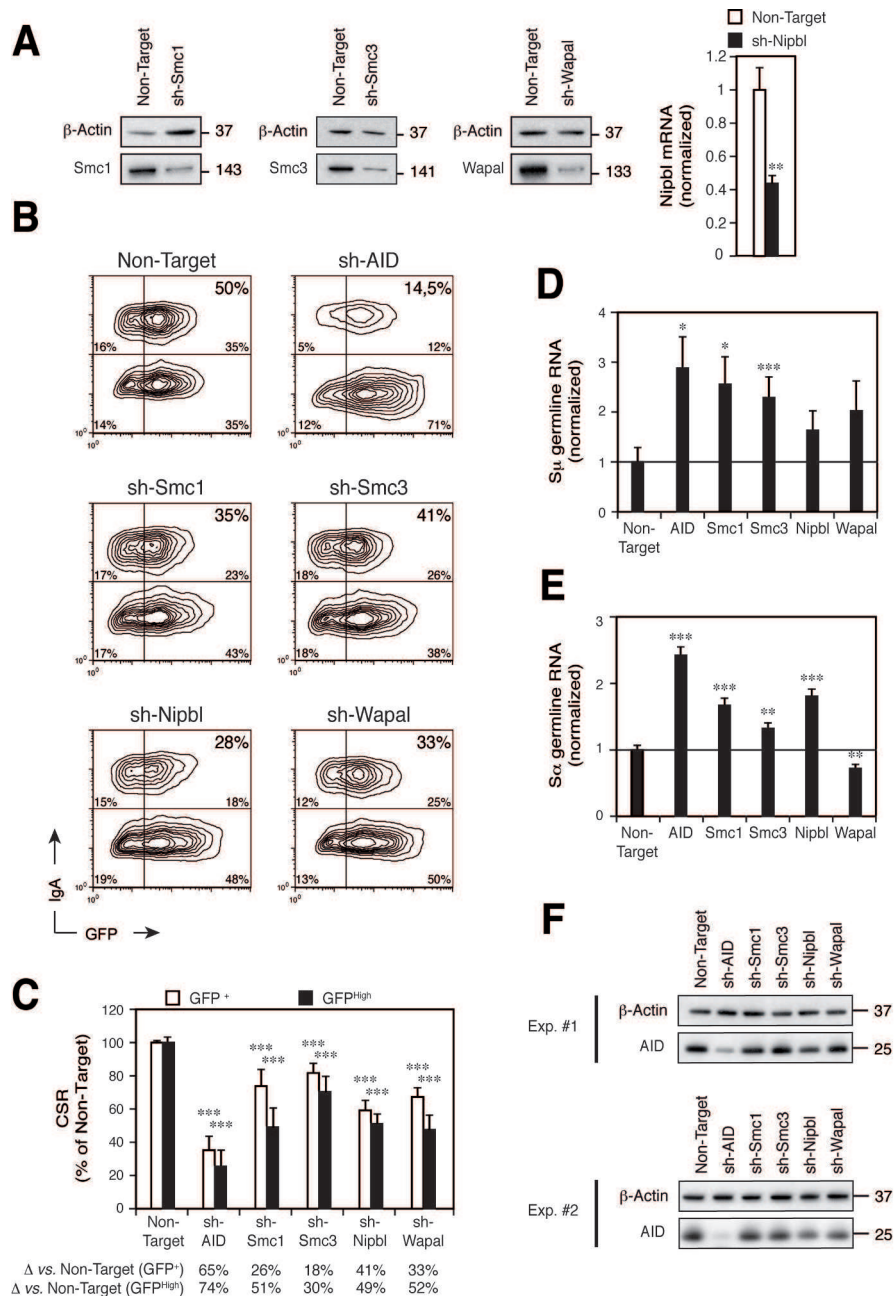
Nuclear extracts (**A and C**) and chromatin fractions (**B**) prepared from CH12 cells expressing (**A and B**) AID<sup>Flag-HA</sup> or (**C**) EGFP<sup>Flag-HA</sup> were immunoprecipitated and blotted with antibodies specific for Flag, AID, Smc1, Smc3, Wapal and Nipbl. Note that the Nipbl antibody works only on immunoprecipitation. Input represents 1% of material used. Note also that only a fraction of AID associates with cohesin subunits. Theoretical molecular weights in kilodaltons (kDa) are indicated. Data are representative of three independent experiments.



**Figure 2. Smc1 and Smc3 are dynamically recruited to the IgH locus during CSR**

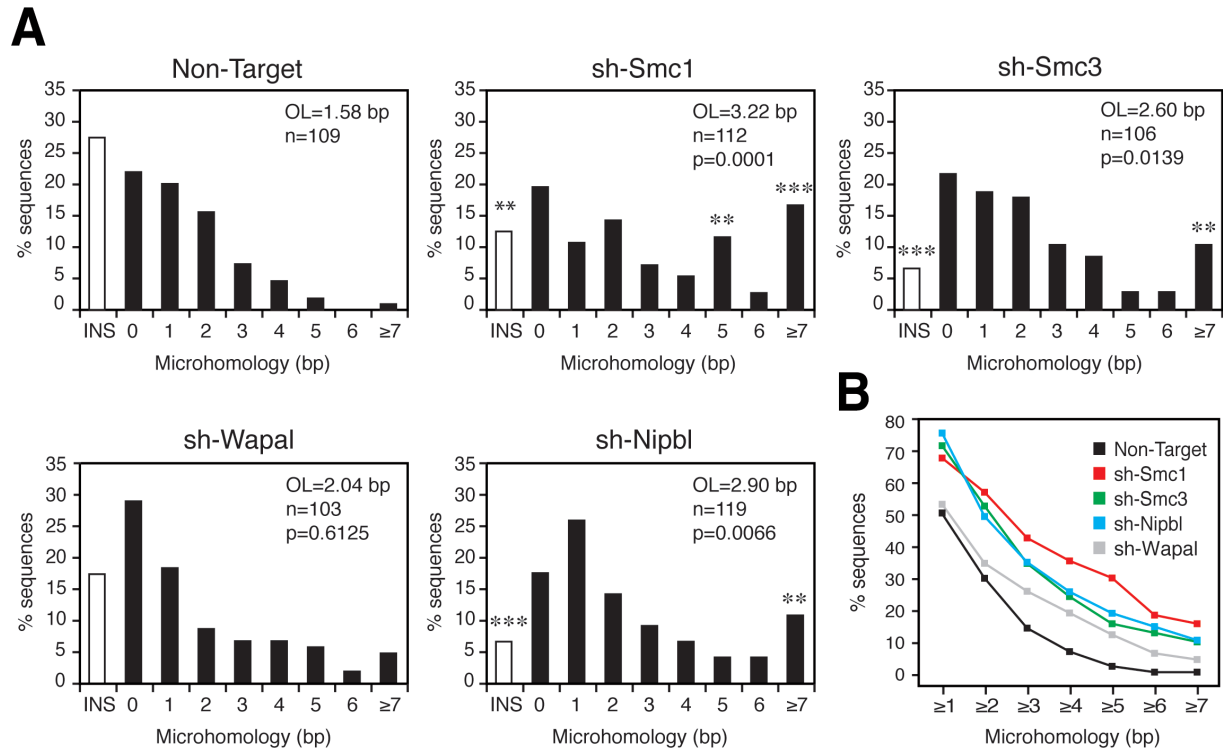
UCSC genome browser screenshots showing the ChIP-Seq binding profiles of CTCF, Smc1, Smc3 and IgG (negative control) at the IgH locus (chr12:114,438,857-114,669,149) in (A) resting and (B) activated (with LPS + IL-4) B cells isolated from wild-type mice. A schematic map of the IgH locus indicates the switch regions (black boxes), the constant region exons (white boxes), the  $E_{\mu}$  enhancer and the DNaseI hypersensitive sites (hs) located in the 3' regulatory region (3'RR). Similar ChIP-Seq profiles were observed in an additional biological replicate experiment for Smc3 that was conducted in resting and activated B cells (data not shown). ChIP-Seq results were verified by analytical-scale ChIP-qPCR experiments in (C) resting and (D) activated B cells. Results are expressed as % input and are representative of two independent biological replicate experiments. Statistical significance vs.  $S_{\gamma 3}$  (two-tailed Student's t-test) is indicated: \*,  $p \leq 0.05$ ; \*\*,  $p \leq 0.01$ ; \*\*\*,  $p \leq 0.001$ . Additional statistical analyses across the locus and between resting and activated B cells are shown in Table S4. (E) ChIP analysis for AID occupancy at the  $S_{\mu}$  and  $S_{\gamma 1}$  switch regions in wild-type and  $AID^{Cre/Cre}$  B cells cultured in vitro with LPS+IL-4 for 60h. Results are expressed as % input. Statistical significance vs.  $AID^{Cre/Cre}$  was determined by a two-tailed Student's t-test. \*,  $p \leq 0.05$ ; \*\*,  $p \leq 0.01$ . Results are representative of four independent experiments.





**Figure 3. CSR is impaired by the knockdown of cohesin subunits**

(A) Western blot for  $\beta$ -Actin, Smc1, Smc3 and Wapal and qRT-PCR for Nipbl transcripts are shown. Expression is normalized to Cd79b and is presented relative to the Non-Target control, set as 1. Statistical significance vs. the Non-Target control (two-tailed Student's *t*-test):  $p=0.0023$ . Data are representative of three experiments. (B) IgA surface and GFP expression as determined by flow cytometry in stimulated CH12 cells transduced (or not) with a lentivirus expressing a GFP reporter and shRNAs specific for AID, Smc1, Smc3, Nipbl, Wapal or a Non-Target shRNA negative control. Representative plots from four to eight independent experiments are shown. Percentage of cells in each quadrant is indicated. The percentage of IgA<sup>+</sup> cells in the GFP<sup>+</sup> population is indicated in the upper right quadrant. (C) Percentage (+ s.d.) of CSR relative to the Non-Target shRNA control from four to eight independent experiments by gating on cells expressing GFP (GFP<sup>+</sup>; white bars) or high levels of GFP (GFP<sup>High</sup>; black bars). CSR in cells expressing the Non-Target shRNA control was set to 100%. The difference in CSR efficiency relative to the Non-Target control ( $\Delta$ ) is indicated below. Statistical significance vs. the Non-Target control (two-tailed Student's *t*-test) is indicated: \*\*\*:  $p \leq 0.001$ . qRT-PCR for  $\mu$  (D) and  $\alpha$  (E) germline transcripts in transduced cells stimulated for 48h. Expression was normalized to HPRT mRNA abundance and is presented relative to the Non-Target control, set as 1 (black line). Statistical significance vs. the Non-Target control (two-tailed Student's *t*-test) is indicated: \*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ . (F) Western blot for  $\beta$ -Actin and AID are shown. Data are representative of three independent experiments performed on transduced and activated cells sorted for GFP expression.



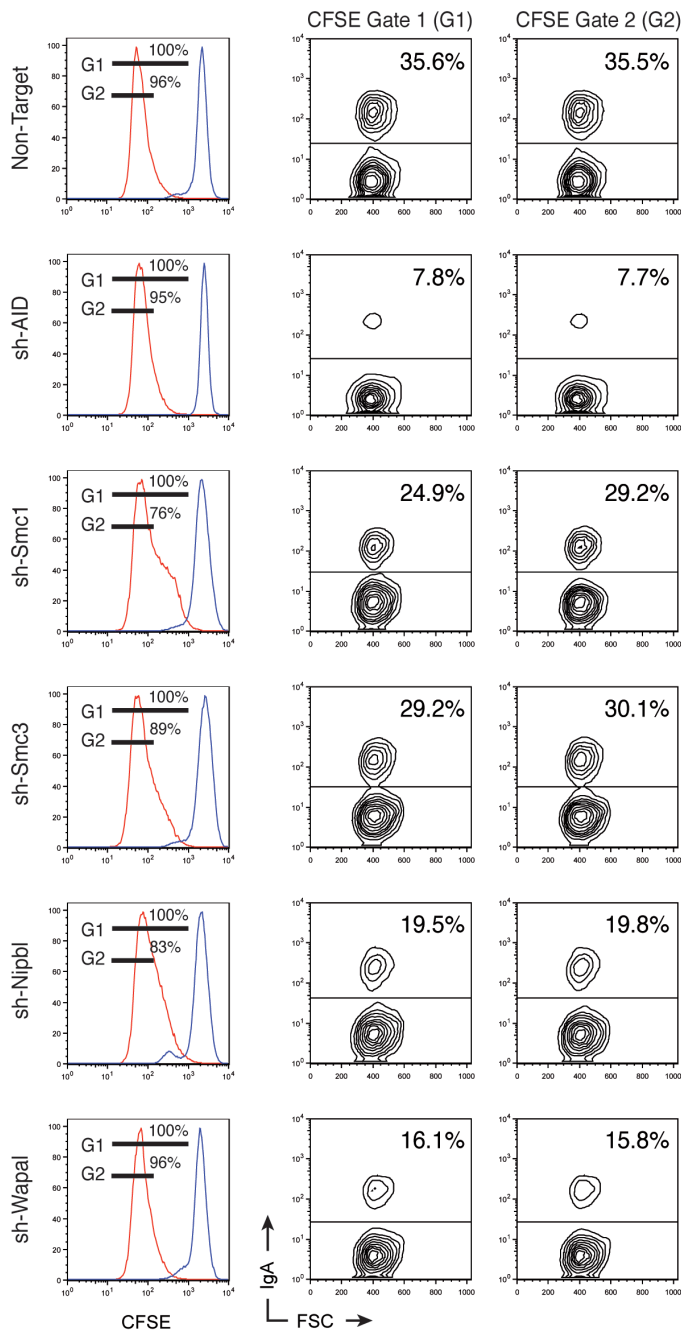
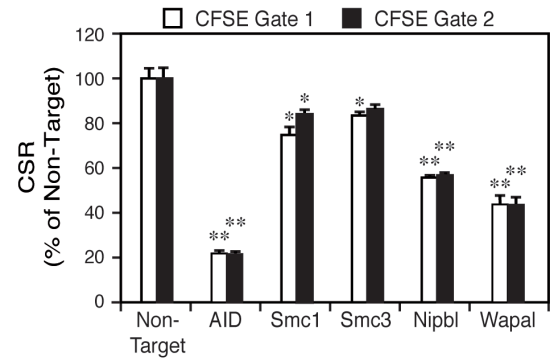
**Figure 4. Knockdown of cohesin affects non-homologous end joining**

**(A)** Histograms showing the percentage of switch junction sequences with indicated nucleotide overlap and obtained from CH12 cells transduced with lentiviruses expressing shRNAs specific for *Smc1*, *Smc3*, *Nipbl*, *Wapal* or a Non-Target negative control and sorted for GFP expression. Number of junctions analyzed ( $n$ ), average length of overlap (OL) and  $p$  values relative to the Non-Target control (Mann-Whitney test) are indicated. White bars indicate the percentage of sequences with small (1-4 nucleotide) insertions. Overlap was determined by identifying the longest region of perfect uninterrupted donor/acceptor identity. Sequences with insertions were not included in the calculation of the average length of overlap. Significant differences relative to the Non-Target control ( $\chi^2$  test) are indicated: \*\*:  $p \leq 0.01$ , \*\*\*:  $p \leq 0.0001$ . Examples of switch junction alignments are shown in Figure S3. **(B)** Line chart showing the cumulative percentage of sequences with a given length of microhomology (bp) and obtained from CH12 cells transduced with lentiviruses expressing shRNAs specific for *Smc1* (red squares), *Smc3* (green squares), *Nipbl* (blue squares), *Wapal* (grey squares) or a Non-Target negative control (black squares) and sorted for GFP expression.

**Table S1. Proteins previously described to associate with AID identified by Flag co-immunoprecipitation coupled to nanoLC-MS/MS and MudPIT analysis**

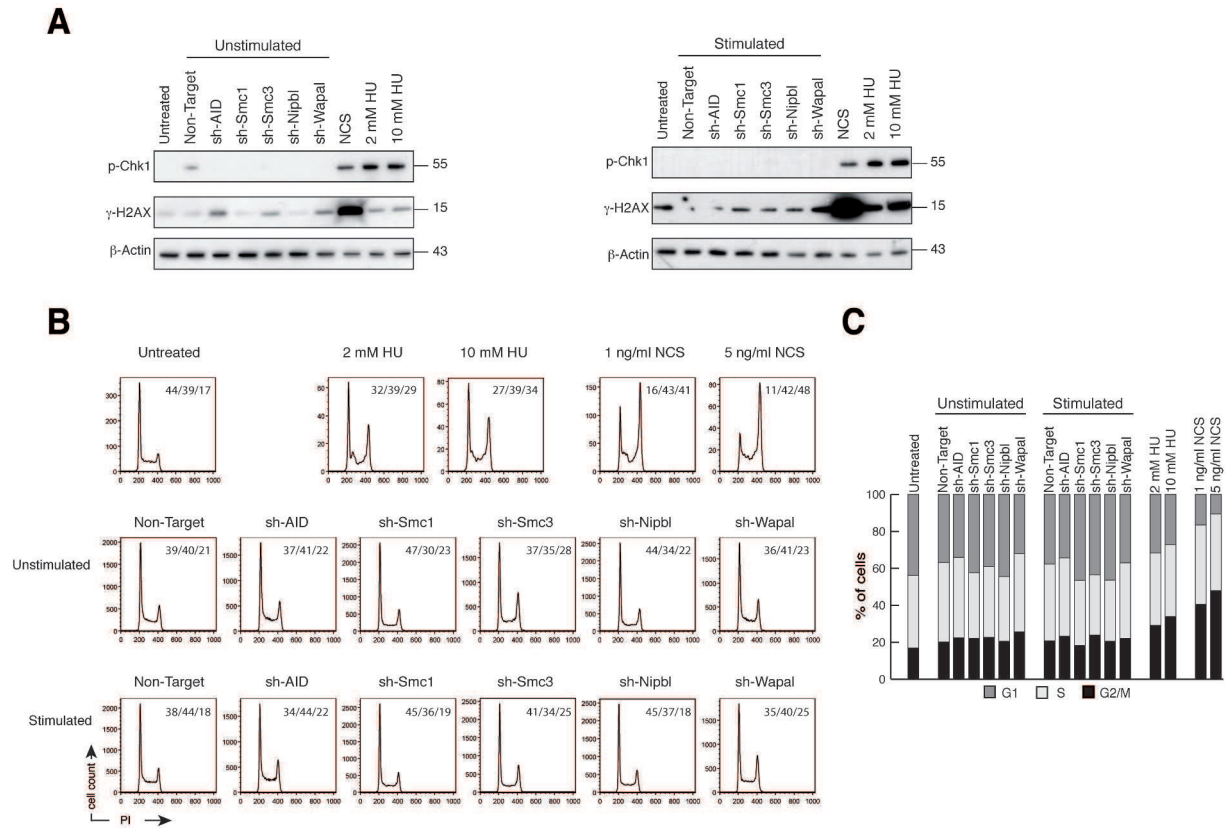
Gene Name (GN), protein description, molecular weight in kilodaltons (MW) and corresponding number of peptides (Pep) and spectral counts (SCs) found are shown on Flag immunoprecipitations conducted on extracts prepared from CH12 cells expressing AID<sup>Flag-HA</sup> or the epitope tags alone (Flag-HA), as a negative control.

GN	Description	MW	Flag-HA		AID <sup>Flag-HA</sup>		Reference
			Pep	SCs	Pep	SCs	
<b>Aicda</b>	<b>Activation-induced cytidine deaminase</b>	<b>24</b>	<b>0</b>	<b>0</b>	<b>17</b>	<b>981</b>	<b>BAIT</b>
<i>Eef1a1</i>	Elongation factor 1-alpha	50.1	6	19	18	238	(Hasler et al., 2011)
<i>Dnaja1</i>	DnaJ homolog subfamily A member 1	44.8	1	1	12	120	(Orthwein et al., 2012)
<i>Dnaja2</i>	DnaJ homolog subfamily A member 2	45.7	0	0	12	60	(Orthwein et al., 2012)
<i>Msh2</i>	DNA mismatch repair protein Msh2	104.1	0	0	16	44	(Ranjit et al., 2011)
<i>Xpo1</i>	Exportin-1	123	0	0	11	20	(Jeevan-Raj et al., 2011)
<i>Yy1</i>	Transcriptional repressor protein YY1	44.7	2	3	8	15	(Zaprazna and Atchison, 2012)
<i>Hsp90ab1</i>	Heat shock protein HSP 90-beta	83.3	1	5	5	15	(Orthwein et al., 2010)
<i>Ywhaz</i>	14-3-3 protein zeta/delta	27.8	0	0	4	14	(Xu et al., 2010)
<i>Ywhae</i>	14-3-3 protein epsilon	29.2	3	4	4	14	(Xu et al., 2010)
<i>Ywhaq</i>	14-3-3 protein theta	27.8	0	0	3	13	(Xu et al., 2010)
<i>Ywhab</i>	14-3-3 protein beta/alpha	28.1	0	0	3	12	(Xu et al., 2010)
<i>Trim28</i>	Transcription intermediary factor 1-beta	88.8	0	0	9	11	(Jeevan-Raj et al., 2011)
<i>Polr2A</i>	DNA-directed RNA polymerase II subunit RPB1	217	1	3	2	10	(Nambu et al., 2003)
<i>Prkdc</i>	DNA-dependent protein kinase catalytic subunit	471.1	4	5	8	9	(Wu et al., 2005)
<i>Psmc3</i>	Proteasome activator complex subunit 3 (REG-γ)	29.5	0	0	2	9	(Uchimura et al., 2011)
<i>Polr2b</i>	DNA-directed RNA polymerase II subunit RPB2	133.8	1	1	1	6	(Nambu et al., 2003)
<i>Cdc73</i>	Parafibromin	60.5	0	0	4	5	(Willmann et al., 2012)
<i>Ctr9</i>	RNA polymerase-associated protein CTR9 homolog	133.3	3	3	2	3	(Willmann et al., 2012)
<i>Supt16h</i>	FACT complex subunit SPT16	119.7	0	0	2	2	(Willmann et al., 2012)
<i>Ssrp1</i>	FACT complex subunit SSRP1	80.8	1	1	2	2	(Willmann et al., 2012)
<i>Ctnnb1</i>	Beta-catenin-like protein 1	64.9	1	1	2	2	(Conticello et al., 2008)
<i>Rpa1</i>	Replication protein A 70 kDa DNA-binding subunit	69	0	0	1	2	(Chaudhuri et al., 2004)
<i>Supt5h</i>	Transcription elongation factor SPT5	120.6	0	0	1	1	(Pavri et al., 2010)

**A****B**

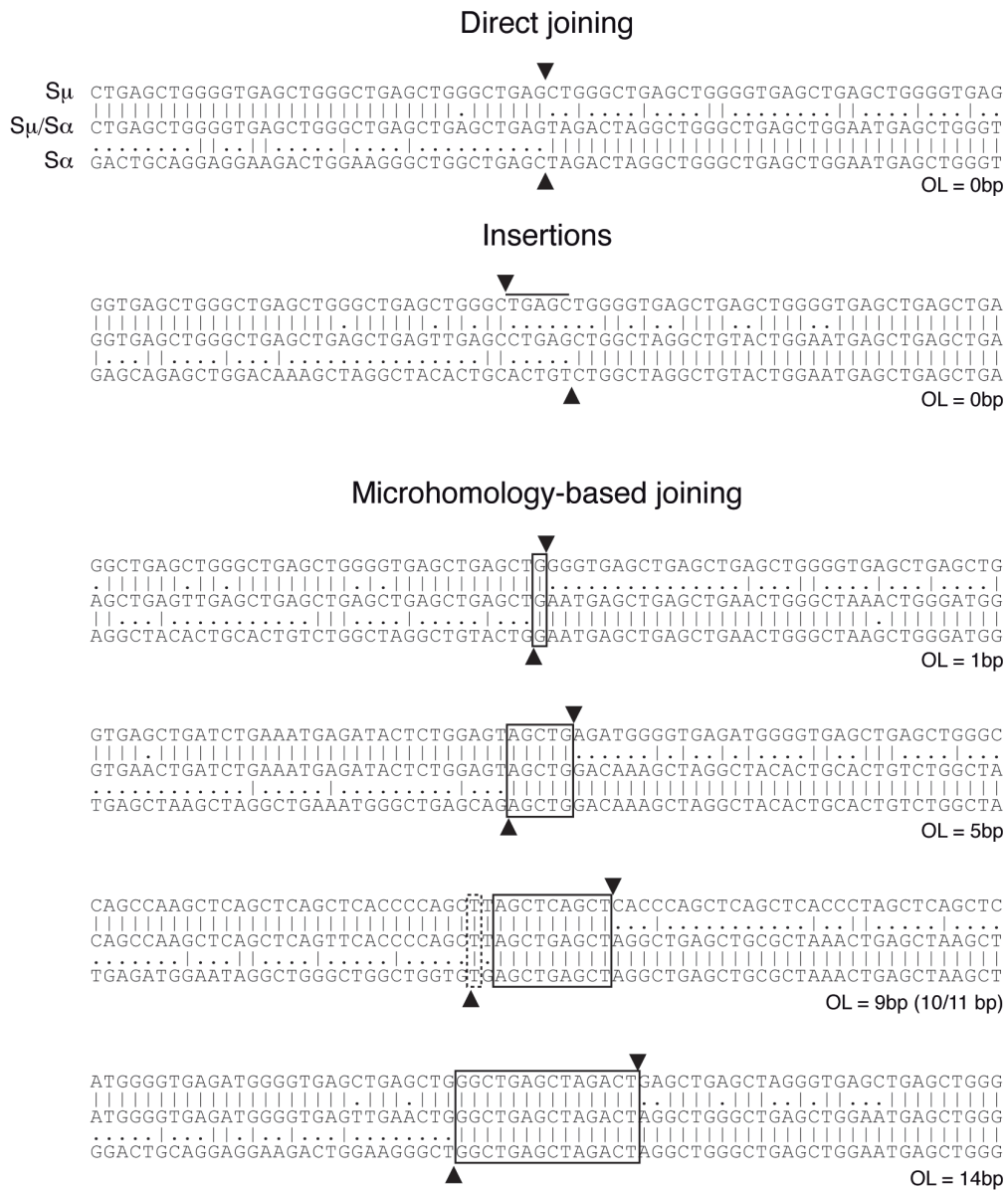
### Figure S1. Proliferation and CSR after knockdown of cohesin

(A) IgA surface expression as determined by flow cytometry in stimulated CH12 cells labeled with CFSE and transduced with a lentivirus (lacking a GFP reporter) and expressing shRNAs specific for AID, Smc1, Smc3, Nipbl and Wapal or a Non-Target shRNA negative control. Histograms depicting CFSE intensity (left panel) at day 0 (shown in blue) and after 48h (shown in red) are shown. Gates and percentage of cells gated are indicated. Gate 1 comprises all dividing cells whereas gate 2 includes only cells having proliferated equally. IgA surface expression analysis on gate 1 (middle panel) and gate 2 (right panel) is shown. Numbers within the plots indicate the percentage of IgA positive cells. Representative histograms and plots from two independent experiments are shown. (B) Percentage (+ s.d.) of CSR relative to the Non-Target shRNA control from two independent experiments analyzed on cells gated on gate 1 (white bars) or on gate 2 (black bars). CSR in cells expressing the Non-Target shRNA control was set to 100%. Statistical significance vs. the Non-Target control was determined by a two-tailed Student's t-test. \*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ . Analysis of CSR in the population of equally dividing cells (gate 2) still shows a defect in CSR efficiency upon cohesin knockdown.



**Figure S2. Cell cycle analysis and checkpoint activation after knockdown of cohesin**

(A) Western blot analysis using antibodies specific for the phosphorylated form of the Chk1 kinase (p-Chk1) and the histone variant H2AX ( $\gamma$ -H2AX) and  $\beta$ -Actin performed in CH12 cells treated or not with hydroxyurea (HU; 2 or 10 mM) or neocarcinostatin (NCS; 200 ng/ml) and CH12 transduced with lentiviruses expressing shRNAs specific for AID, Smc1, Smc3, Nipbl and Wapal or a Non-Target shRNA negative control and sorted for GFP expression. (B) Representative histograms of DNA content flow cytometry analysis as determined by propidium iodide (PI) incorporation in cells described in (A) before (unstimulated) or after 48h of stimulation (stimulated). Percentage of cells in G1/S/G2-M is indicated within the histogram. (C) Histograms showing the cell cycle distribution in cells analyzed in (B). Data are representative from two to three independent experiments.



**Figure S3. Examples of switch junction alignments**

Three-wise alignments where the donor switch region ( $S_{\mu}$ ; top), the switch junction (middle) and the acceptor switch region ( $S_{\alpha}$ ; bottom) sequences are shown. Microhomology was determined by identifying the longest region of perfect uninterrupted donor/acceptor identity. Solid and dashed boxes indicate perfect homology and allowing 1 bp mismatch, respectively. The length of overlap in base pairs is indicated on the bottom right. Filled arrows indicate breakpoints.

**Table S2. Antibodies used**

<b>Antibody</b>	<b>Clone</b>	<b>Source</b>	<b>Use*</b>
AID	Strasbg 9 (AID-2E11)	IGBMC (Jeevan-Raj et al., 2011)	WB, IP
AID	Polyclonal	IGBMC	ChIP
Flag	M2	Sigma	WB, IP
Smc1	A300-055A	Bethyl	WB, IP, ChIP
Smc3	ab9263	Abcam	WB, IP, ChIP
Nipbl	A301-779A	Bethyl	IP
Wapal	A300-268A	Bethyl	WB, IP
CTCF	07-729	Millipore	ChIP
$\beta$ -Actin	A1978	Sigma	WB
$\gamma$ -H2AX	JBW301	Millipore	WB
p-Chk1	133D3	Cell Signaling	WB

\* WB: western blot; IP: immunoprecipitation; ChIP: chromatin immunoprecipitation

**Table S3. Primers and probes used**

**Germline transcripts**

Primer	Sequence (5'-3')	Probe or Reference
I $\mu$ -C $\mu$ -Fwd	ACCTGGGAATGTATGGTTGTGGCTT	(Jeevan-Raj et al., 2011)
I $\mu$ -C $\mu$ Rev	TCTGAACCTTCAAGGATGCTCTTG	(Jeevan-Raj et al., 2011)
I $\alpha$ -C $\alpha$ Fwd*	GGAGACTCCCAGGCTAGACA	UPL probe #27
I $\alpha$ -C $\alpha$ Rev*	CGGAAGGGAAGTAATCGTGA	UPL probe #27

**Switch junctions**

Primer	Sequence (5'-3')	Probe or Reference
S $\mu$ -Fwd	AACAAGCTTGGCTTAACCGAGATGAGCC	(Schrader et al., 2002)
C $\alpha$ -Rev	CCGGAATTCCTCAGTGCAACTCTATCTAGGTCT	(Ehrenstein and Neuberger, 1999)

**Knockdown**

Primer	Sequence (5'-3')	Probe or Reference
Nipbl-Fwd*	CCCTTAAGATCTCCTCAACCAG	UPL probe #2
Nipbl-Rev*	TGTAGAATTAAGGTGGTCTTGAGC	UPL probe #2
CD79b-Fwd*	TGGTGCTGTCTTCCATGC	UPL probe #18
CD79b-Rev*	TTGCTGGTACCGGCTCAC	UPL probe #18
HPRT-Fwd*	GTCACGGGGGACATAAAAA	UPL probe #22
HPRT-Rev*	CAACAATCAAGACATTTCTTTCCA	UPL probe #22

**ChIP**

Primer	Sequence (5'-3')	Probe or Reference
E $\mu$ -Fwd	GGGAGTGAGGCTCTCTCATA	(Wang et al., 2009)
E $\mu$ -Rev	ACCACAGCTACAAGTTTACCTA	(Wang et al., 2009)
5'S $\mu$ -Fwd1	TAAAATGCGCTAAACTGAGGTGATTACT	(Kuang et al., 2009)
5'S $\mu$ -Rev1	CATCTCAGCTCAGAACAGTCCAGTG	(Kuang et al., 2009)
5'S $\mu$ -Fwd2	TAGTAAGCGAGGCTCTAAAAGCAT	(Pavri et al., 2010)
5'S $\mu$ -Rev2	AGAACAGTCCAGTGTAGGCAGTAGA	(Pavri et al., 2010)
3'S $\mu$ -Fwd	CTGAATGAGTTTACCAGGCC	(Wang et al., 2006)
3'S $\mu$ -Rev	GCCTGTCTGCTTGGCTTC	(Wang et al., 2006)
C $\mu$ -Fwd*	GTTCTGTGCCTCCGTCTAGC	
C $\mu$ -Rev*	AGCATTTGCATAAGGGTTGG	
S $\gamma$ 3-Fwd	GCTGAGAGTATGCACAGCCA	(Wang et al., 2006)
S $\gamma$ 3-Rev	GGATCATGGAAACTCCTCCG	(Wang et al., 2006)
S $\gamma$ 1-Fwd	GGAGGTCCAGTTGAGTGTCTTTAG	(Muramatsu et al., 2000)
S $\gamma$ 1-Rev	TTGTTATCCCCATCCTGTACCT	(Muramatsu et al., 2000)
C $\alpha$ -Fwd*	CTCCTGTCTCACAGGCCTTC	
C $\alpha$ -Rev*	CATGGGCCTTTACTCCACTC	
Hs6,7-Fwd*	CCCTGGTGACCATGTGTGT	
Hs6,7-Rev*	TCTGGGTCTGTTTTGTTACTGAAA	

Primers designed in this study are marked with an \*



**Table S4. ChIP Statistics (Two-tailed Student's t-test)**

Blue highlight indicates statistical significance.

**MS1**

ms1g	STP1	STP2	Cp	Sp	Sp1	Cs	ms1F
Sp	0.1783	0.0206	0.0006	0.0165	0.0045	0.0008	0.0002
STP1		0.1253	0.0817	0.0589	0.0679	0.0008	0.0003
STP2			0.2372	0.8895	0.4439	0.0068	0.0027
Cp				0.1692	0.0239	0.0074	0.0029
Sp					0.5373	0.0010	0.0003
Sp1						0.0009	0.0003
Cs							0.7355

**dm**

dm	Su5'	Su3'	Cu	Sg3	Sg1	Ca	dmF
Sp	0.0317	0.0006	0.0336	0.2580	0.0308	0.0152	0.0013
STP1		<0.0001	0.6838	0.0388	0.7609	0.0057	0.0006
STP2			<0.0001	<0.0001	0.0002	0.0001	0.0284
Cp				0.0179	0.5172	0.0012	0.0003
Sp					0.0606	0.0033	0.0011
Sp1						0.0102	0.0011
Cs							0.0261

**ms1g vs dm**

ms1g	Sp	STP1	STP2	Cp	Sp	Sp1	Cs	dmF
ms1g	0.1386	0.0002	<0.0001	0.0005	0.0006	0.0004	0.7394	0.0142

**MS3**

ms3g	STP1	STP2	Cp	Sp	Sp1	Cs	ms3F
Sp	0.0035	0.2136	0.4362	0.4176	0.0147	0.0004	<0.0001
STP1		0.0051	0.0011	0.0089	0.0099	0.0003	<0.0001
STP2			0.0331	0.0921	0.0052	0.0038	0.0006
Cp				0.1211	<0.0001	0.0004	<0.0001
Sp					0.0007	0.0034	0.0005
Sp1						0.0003	<0.0001
Cs							0.2814

**dm**

dm	Su5'	STP1	Cu	Sg3	Sg1	Ca	dmF
Sp	0.0019	<0.0001	0.0018	0.9686	0.0012	0.0007	0.0004
STP1		0.0057	0.0418	0.0031	0.0923	0.0523	0.0044
STP2			0.1916	0.0003	0.0037	0.0573	0.1521
Cp				0.0027	0.0335	0.5436	0.0859
Sp					0.0142	0.0015	0.0006
Sp1						0.0207	0.0094
Cs							0.0432

**ms3g vs dm**

ms3g	Sp	STP1	STP2	Cp	Sp	Sp1	Cs	dmF
ms3g	0.3172	0.0008	0.0021	0.0019	1.0000	0.0007	0.0226	0.5388

**MS4**

ms4g	STP1	STP2	Cp	Sp	Sp1	Cs	ms4F
Sp	ND	ND	ND	ND	ND	ND	ND
STP1		0.8032	0.0122	0.0086	0.0755	0.0025	0.0059
STP2			0.9086	0.8551	0.9393	0.6766	0.2214
Cp				0.0324	0.0717	0.0341	0.0078
Sp					0.0404	0.0529	0.0083
Sp1						0.0078	0.0071
Cs							0.0119

**dm**

dm	Smu570571	Smu862863	Cmu	Sg3	Sg1	Ca	dmF
Sp	0.0862	0.0042	0.0407	0.9352	0.0404	0.0465	0.0013
STP1		0.1026	0.0163	0.0985	0.2332	0.1449	0.0048
STP2			0.0124	0.5445	0.0632	0.0501	0.0013
Cp				0.1669	0.0053	0.0078	<0.0001
Sp					0.1012	0.0489	0.0016
Sp1						0.0685	0.0017
Cs							0.0207

**ms4g vs dm**

ms4g	Sp	STP1	STP2	Cp	Sp	Sp1	Cs	dmF
ms4g	ND	0.0079	0.8504	0.0302	1.0000	0.0322	0.3909	0.0141

## Supplementary references

- Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* **430**, 992-998.
- Coticello, S.G., Ganesh, K., Xue, K., Lu, M., Rada, C., and Neuberger, M.S. (2008). Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNB1. *Mol Cell* **31**, 474-484.
- Ehrenstein, M.R., and Neuberger, M.S. (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *Embo J* **18**, 3484-3490.
- Hasler, J., Rada, C., and Neuberger, M.S. (2011). Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A). *Proc Natl Acad Sci U S A*.
- Jeevan-Raj, B.P., Robert, I., Heyer, V., Page, A., Wang, J.H., Cammas, F., Alt, F.W., Losson, R., and Reina-San-Martin, B. (2011). Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *J Exp Med* **208**, 1649-1660.
- Kuang, F.L., Luo, Z., and Scharff, M.D. (2009). H3 trimethyl K9 and H3 acetyl K9 chromatin modifications are associated with class switch recombination. *Proc Natl Acad Sci U S A* **106**, 5288-5293.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553-563.
- Nambu, Y., Sugai, M., Gonda, H., Lee, C.G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* **302**, 2137-2140.
- Orthwein, A., Patenaude, A.M., Affar el, B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *J Exp Med* **207**, 2751-2765.
- Orthwein, A., Zahn, A., Methot, S.P., Godin, D., Coticello, S.G., Terada, K., and Di Noia, J.M. (2012). Optimal functional levels of activation-induced deaminase specifically require the Hsp40 DnaJa1. *Embo J* **31**, 679-691.
- Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina-San-Martin, B., Barreto, V., *et al.* (2010). Activation-Induced Cytidine Deaminase Targets DNA at Sites of RNA Polymerase II Stalling by Interaction with Spt5. *Cell* **143**, 122-133.
- Ranjit, S., Khair, L., Linehan, E.K., Ucher, A.J., Chakrabarti, M., Schrader, C.E., and Stavnezer, J. (2011). AID binds cooperatively with UNG and Msh2-Msh6 to Ig switch regions dependent upon the AID C terminus. *J Immunol* **187**, 2464-2475.
- Schrader, C.E., Vardo, J., and Stavnezer, J. (2002). Role for mismatch repair proteins Msh2, Mlh1, and Pms2 in immunoglobulin class switching shown by sequence analysis of recombination junctions. *J Exp Med* **195**, 367-373.
- Uchimura, Y., Barton, L.F., Rada, C., and Neuberger, M.S. (2011). REG-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. *J Exp Med*.
- Wang, L., Whang, N., Wuerffel, R., and Kenter, A.L. (2006). AID-dependent histone acetylation is detected in immunoglobulin S regions. *J Exp Med* **203**, 215-226.
- Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J Exp Med* **206**, 1817-1830.
- Willmann, K.L., Milosevic, S., Pauklin, S., Schmitz, K.M., Rangam, G., Simon, M.T., Maslen, S., Skehel, M., Robert, I., Heyer, V., *et al.* (2012). A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *J Exp Med* **209**, 2099-2111.
- Wu, X., Geraldes, P., Platt, J.L., and Cascalho, M. (2005). The double-edged sword of activation-induced cytidine deaminase. *J Immunol* **174**, 934-941.
- 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.
- Zaprazna, K., and Atchison, M.L. (2012). YY1 controls immunoglobulin class switch recombination and nuclear activation-induced deaminase levels. *Mol Cell Biol* **32**, 1542-1554.

# **DISCUSSION**

## 1. Patient analysis and transcriptome profiling: the dark side of the approach

The experimental approaches that can be applied to address a scientific question might be very different but the opportunity to analyze patients affected by a specific pathology, with the aim to characterize the molecular defect underneath, is a powerful tool. One clear example is AID: although the deaminase was identified by cDNA subtraction screen of unstimulated and stimulated CH12 B cells (Muramatsu et al., 1999), its role in switching and somatic hypermutation has been clarified by the study of Revy et al. based on patients affected by hyper-IgM syndrome, and complemented by the work of Muramatsu and collaborators on AID knockout mice (Muramatsu et al., 2000; Revy et al., 2000). The advantage given by this approach represented the reason behind our study, focused on the CSR-specific AID cofactors. With the aim to better understand the mechanisms regulating CSR, we performed a transcriptome and proteome profiling of B cells obtained from CSR-ID patients harboring a CSR-specific defect, and we complemented these analyses with the proteome identified in CH12 cells overexpressing AID or its C-terminal domain. Whereas for the identification of potential candidates the transcriptome and genome sequencing results are pivotal, the proteome and *in vitro* characterization by using the CH12 mouse cell line can reinforce the potential discoveries and allow us to address the function of candidate genes.

Nevertheless, in this study we encountered some limitations, which had an impact on the results. The first one is represented by the limited number of samples: considering the gene polymorphisms and the huge variability existing between individuals, a cohort of three patient-derived B lines as well as three controls used for transcriptome profiling might not be sufficient to focus on significant differences which are intrinsic of the phenotype described. Furthermore, our experiment was initially including a fourth patient-derived line but a technical issue, which occurred during the cDNA library preparation, made it unavailable for further analysis. Although we succeeded to sequence it, an improvement of the Illumina sequence annotation database impaired us from pooling these data with the previously obtained data set. However, we included this cell line in the western blots we performed.

Another limitation was represented by the lack of information about the patients and controls. Were they consanguineous? Did they match for sex and/or age? Did they have the same clinical profile? We assumed that the cell lines analyzed were obtained from not related individuals, but further details about them would have allowed us to modify the analysis: we could have pooled related samples before comparison with the controls in order to perform a better comparison. We also hypothesized that, within the group of patients analyzed and despite the homogeneity detected, there might be a further sub-division, due possibly to a difference on the patients' clinical profile. We analyzed the sequencing data considering each of the CSR-ID line as independent condition, thus comparing each individual line to the group of three healthy donors and to the AID deficient line (data not shown). However, our attempt did not lead to any further improvement in the "cofactor discovery" as the variability inter-samples made the data interpretation even more difficult than before.

The technical limitations mentioned so far cannot be solved neither improved; however, the use of mouse models could be useful to complement our analysis. The transcriptome profiling of human B

lines provided information about those genes regulated upon CSR and SHM and, by comparing the data obtained from CSR-ID patient lines and AID<sup>-/-</sup> line with the control ones it is possible to identify those genes dependent on AID expression and required for CSR. However, the transcriptome profiling of mouse B cells stimulated or not to undergo CSR could complement this approach, as between mouse and human exists a strong similarity in terms of mechanisms and factors involved in antibody diversification. By using WT and AID<sup>-/-</sup> B cells, to mimic the condition of the healthy donor controls and AID-deficient patients, and by analyzing the transcriptome of resting B cells, we would obtain the background expression of mouse B cells and, between cells proficient or deficient for AID, we do not expect to observe any significant differences as AID expression is regulated upon activation. On the other hand, the comparison between differentially regulated genes in WT and AID<sup>-/-</sup> B cells prior to or upon activation would allow us to select CSR and AID-dependent genes. Furthermore, additional comparison with the human samples would allow us to cut-off those genes deregulated in patients and dependent on the experimental condition, such as EBV infection-dependent genes.

Additionally, to complement the patients' condition – namely AID expression and impairment in CSR despite normal SHM – we could use resting B cells isolated from AID-deficient mice and transduced with a retrovirus expressing a C-terminal truncated form of AID as fused to the GFP reporter genes and to the Flag and HA tags (AID<sup>(1-116)</sup>-GFP-Flag-HA). Upon cell sorting of GFP<sup>+</sup> B cells, we would be able to obtain a population of cells which express AID lacking its C-terminal domain and which are able to undergo SHM but not CSR, as previously described (Barreto et al., 2003). As control for this condition, we could use mouse AID<sup>-/-</sup> cells transduced with a retrovirus expressing GFP-Flag-HA as well as AID<sup>-/-</sup> cells overexpressing AID, as further control for the retrovirus-induced genes, and sorted for GFP expression. Upon stimulation and transcriptome analysis, we would obtain the deregulated genes dependent on AID activity but not related to switching.

Concerning the approach to adopt for the additional transcriptome profiling, next generation sequencing advances have replaced the microarray and DGE method with RNA-Seq, which allows a reproducible, sensitive and unbiased transcripts detection. Thus, this experiment would complement the analysis we performed on human B cell lines and provide new insights in switching regulation.

## 2. Validated or not validated: that is the question

The large-scale data analyses (–omics approaches) have revolutionized the way researchers can address a specific question. The opportunity to obtain a large set of data that represent a kind of “picture” of the condition analyzed; furthermore, the ability to compare set derived from different approaches and their reproducibility can be considered as one of the most important technological advancement of the last years. Concerning my work, we used these approaches to analyze the transcriptome and the proteome of B cells but, if on one hand the huge amount of data provides many opportunities to confirm the hypothesis, on the other hand technical issues and experimental conditions may reflect on the results obtained, thus the data validation is a required step to get a

conclusion from the experiment. When we undertook the transcriptome profiling of EBV-immortalized human B lines from CSR-ID patients and controls, the DGE represented the “new technology” available. Unlike microarray technology, which allows the quantification of transcripts coding for annotated genes and thus requires an upstream knowledge of what one is going to look at, DGE is a sequence-based approach, which then allows the identification of either known or unknown transcripts. Moreover, an additional advantage is the quantitative aspect of this technique, where the detection of a certain number of tags directly reflects the abundance of the identified transcript. However, as this field is evolving quite fast, the DGE has been available only for a relatively short period of time and now has been replaced by the RNA-Seq.

In order to validate the data obtained by DGE, we sorted the deregulated genes obtained from each of the three conditions analyzed (patients vs. controls, patients vs. AID<sup>-/-</sup> and controls vs. AID<sup>-/-</sup>, whose example is shown in Tables 5, 6 and 7) and we chose random genes from the most upregulated and downregulated to perform RT-qPCR. The choice of the genes presented in the results has been also influenced by the amplification efficiency, as some of them, despite many attempts, did not provide any result suitable for relative quantification. Although the expression of most of the genes we assessed by RT-qPCR reflected the profile detected by DGE, we observed some incongruences (Figures 20B and 21C). The expression of *PSMA4* did not reflect the tags detection in any of the conditions analyzed. Additionally, we also observed an opposite behavior for one of the control lines we included in this assay but which was not included in the transcriptome profiling (Ctr #1, Figures 20F and 21D). While, according to the *TESC* and *PRF1* expression levels, we verified the downregulation of these genes in controls when compared to the AID<sup>-/-</sup> line, for Ctr #1 cell line we detected an overexpression of *TESC*, in sharp contrast with the others (Figure 20F). In line with this result, we observed a strong reduction in *PFN2* expression in Ctr #1 (Figure 21D), comparable to the one detected in patients, where *PFN2* was supposed to be downregulated and which was opposite to the higher expression detected in the other two control lines.

Although these results suggest that differences between the transcriptome analysis and the actual amount of a particular transcript must be taken into account, the limited number of genes assessed for validation must also be considered, as a higher numbers of assays would provide a more precise overview of the “expression trend” within the samples analyzed. Nevertheless, in our analysis, we also identified genes whose expression level assessed by RT-qPCR was reflecting the abundance of tags detected by DGE and thus allowed further data analysis.

### 3. Missing factor: is it really downregulated?

In order to identify the molecular defect leading to the CSR-ID observed in the patients' B lines analyzed – CSR impairment associated to normal AID expression and normal mutation frequency and pattern at the IgV regions – we performed a transcriptome and proteome profiling, hypothesizing that a CSR-specific factor able to interact with the C-terminus of AID and to target the deaminase to the S

regions would be “missing” or expressed at lower levels in patients when compared to controls. The plethora of AID regulators and the numerous studies based on either knockout models or gene silencing showed how reduced levels of one of these regulators can impair switching, somatic hypermutation or gene conversion. However, we cannot exclude completely the opposite hypothesis: that it may be the case of overexpression. What supports this alternative is that our collaborators sequenced in patients those genes coding for known AID partners and/or DNA repair proteins but no mutations were found. Thus, we may suppose that, if we consider a known and “unsuspectable” AID partner, its stability might be altered in case of reduced levels of a negative regulator. By considering AID as example, has been shown as its turnover depends on the action of miRs such as miR-181b, miR-155 and miR-93 (Borchert et al., 2011; de Yébenes et al., 2008; Dorsett et al., 2008), whose expression is tightly regulated in resting or activated B cells. As miRs lack the poly(A) tail, DGE technology or any other based on the selection of polyadenylated transcripts is not suitable for this type of analysis but RNA-Seq profiling, instead, would allow to address this question. Additionally, we could reconsider our analysis by focusing on upregulated transcripts; although, DGE profiling does not provide any information about the rate of transcription and/or translation, the mRNA stability or protein-protein interaction but additional analyses on proteome data by focusing on overexpressed proteins could help to overcome this issue.

#### 4. The importance of being within the nucleus

The transcriptome profiling we performed on human B lines was based on the assumption that if a mutation in a gene coding for the factor(s) deregulated in CSD-ID patients would affect the gene expression level, we could detect it by DGE. However, if the mutation is located in an intron, or has no effect on the transcription of this gene, we will not be able to identify it. So we performed, as complementary approach, the proteome profiling of human B lines and, in parallel, of the mouse CH12 B line overexpressing AID or its C-terminal domain. By extracting nuclear protein we immunoprecipitated AID and looked at its interactors.

Considering that AID shuttles between the nucleus and the cytoplasm (Brar et al., 2004; Geisberger et al., 2009; Ito et al., 2004; McBride et al., 2004; Patenaude et al., 2009) and that it is retained mainly in the cytoplasm (Brar et al., 2004; Geisberger et al., 2009; Ito et al., 2004; McBride et al., 2004), where it is stabilized by eEF1A (Hasler et al., 2011) and the chaperones Hsp90 and Hsp40 Dnaja1 (Orthwein et al., 2010; Orthwein et al., 2012), while in the nucleus its availability is maintained by the interaction with YY1 (Zaprazna and Atchison, 2012), the trouble of “where to look at” rises. As AID exerts its deamination activity in the nucleus, this makes the nuclear proteome the best “environment” where we can expect to successfully identify new AID interactors. However, numerous studies have shown that the cytoplasmic compartment still provides good candidates that, although identified in a “neutral environment”, can indeed regulate AID localization to the target regions. One clear example is Spt6, identified in the cytoplasm of AID-overexpressing CH12 B cells by Okazaki et al. (Okazaki et al., 2011). However, it is difficult to decide *a priori* which approach would be the best, as the purity of the

fractions and, in case of immunoprecipitation, the stringency of the conditions reflect on the AID interactome and only functional studies can rule out whether the candidate identified is playing any role in antibody diversification.

## 5. Does AID size and domains matter?

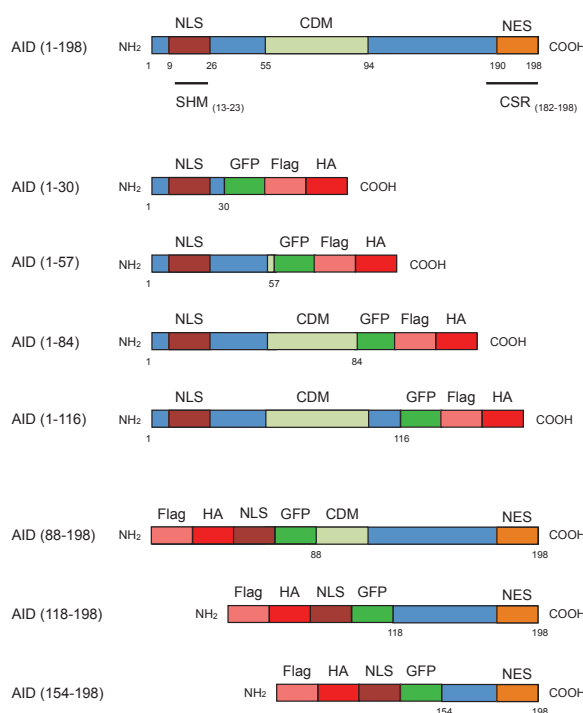
One of the unanswered questions about AID is: how such a small protein is able to interact with such a large number of factors? In addition to space and time constraints – and to the need of regulating its deaminase activity – which might influence the cooperation with certain factors, how is possible that it seems to be in a complex with so many proteins within the cell? AID is a relatively small protein of 198 residues and about 25 KDa. So far is not well defined whether it acts as a monomer or a dimer: Wang and collaborators in 2006 published an article about AID dimerization which was later retracted (Wang et al., 2006); one year later, Prochnow and colleagues performed structural studies on APOBEC2 in order to predict, according to sequence homology, the structure of AID. By expressing GST-tagged AID harboring mutations in those residues corresponding to APOBEC2 tetramerization domain, they observed impairment in AID deaminase activity, suggesting that mutation of these residues affects AID activity (Prochnow et al., 2007). Furthermore, a third report displayed that AID acts as a monomer on ssDNA (Brar et al., 2008), while mutagenesis of those residues located in the region predicted to be required for AID dimerization impairs oligomerization as well as AID nuclear import (Patenaude et al., 2009). This debate has not been solved yet and this “stream” of research is on standby, also due to the fact that AID structure is not available; thus researchers are trying to focus on other aspects of AID regulation and intrinsic properties, which are most likely easier to address.

Another important issue, which is directly reflecting on one of the main assumptions we made throughout this work, is the separation of roles in SHM and CSR according to AID domains requirements. Studies based on mutagenesis, as well as patients harboring AID mutations and affected by CSR-immunodeficiencies, displayed that the N-terminal and C-terminal domain of AID are differentially required for the two antibody diversification mechanisms. However, it appears that this division is not so strict and that certain of the factors whose involvement has been described for CSR are interacting with the N-terminus of the deaminase. This is the case for Spt6, which has been shown to be dispensable for SHM despite its binding to the residues 2-26, spanning the NLS (Okazaki et al., 2011), and also for Paf1, where the generation of AID/APOBEC2 chimeras showed that the region spanning residues 19-85 is critical for Paf1/AID interaction in HEK293 cells (see chapter IV, Figure 3D, Willmann et al., 2012). In the latter case the requirement of the PAF complex for SHM has not been addressed yet, but still shows that the factors important in regulating AID during CSR do not exclusively interact with its C-terminus.

This evidence brings to the key question related to this project: is the “missing factor” in CSR-ID patients binding to AID C-terminus? As we identified either proteins interacting with full length mouse AID and those bound to its last 17 residues, we did not introduce any bias in our analysis, as shown by the list of known AID interactors obtained upon overexpression of Flag-HA-AID<sup>(1-198)</sup> in CH12 B



cells. Additional experiments we could perform to solve “what is binding where” is to overexpress N-terminal and C-terminal AID truncations in CH12 B cells and, upon Flag IP and comparison of the different pattern identified, define the differential interactome for each form of the protein. Based on the work from Patenaude and co-workers (Patenaude et al., 2009), we generated the mouse AID constructs depicted in Figure 32 as well as two additional controls to the GFP-Flag-HA construct, based on APOBEC2 sequence (hAPOBEC2<sup>(1-84)</sup>-GFP-Flag-HA and Flag-HA-NLS-GFP-APOBEC2<sup>(88-198)</sup>). This experiment would allow us to map the specific interaction of the proteins identified to the specific AID domain and further functional studies would help to clarify whether a SHM or CSR-specific interactome exists.



**Figure 32. AID N-terminal and C-terminal truncations**

Representation of AID N-terminal and C-terminal truncations generated in our laboratory. All the constructs harbor a Flag-HA tag and a GFP reporter; additionally, the ones lacking the N-terminus of AID harbor a nuclear localization signal (NLS) for proper localization.

## 6. The short story of Spt6

In order to identify the deregulated factor(s) responsible of the CSR-ID phenotype observed in patients, we integrated data obtained from the transcriptome and the proteome analysis, and we identified Spt6. Strikingly, Spt6 lower expression resulted from the proteome screening, whereas no differential gene expression was detected. Although, as three isoforms have been described for Spt6, to rule out that a compensatory mechanism could account for the expression level detected by

transcriptome, we analyzed the DGE tag sequences in order to distinguish between the three isoforms. However, we were not able to discriminate them and we hypothesized that Spt6 deregulation was post-transcriptional and thus due to a lower translation rate or protein instability.

Moreover, when we verified the protein expression levels in human B lines by western blot, we observed a difference between different experiments, which still suggested lower Spt6 levels in CSR-ID B cells (Figure 24). Nevertheless, the co-immunoprecipitation experiments we performed in CH12 B cells overexpressing full-length AID showed that AID and Spt6 were in the same complex (see chapter IV, Figure 2B, Willmann et al., 2012). Most importantly, we found that also Spt5 and the RNA polymerase II were interacting with AID and, in light of the previously described interactions between the latter and the deaminase and their role in CSR (Nambu et al., 2003; Pavri et al., 2010), we decided to better characterize Spt6 by gene silencing in CH12 B cells.

Unfortunately, the experiments we performed with recombinant retroviruses expressing Spt6 shRNAs did not provide the expected efficiency in terms of Spt6 depletion, thus we were not able to identify a clear trend in CSR efficiency (either an increase or a decrease); so we decided to use recombinant lentiviruses. As Rushad Pavri and collaborators identified Spt5 through shRNA library screening, this system was supposed to be a solution. However, it took a long time to optimize the conditions for an optimal transfection, transduction and selection of the cell lines. Sorting transduced cells expressing the GFP reporter prior to or after stimulation represented a pivotal step to evaluate the expression levels of the protein whose coding gene was targeted by the shRNA and the consequent ability of those cells to undergo switching, as well as AID expression and switch regions germline transcription. Although we applied these improvements to the characterization of the cohesin complex in CSR, and, partially, to the Smc5/6 complex, this has not been the case for Spt6, as the assays we performed were pioneer experiments.

Moreover, the publication from the laboratory of Tasuku Honjo showed that Spt6 plays a role in CSR (Okazaki et al., 2011) but, even if the methodology used is quite similar (identification by proteome screening and functional characterization through knockdown), there are some differences that might be underlined. As already mentioned, Okazaki et al. analyzed the cytoplasmic compartment for AID interactors, whereas we focused on the nuclear one, in light of the fact that the deamination reaction occurs within the nucleus and thus is most likely there that we expect to find factors required to target AID to the S regions. Moreover, in the first experiment displayed by Okazaki et al. and performed on CH12 B cells, they observed that siRNA-mediated Spt6 silencing induced a defect in CSR but, in parallel, impaired GLT at the acceptor S region as well as AID expression, detected by qPCR. Interestingly, upon Wapal knockdown in CH12 B cells, we detected a significant reduction in GLT $\alpha$  transcription as well, leading to the conclusion that the lower recombination efficiency upon Wapal silencing could be due to a reduced availability of the acceptor S region and not to an intrinsic requirement of Wapal for CSR.

By comparing our work to the one performed by other laboratories, it emerges that the knockdown strategy can be considered, on one hand, as a relatively easy tool to assess the function of a gene of interest in CSR but, on the other hand, the stability of this system and the consequences at the cellular level such as off-target effects, which are difficult to figure out, lead to the observation that gene

inactivation is still the best strategy - when applicable - to specifically address the function of the gene coding for the protein of interest.

## **7. The PAF complex, AID and transcription-associated factors: a complicated relationship**

The work we performed on Spt6 could not go further due to the Okazaki publication, but collaboration with the laboratory of Dr. Svend Petersen-Mahrt allowed us to contribute, with our data, to delineate a global picture of AID regulation mediated by transcription-associated factors and which was focused on the PAF complex.

The PAF complex has been identified through an elegant knock-in strategy applied to the DT40 chicken B cell line, that was based on endogenous tagged AID which allowed the identification of chromatin-related AID factors in a physiological context. This led to the identification of previously described partners, such as Pol II, Spt5, Spt6, the RNA exosome and the FACT complex and, in addition, the PAF subunits: Paf1, Leo1, Ctr9 and Cdc73. The subunit Rtf1 was not identified in this screening, possibly due to a more labile interaction. Studies performed on the yPAF in order to dissect the molecular interactions of each of its subunits evidenced that if, on one hand, Cdc73 seems to be the direct link between the PAF complex and the RNA polymerase II (Shi et al., 1997), on the other hand Rtf1 might interact with Spt5 (Squazzo et al., 2002). However, the situation seems to be a bit different in hPAF, where Paf1, and to a lesser extent Leo1, mediate the interaction with the RNA polymerase II (Kim et al., 2010). Concerning PAF/AID interaction, by analyzing DT40 and mouse CH12 B cells, we confirmed by immunoprecipitation that AID was in a complex with PAF, Spt5, Spt6 and the Pol II (see chapter IV, Figure 2, Willmann et al., 2012), as also indicated by the reciprocal co-IPs which have not been included in the final version of the publication. Moreover, we found a direct association between Paf1 and AID, suggesting that the interactions described for the hPAF might reflect the link between AID, PAF and the RNA polymerase II we observed in B cells.

These observations allowed us to address the role of PAF in switching by using CH12 B cells. By silencing the PAF subunits identified in the screening (Paf1, Leo1, Cdc73 and Ctr9) we observed an impaired CSR in all the conditions tested, but that in case of Paf1, Cdc73 and Ctr9 the defect observed could be the consequence of reduced GLT or AID expression. The only subunit which seemed to possess an intrinsic function in CSR was Leo1, and this was confirmed by ChIP experiments conducted on knockdown lines where we observed that AID recruitment was reduced upon Leo1 knockdown, suggesting that the PAF complex is involved in AID targeting to the DNA through interaction with Spt5 and the RNA polymerase II.

Moreover, many experiments can be proposed to dissect the role of PAF in antibody diversification. Leo1 knockdown should be assessed in primary B cells stimulated *ex vivo*, to confirm the effects observed in the CH12 B line. Additionally, we could assess the role of Leo1 in IGC, by generating a Leo1 knockout DT40 B line, and also in SHM, through a Leo1 knockout mouse model. Additionally, an

involvement of PAF in DNA repair cannot be excluded: as the interactome of the single subunits has not been completely clarified, it could be possible that the “binding platform” created during transcription might be required also for the interaction with repair proteins. The candidates could be part of BER or MMR pathway, such as UNG or MSH2/6, which are responsible of initiating the repair cascade, as well as the components of the NHEJ pathway in case a differential recruitment might affect the choice between the classical and alternative pathway. In this latter case, Leo1 activity might reflect on the microhomology length at the S regions and, to address this point, junction analysis in CH12 B cells upon Leo1 knockdown would be required. Moreover, it would be interesting to apply the knock-in strategy used in DT40 to CH12 B cells, to identify those proteins which bind to endogenous AID and thus reduce the bias given by its overexpression.

## 8. AID targeting: what does it mean?

The identification of the PAF complex and characterization of the role of its subunit Leo1 in CSR has brought a “fresh perspective” in the effort to understand how AID is targeted to the DNA. If, on one hand, the stalled RNA polymerase II fits with the idea that the GC-rich S regions temporarily impair its processing, on the other hand it is also true that AID does not target all transcribed genes, underlying a sort of “defense mechanism” against constitutive and genome-wide mutagenesis which would most probably affect cell viability. Thus this means that there is still a piece of the puzzle missing, an additional level of regulation.

ChIP experiments have been pivotal in clarifying AID recruitment to the Ig loci and the importance of specific factors. However, these experiments provide a kind of “snapshot”, at a particular time, of the status of this interaction(s), and thus we cannot rule out whether the interaction with a known factor occurred before or instead occurred, for instance, at the S regions already occupied by the RNA polymerase II and its associated factors. Furthermore, it is difficult to differentiate between AID targeting and tethering/retention and, adding an additional level of complexity, the difference in targeting between the S regions and the V regions has not been completely elucidated.

Genome-wide ChIP analysis provided additional information to the “classical” chromatin immunoprecipitation approaches. Concerning the PAF complex, it would be interesting to perform a Leo1 ChIP-Seq as well as Spt6 ChIP-Seq experiment in resting and activated mouse B cells to compare the profile obtained with the one described for Spt5, AID and the Pol II (Pavri et al., 2010; Yamane et al., 2011), to further address Leo1 and Spt6 role as targeting factors. Additionally, performing AID ChIP-Seq in stimulated B cells upon Leo1 knockdown would show us whether the genomic occupancy of AID in cells undergoing switching is altered specifically at the S regions (as observed for the donor S region) or it results in an imbalanced targeting throughout the genome; moreover, this experiment could also complement the SHM characterization.

Although the picture describing the exact mechanism of AID targeting is not complete, our work contributed to clarify the association between AID and the transcription-related factors, providing a base for further investigations.

## 9. The Smc5/6 complex: a functional “divorce”?

Our investigations on the Smc5/6 complex started with the discovery of a mutation in the *SMC5* gene found in a CSR-ID patient, which let us wonder whether Smc5 could be the “missing CSR-specific factor” we were looking for. As no information about the mutation was given, we assessed the Smc5 expression level in human B cell lines and, detecting a variable expression, we silenced *Smc5* in CH12 B cells in order to directly assess its role in CSR. As already mentioned, unlike the experiments performed to characterize Spt6, those on Smc5 took advantage of the improvements of our lentivirus-based knockdown strategy allowing a higher efficiency. Although, we observed a “recovery” of cells upon knockdown which results in the re-expression of Smc5 in silenced lines. We could not assess germline transcription in these conditions, as cell sorting for GFP expression would be required. Nevertheless, upon Smc5 knockdown we observed either a partial impairment or an enhancement of CSR, which was more evident after 72h stimulation, and these results suggested that Smc5 was not required for efficient recombination. However, we cannot exclude that the redundancy of certain factors, able to compensate Smc5 activity, might mask the effect of its depletion on CSR efficiency. Additionally, as Smc5 is in a complex with Smc6, we silenced the latter as well in CH12 B cells for functional characterization. Surprisingly, we observed that Smc6 depletion led to a defect in CSR, consistent after 48h and 72h of stimulation, and although GLT of donor and acceptor S regions must be assessed, this effect is independent of AID as we detected comparable protein levels between the non-target transduced line and the Smc6 knockdown lines. Thus, these results suggest that Smc6, unlike Smc5, might be required for CSR.

The tricky question these experiments rise is: why the two members of the heterodimer seem to display a different behavior? Which properties are harbored by Smc6 which are not shared with Smc5? If we consider their primary structure, Smc family members share the same domain organization and, although phylogenetic studies suggest a divergence between the Smc5/6 proteins and the other Smc family members (Cobbe and Heck, 2004), a sequence-specificity would be unlikely. However, a difference can be identified in the non-Smc element (Nse) accessory proteins binding, as Nse2, a SUMO ligase, has been described to interact specifically with Smc5 (Sergeant et al., 2005). The generation of Nse2/Smc5 double knockout in DT40 cells revealed that Nse2 is required for Smc5 stability but not for Smc5/6 association (Kliszczak et al., 2012). Furthermore, knockdown of Smc5 or MMS21 (human Nse2) in HeLa cells leads to a mitotic defect due to prematurely separated chromosomes, whereas Smc6 knockdown does not exert the same effect (Behlke-Steinert et al., 2009). This suggests that during CSR there might be a functional disconnection between the two members of the heterodimer, which is reflected in a more pronounced effect upon Smc6 knockdown.

Interestingly, although Smc5/6 proteins have been described as essential in yeast, Smc5 knockout in the DT40 B cell line displayed no defects in cell viability but reduced proliferation (Stephan et al., 2011a). However, a recent work from Ju et al. showed that the loss of Smc5 in chicken B cells cannot be compared to the loss of Smc6 in mouse. In the attempt to generate Smc6 knockout mice, they observed that loss of Smc6 was embryonic lethal and thus they generated Smc6 mutant mice harboring the S994A missense mutation (Ju et al., 2013). This mutation is the equivalent of the S1045A Smc6 mutant identified in *S. pombe*, and has been proposed to affect the ATP hydrolysis function of Smc6 and thus disconnecting its role in DNA repair from any additional role, as these yeast mutants display DNA damage sensitivity associated to normal viability (Fousteri and Lehmann, 2000). Smc6 knock-in mice were viable and displayed a minor effect in hematopoiesis; moreover, Ig isotype analysis in blood plasma revealed a reduced amount of IgG2 antibody which suggests a possible role of Smc6 in the immune system (Ju et al., 2013).

Thus, further experiments will be required to delineate the exact function of the Smc5/6 complex in antibody diversification. First, cell proliferation and cell cycle progression has to be assessed upon Smc5 and Smc6 knockdown, especially for the latter, in order to rule out, as we did upon cohesin knockdown, that a proliferation defect might account for the reduction of CSR observed. Furthermore, we need to verify if Smc5 and Smc6 are in the same complex as AID in B cells by co-immunoprecipitation experiments. Moreover, germline transcription in Smc6 knockdown cells must be assessed, to exclude the involvement of reduced transcription in the phenotype observed, and ChIP experiments would reveal whether the Smc5/6 complex is recruited to the IgH locus. By taking into account the phenotype of Smc6 knock-in mice, it would be interesting to better characterize this mouse model and study the ability of primary B cells to undergo CSR to all isotypes *ex vivo*, as well as to assess SHM *in vivo*. However, we need to consider that the characterization of the Smc6 S994A mutant mouse might provide different information compared to a knockout model, and that we cannot exclude that the expression of the protein, although not functional, could partially rescue other additional defects due to a loss of protein-protein interaction. As alternative, Smc5 and Smc6 knockout CH12 B cells could be generated by taking advantage of the clustered, regularly interspaced, short palindromic repeats (CRISPR)-CRISPR-associated protein (Cas) system. Is it possible that Smc6 depletion might affect cell survival, as observed in mice, but, on the other hand, the results obtained with DT40 B cells lacking Smc5 expression encourage to use this strategy and clarify the role of Smc5/6 complex in class switch recombination. Furthermore, as the Smc5/6 complex is required for DNA repair through HR (Kegel and Sjogren, 2010), in order to assess whether Smc6 could cooperate also with the NHEJ repair pathway, junction analysis of S regions in stimulated and Smc6 silenced CH12 B cells would be pivotal to clarify this point.

## 10. The cohesin complex in CSR regulation: long-range interactions, repair or both?

When we analyzed the proteome of CH12 B cells expressing tagged full length AID, we found all three Smc complexes: cohesins (Smc1/3), condensins (Smc2/4) and, as previously discussed, the Smc5/6 complex. These complexes are involved in chromosome dynamics during mitosis and meiosis and in DNA repair and, additionally, cohesins have been described to mediate the long-range interactions occurring respectively at the IgH locus in pro-B cells (Degner et al., 2011) and at the TCR $\alpha$  loci in differentiating T cells (Seitan et al., 2011). Furthermore, interactions between promoters and enhancers have been described as required for CSR, as the conformation of the IgH locus in resting and stimulated B cells undergoes dynamic changes (Kenter et al., 2012). Thus, we decided to assess whether the cohesin complex was involved in CSR regulation. We identified Smc1, Smc3, Nipbl and Wapal as in a complex with AID by co-immunoprecipitation experiments, and these results allowed us to investigate whether cohesins might mediate the interactions between the IgH locus regulatory elements in cells undergoing CSR. By performing ChIP-Seq experiments, we observed that in resting B cells CTCF, Smc1 and Smc3 colocalize at the 3'RR and at the C $\alpha$  exon, while, upon activation, cohesins are dynamically recruited to the S $\mu$ -C $\mu$  region independently on CTCF. These results suggested that cohesins are involved in the long-range interactions occurring in cells poised for recombination, and to assess whether they might play a direct role in regulating CSR we performed their functional characterization in CH12 B cells. We detected a CSR defect for all the subunits of the cohesin complex, and observed that this impairment was AID and GLT-independent in cells with silenced expression of Smc1, Smc3 and Nipbl. Wapal knockdown led, instead, to a reduced transcription at S $\alpha$  region, suggesting that the impairment in GLT might be the cause of defective recombination. Furthermore, as cohesins have been also described as involved in DNA repair (Birkenbihl and Subramani, 1992), we assessed whether the CSR impairment observed upon depletion of Smc1, Smc3, Nipbl and Wapal was due to an aberrant resolution of DSBs generated at the S regions. Thus, by sequencing the S $\mu$ -S $\alpha$  junctions in CH12 B cells undergoing CSR, we observed a bias towards the use of longer microhomologies upon Smc1, Smc3 and Nipbl knockdown. This result suggested the involvement of the A-NHEJ pathway in the repair step of CSR.

However, there are many questions that need to be answered. The first one concerns the looping structure observed at the IgH locus. According to the model proposed by Wuerffel and Kenter (Wuerffel et al., 2007), in resting B cells the E $\mu$  enhancer and the 3'RR are engaged in a loop structure, which undergoes dynamic changes upon B cells activation as the acceptor S region is actively recruited in proximity of E $\mu$  to favor transcription and recombination. However, our data show that cohesins are not involved in the interaction between the 3'RR and E $\mu$ , as no cohesin binding was identified at the 5' of the locus. Moreover, in activated B cells, whereas we detected an active recruitment at the S $\mu$ -C $\mu$  region we did not succeed in identifying cohesins at the acceptor S region S $\gamma$ 1. This leads to two alternatives: a) cohesins are indeed recruited to the acceptor S region but our experimental conditions did not allow us to verify it; b) cohesins are not involved in the recruitment of

the acceptor S region and thus an additional unknown protein or complex might be required for this interaction. To address this point, 4C-Seq experiments on unstimulated and stimulated CH12 cells silenced for the cohesin complex would help to clarify whether there is an interaction and, additionally, if this interaction is disrupted when one of the cohesin complex components is missing.

Interestingly, Potts et al. performed knockdown experiments for Smc5 and MMS21 (Nse2) in HeLa cells and showed impaired recruitment of Smc1 and Scc1 at DSBs, proposing a model in which Smc5/6 complex is involved in the recruitment of Smc1/3 to DSBs (Potts et al., 2006). Although these results are based on the characterization of Smc5 and Nse2 depletion, it would therefore be interesting, once assessed whether Smc6 is recruited to the IgH locus, to compare the binding profile of Smc6 and Smc1/3 by ChIP-Seq.

While performing cohesin knockdown experiments we observed a direct relationship between knockdown efficiency and defect on cell proliferation. This might be explained by the fact that, optimizing the experimental condition, the depletion was more efficient and thus resulted in a lower proliferation rate upon Smc1, Smc3 and Nipbl knockdown. In this case, Wapal depletion did not show the same profile of the other members of the complex, confirming the requirement of cohesins for proper cell division. Interestingly, when we performed cell cycle analysis and verified cell cycle checkpoints activation, we did not observe any significant differences in either unstimulated or stimulated CH12 B cells depleted for cohesins. This result suggests that even residual levels of cohesins are still efficient to ensure the basal cell functions.

Upon cohesins knockdown we observed a similar profile in terms of CSR defect, which was not due to a lower AID expression. However, when we verified S regions germline transcription, we observed again a different profile for Wapal. Depletion of the unloading cohesin factor leads to a reduced transcription at S $\alpha$ , and might suggest that retention of cohesins at the IgH locus could be involved in gene expression regulation. However, Smc1, Smc3 and Nipbl depletion resulted in increased germline transcription at the S regions, suggesting that the cohesin complex is not involved in transcriptional regulation at these loci. Is it possible, however, that the effect observed upon Wapal depletion can be due to some off-target effects exerted by the shRNA.

Moreover, Wapal depletion had no significant effect in the DSBs repair, suggesting, in this case, that the presence of the cohesin complex is allowing the proper resolution of the breaks through the C-NHEJ pathway. As the initial steps of AID-mediated DSBs formation depend on UNG and MSH2/6, it would be interesting to assess the expression level of these proteins upon Smc1, Smc3 and Nipbl knockdown. Moreover, cohesins have been proposed to recruit 53BP1 to DSBs (Watrin and Peters, 2009) and Smc1 has been described as involved in both HR and NHEJ repair pathways (Schar et al., 2004). As 53BP1 deficiency results in increased intra-switch recombination between S $\mu$  and S $\gamma$ 1 regions (Reina-San-Martin et al., 2007), and recent reports propose 53BP1 as crucial in the choice between the C-NHEJ and the A-NHEJ which favors end resection and intra-switch recombination (Bothmer et al., 2010), it could be possible that cohesins recruit 53BP1 at the damaged DNA during CSR and thus indirectly influence the choice of the DNA repair pathway during CSR. As Rad51 is involved in end resection, to address this point we could perform Rad51 ChIP upon cohesin knockdown to verify whether depletion of cohesins favors Rad51 recruitment at the IgH locus and the



alternative repair pathway. Thus, our results suggest that cohesin complex might regulate CSR at two levels: by modifying the 3D structure of the IgH locus and by influencing the DNA repair occurring at DNA breaks.

## Working model for CSR

The mechanisms that regulate class switch recombination have been the object of intensive studies during these last years. Dissection of the sequential steps of the reaction, as well as the tight recruitment of AID to the S regions and the interplay between recombination and transcription is pivotal to clarify the mutagenic activity of AID, in light of its off-target activity which leads to pathological consequences.

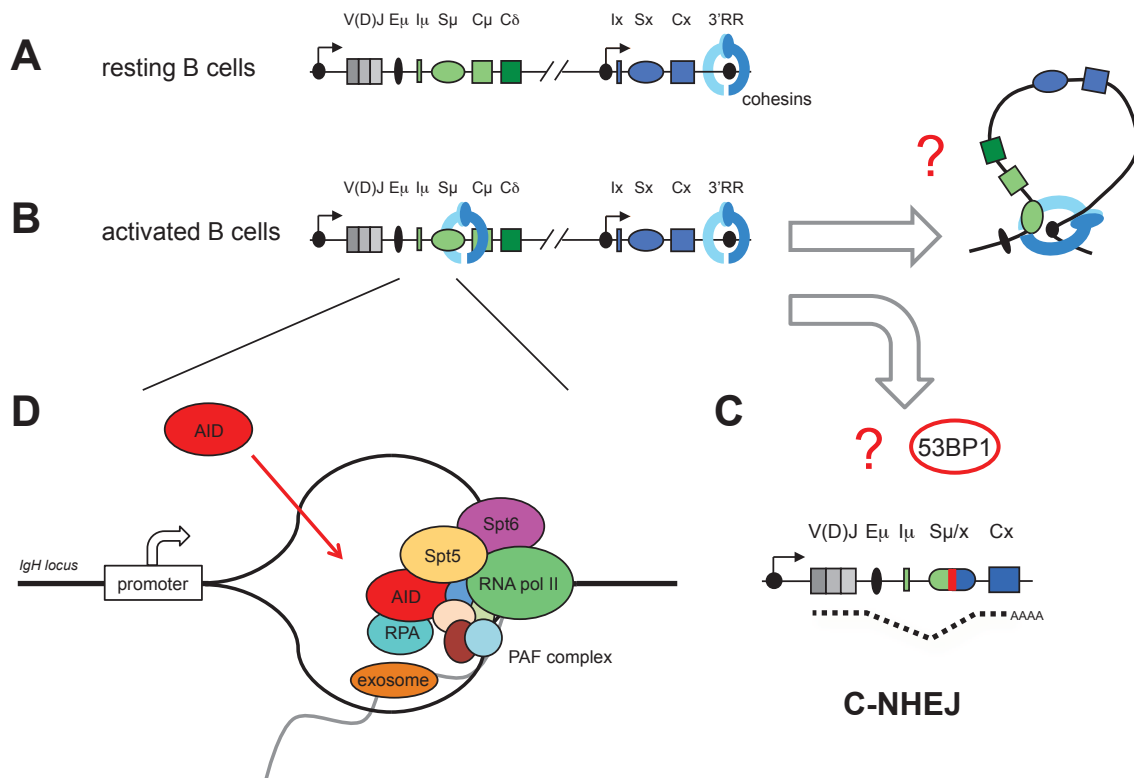
With this study, we provided new insights within the long-range rearrangements occurring at the IgH locus as well as AID interaction with factors involved in transcription and chromatin remodeling. By focusing on CSR-specific cofactors through the analysis of B cells isolated from CSR-ID patients harboring a specific defect in CSR and mouse B cells overexpressing AID or its C-terminal domain specifically, we found the cohesin complex, the Smc5/6 complex and Spt6.

By focusing on the IgH locus organization, we show that the cohesin complex is actively recruited to the IgH locus upon activation, and Smc1/3 occupancy at the 3'RR and at the S<sub>μ</sub>-C<sub>μ</sub> region suggests an interaction between these two regions of the locus, although this hypothesis will be confirmed by 4C-Seq experiments. Moreover, loss of Smc1, Smc3 and of the loading factor Nipbl results in impaired class switching, which is independent on AID expression and donor and acceptor germline transcription, showing the requirement of these proteins for proper recombination during antibody diversification. The analysis of switch junctions reveals that Smc1, Smc3 and Nipbl depletion leads to a microhomology-based repair of AID-induced DSBs, involving cohesins in the choice between the C-NHEJ or A-NHEJ pathway.

If on one hand the cohesin complex appears as regulating the global structure of the locus and the outcome of recombination, on the other hand our investigations focused on Spt6 and the PAF complex contribute to delineate the dynamics of AID association to the S regions. We show that AID associates with Spt5, Spt6, the PAF complex and the RNA polymerase II, and that the specific interaction between AID and PAF is mediated by the Paf1 subunit of the complex. Moreover, CH12 B cell depleted of Leo1 display impairment in CSR, which is not due to a lower AID expression nor to an impaired S regions transcription. Furthermore, Leo1 depletion impairs AID recruitment to the S<sub>μ</sub> region.

We also show that the Smc5/6 complex is required for CSR, and that a prominent effect on recombination is observed upon Smc6 depletion in CH12 B cells. Whether this complex acts in association to cohesins or is involved in DNA repair, will be clarified by further investigations.

Thus, we propose a model where cohesins are actively recruited to the donor S region upon activation (Figure 33B), and determine the outcome of recombination by favoring DSBs repair through NHEJ, whose hallmark are short microhomologies at the junctions (Figure 33C). Moreover, we propose that AID is retained at the S regions through a complex including the RNA polymerase II and Spt5 as well as Spt6 and the PAF complex (Figure 33D), adding additional pieces to the puzzle of AID regulation in antibody diversification.



**Figure 33. Working model for CSR**

With our study, we show that **(A)** the cohesin complex is present at the 3'RR of the IgH locus in resting B cells and **(B)** is actively recruited to the S $\mu$ -C $\mu$  region upon activation; this might result in a long-range interaction between the S region and the 3'RR, and further experiments will clarify this point. Moreover, **(C)** cohesins are required for DNA repair of AID-induced DSBs through C-NHEJ pathway whose hallmark are short microhomologies at the S junctions, and this regulation might be mediated by 53BP1 which inhibits end resection and microhomology-based repair through A-NHEJ. **(D)** The identification of CSR-specific factors as Spt6 and the contribution to the PAF complex characterization show that AID is present at the S regions in a complex that includes the RNA polymerase II, the elongation factor Spt5 as well as Spt6 and the PAF complex, linking transcription and chromatin modifications as concerted regulation of AID-mediated deamination.

## **General conclusions**

AID was discovered more than ten years ago but, in spite of the progresses made in the understanding of its role and functions, there are still opened questions that need an answer. With the aim to better understand the regulation occurring during class switching, and the specificity of AID targeting to the Ig loci, we focused on its CSR-specific partners. The collaboration with other laboratories allowed us to identify new molecular players required for different steps of the recombination reaction. Thus, understanding the dynamics of AID regulation will provide new insights in antibody diversification mechanisms, and our results provide a base for further investigations.

## **LITERATURE CITED**

## LITERATURE CITED

- Adelman, K., Wei, W., Ardehali, M.B., Werner, J., Zhu, B., Reinberg, D., and Lis, J.T. (2006). *Drosophila* Paf1 modulates chromatin structure at actively transcribed genes. *Molecular and cellular biology* 26, 250-260.
- Adkins, M.W., and Tyler, J.K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Mol Cell* 21, 405-416.
- Agematsu, K., Nagumo, H., Shinozaki, K., Hokibara, S., Yasui, K., Terada, K., Kawamura, N., Toba, T., Nonoyama, S., Ochs, H.D., *et al.* (1998). Absence of IgD-CD27(+) memory B cell population in X-linked hyper-IgM syndrome. *J Clin Invest* 102, 853-860.
- Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 46, 115-124.
- Allen, R.C., Armitage, R.J., Conley, M.E., Rosenblatt, H., Jenkins, N.A., Copeland, N.G., Bedell, M.A., Edelhoff, S., Distèche, C.M., Simoneaux, D.K., *et al.* (1993). CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. *Science* 259, 990-993.
- Alt, F.W., and Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proceedings of the National Academy of Sciences of the United States of America* 79, 4118-4122.
- 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. *The EMBO journal* 3, 1209-1219.
- Anderson, L., Henderson, C., and Adachi, Y. (2001). Phosphorylation and rapid relocalization of 53BP1 to nuclear foci upon DNA damage. *Molecular and cellular biology* 21, 1719-1729.
- Andrews, E.A., Palecek, J., Sergeant, J., Taylor, E., Lehmann, A.R., and Watts, F.Z. (2005). Nse2, a component of the Smc5-6 complex, is a SUMO ligase required for the response to DNA damage. *Molecular and cellular biology* 25, 185-196.
- Andrulis, E.D., Guzman, E., Doring, P., Werner, J., and Lis, J.T. (2000). High-resolution localization of *Drosophila* Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev* 14, 2635-2649.
- Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., and Lis, J.T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* 420, 837-841.
- Aoufouchi, S., Faili, A., Zober, C., D'Orlando, O., Weller, S., Weill, J.C., and Reynaud, C.A. (2008). Proteasomal degradation restricts the nuclear lifespan of AID. *The Journal of experimental medicine* 205, 1357-1368.
- Arakawa, H., Hauschild, J., and Buerstedde, J.M. (2002). Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* 295, 1301-1306.
- Ardehali, M.B., Yao, J., Adelman, K., Fuda, N.J., Petesch, S.J., Webb, W.W., and Lis, J.T. (2009). Spt6 enhances the elongation rate of RNA polymerase II in vivo. *The EMBO journal* 28, 1067-1077.
- Aruffo, A., Farrington, M., Hollenbaugh, D., Li, X., Milatovich, A., Nonoyama, S., Bajorath, J., Grosmaire, L.S., Stenkamp, R., Neubauer, M., *et al.* (1993). The CD40 ligand, gp39, is defective in activated T cells from patients with X-linked hyper-IgM syndrome. *Cell* 72, 291-300.
- Arumugam, P., Gruber, S., Tanaka, K., Haering, C.H., Mechtler, K., and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. *Curr Biol* 13, 1941-1953.
- Audebert, M., Salles, B., and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *The Journal of biological chemistry* 279, 55117-55126.
- Baniahmad, C., Nawaz, Z., Baniahmad, A., Gleeson, M.A., Tsai, M.J., and O'Malley, B.W. (1995). Enhancement of human estrogen receptor activity by SPT6: a potential coactivator. *Mol Endocrinol* 9, 34-43.
- Bardwell, P.D., Woo, C.J., Wei, K., Li, Z., Martin, A., Sack, S.Z., Parris, T., Edelmann, W., and Scharff, M.D. (2004). Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. *Nat Immunol* 5, 224-229.
- Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M., and Nussenzweig, M.C. (2003). C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 12, 501-508.

- Basu, U., Chaudhuri, J., Alpert, C., Dutt, S., Ranganath, S., Li, G., Schrum, J.P., Manis, J.P., and Alt, F.W. (2005). The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature*.
- 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.
- Beckouet, F., Hu, B., Roig, M.B., Sutani, T., Komata, M., Uluocak, P., Katis, V.L., Shirahige, K., and Nasmyth, K. (2010). An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. *Mol Cell* **39**, 689-699.
- Begum, N.A., Kinoshita, K., Muramatsu, M., Nagaoka, H., Shinkura, R., and Honjo, T. (2004). De novo protein synthesis is required for activation-induced cytidine deaminase-dependent DNA cleavage in immunoglobulin class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 13003-13007.
- Behlke-Steinert, S., Touat-Todeschini, L., Skoufias, D.A., and Margolis, R.L. (2009). SMC5 and MMS21 are required for chromosome cohesion and mitotic progression. *Cell Cycle* **8**, 2211-2218.
- Bertocci, B., Quint, L., Delbos, F., Garcia, C., Reynaud, C.A., and Weill, J.C. (1998). Probing immunoglobulin gene hypermutation with microsatellites suggests a nonreplicative short patch DNA synthesis process. *Immunity* **9**, 257-265.
- Bertolino, E., Reddy, K., Medina, K.L., Parganas, E., Ihle, J., and Singh, H. (2005). Regulation of interleukin 7-dependent immunoglobulin heavy-chain variable gene rearrangements by transcription factor STAT5. *Nat Immunol* **6**, 836-843.
- Betz, A.G., Rada, C., Pannell, R., Milstein, C., and Neuberger, M.S. (1993). Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 2385-2388.
- Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042-1047.
- Bhutani, N., Decker, M.N., Brady, J.J., Bussat, R.T., Burns, D.M., Corbel, S.Y., and Blau, H.M. (2013). A critical role for AID in the initiation of reprogramming to induced pluripotent stem cells. *FASEB J* **27**, 1107-1113.
- Birkenbihl, R.P., and Subramani, S. (1992). Cloning and characterization of rad21 an essential gene of *Schizosaccharomyces pombe* involved in DNA double-strand-break repair. *Nucleic acids research* **20**, 6605-6611.
- Boboila, C., Alt, F.W., and Schwer, B. (2012a). Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. *Advances in immunology* **116**, 1-49.
- Boboila, C., Oksenysh, V., Gostissa, M., Wang, J.H., Zha, S., Zhang, Y., Chai, H., Lee, C.S., Jankovic, M., Saez, L.M., *et al.* (2012b). Robust chromosomal DNA repair via alternative end-joining in the absence of X-ray repair cross-complementing protein 1 (XRCC1). *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2473-2478.
- 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. *The Journal of experimental medicine* **207**, 417-427.
- Bolland, D.J., Wood, A.L., Afshar, R., Featherstone, K., Oltz, E.M., and Corcoran, A.E. (2007). Antisense intergenic transcription precedes Igh D-to-J recombination and is controlled by the intronic enhancer Emu. *Molecular and cellular biology* **27**, 5523-5533.
- 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.
- Borchert, G.M., Holton, N.W., and Larson, E.D. (2011). Repression of human activation induced cytidine deaminase by miR-93 and miR-155. *BMC Cancer* **11**, 347.
- Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Rolef Ben-Shahar, T., and Uhlmann, F. (2010). Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. *Mol Cell* **39**, 677-688.
- Bortvin, A., and Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science* **272**, 1473-1476.
- Bose, T., and Gerton, J.L. (2010). Cohesinopathies, gene expression, and chromatin organization. *J Cell Biol* **189**, 201-210.

- Bosma, G.C., Kim, J., Urich, T., Fath, D.M., Cotticelli, M.G., Ruetsch, N.R., Radic, M.Z., and Bosma, M.J. (2002). DNA-dependent protein kinase activity is not required for immunoglobulin class switching. *The Journal of experimental medicine* 196, 1483-1495.
- Bothmer, A., Robbiani, D.F., Di Virgilio, M., Bunting, S.F., Klein, I.A., Feldhahn, N., Barlow, J., Chen, H.T., Bosque, D., Callen, E., *et al.* (2011). Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1. *Mol Cell* 42, 319-329.
- Bothmer, A., Robbiani, D.F., Feldhahn, N., Gazumyan, A., Nussenzweig, A., and Nussenzweig, M.C. (2010). 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *The Journal of experimental medicine* 207, 855-865.
- Bothmer, A., Rommel, P.C., Gazumyan, A., Polato, F., Reczek, C.R., Muellenbeck, M.F., Schaezlein, S., Edelmann, W., Chen, P.L., Brosh, R.M., Jr., *et al.* (2013). Mechanism of DNA resection during intrachromosomal recombination and immunoglobulin class switching. *The Journal of experimental medicine* 210, 115-123.
- Bottaro, A., Lansford, R., Xu, L., Zhang, J., Rothman, P., and Alt, F.W. (1994). S region transcription per se promotes basal IgE class switch recombination but additional factors regulate the efficiency of the process. *The EMBO journal* 13, 665-674.
- 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. *Proceedings of the National Academy of Sciences of the United States of America*.
- Brar, S.S., Watson, M., and Diaz, M. (2004). Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *The Journal of biological chemistry* 279, 26395-26401.
- Breitfeld, D., Ohl, L., Kremmer, E., Ellwart, J., Sallusto, F., Lipp, M., and Forster, R. (2000). Follicular B helper T cells express CXC chemokine receptor 5, localize to B cell follicles, and support immunoglobulin production. *The Journal of experimental medicine* 192, 1545-1552.
- Burma, S., Chen, B.P., Murphy, M., Kurimasa, A., and Chen, D.J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *The Journal of biological chemistry* 276, 42462-42467.
- Busslinger, M. (2004). Transcriptional control of early B cell development. *Annu Rev Immunol* 22, 55-79.
- Caratao, N., Cortesao, C.S., Reis, P.H., Freitas, R.F., Jacob, C.M., Pastorino, A.C., Carneiro-Sampaio, M., and Barreto, V.M. (2013). A novel activation-induced cytidine deaminase (AID) mutation in Brazilian patients with hyper-IgM type 2 syndrome. *Clin Immunol* 148, 279-286.
- Carpten, J.D., Robbins, C.M., Villablanca, A., Forsberg, L., Presciuttini, S., Bailey-Wilson, J., Simonds, W.F., Gillanders, E.M., Kennedy, A.M., Chen, J.D., *et al.* (2002). HRPT2, encoding parafibromin, is mutated in hyperparathyroidism-jaw tumor syndrome. *Nat Genet* 32, 676-680.
- Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592.
- Casellas, R., Nussenzweig, A., Wuerrfel, R., Pelanda, R., Reichlin, A., Suh, H., Qin, X.F., Besmer, E., Kenter, A., Rajewsky, K., *et al.* (1998). Ku80 is required for immunoglobulin isotype switching. *The EMBO journal* 17, 2404-2411.
- Casellas, R., Yamane, A., Kovalchuk, A.L., and Potter, M. (2009). Restricting activation-induced cytidine deaminase tumorigenic activity in B lymphocytes. *Immunology* 126, 316-328.
- Castle, B.E., Kishimoto, K., Stearns, C., Brown, M.L., and Kehry, M.R. (1993). Regulation of expression of the ligand for CD40 on T helper lymphocytes. *Journal of immunology* 151, 1777-1788.
- Catalan, N., Selz, F., Imai, K., Revy, P., Fischer, A., and Durandy, A. (2003). The block in immunoglobulin class switch recombination caused by activation-induced cytidine deaminase deficiency occurs prior to the generation of DNA double strand breaks in switch mu region. *Journal of immunology* 171, 2504-2509.
- Cesare, A.J., and Reddel, R.R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* 11, 319-330.
- Chahwan, R., Edelmann, W., Scharff, M.D., and Roa, S. (2012). AIDing antibody diversity by error-prone mismatch repair. *Semin Immunol* 24, 293-300.
- Chakraborty, T., Chowdhury, D., Keyes, A., Jani, A., Subrahmanyam, R., Ivanova, I., and Sen, R. (2007). Repeat organization and epigenetic regulation of the DH-Cmu domain of the immunoglobulin heavy-chain gene locus. *Mol Cell* 27, 842-850.

- Chatterji, M., Unniraman, S., McBride, K.M., and Schatz, D.G. (2007). Role of activation-induced deaminase protein kinase A phosphorylation sites in Ig gene conversion and somatic hypermutation. *Journal of immunology* 179, 5274-5280.
- Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 4, 541-552.
- Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan, R.T., Datta, A., *et al.* (2007). Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Advances in immunology* 94, 157-214.
- Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 430, 992-998.
- Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422, 726-730.
- Chen, Y., Yamaguchi, Y., Tsugeno, Y., Yamamoto, J., Yamada, T., Nakamura, M., Hisatake, K., and Handa, H. (2009). DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes Dev* 23, 2765-2777.
- Chester, A., Somasekaram, A., Tzimina, M., Jarmuz, A., Gisbourne, J., O'Keefe, R., Scott, J., and Navaratnam, N. (2003). The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *The EMBO journal* 22, 3971-3982.
- Cheung, V., Chua, G., Batada, N.N., Landry, C.R., Michnick, S.W., Hughes, T.R., and Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. *PLoS Biol* 6, e277.
- Chevillard, C., Ozaki, J., Herring, C.D., and Riblet, R. (2002). A three-megabase yeast artificial chromosome contig spanning the C57BL mouse Igh locus. *Journal of immunology* 168, 5659-5666.
- Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., *et al.* (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 147, 107-119.
- Chowdhury, D., and Sen, R. (2003). Transient IL-7/IL-7R signaling provides a mechanism for feedback inhibition of immunoglobulin heavy chain gene rearrangements. *Immunity* 18, 229-241.
- Chowdhury, M., Forouhi, O., Dayal, S., McCloskey, N., Gould, H.J., Felsenfeld, G., and Fear, D.J. (2008). Analysis of intergenic transcription and histone modification across the human immunoglobulin heavy-chain locus. *Proceedings of the National Academy of Sciences of the United States of America* 105, 15872-15877.
- Chu, Y., Simic, R., Warner, M.H., Arndt, K.M., and Prelich, G. (2007). Regulation of histone modification and cryptic transcription by the Bur1 and Paf1 complexes. *The EMBO journal* 26, 4646-4656.
- Clark-Adams, C.D., and Winston, F. (1987). The SPT6 gene is essential for growth and is required for delta-mediated transcription in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 7, 679-686.
- Close, D., Johnson, S.J., Sdano, M.A., McDonald, S.M., Robinson, H., Formosa, T., and Hill, C.P. (2011). Crystal structures of the *S. cerevisiae* Spt6 core and C-terminal tandem SH2 domain. *J Mol Biol* 408, 697-713.
- Cobb, R.M., Oestreich, K.J., Osipovich, O.A., and Oltz, E.M. (2006). Accessibility control of V(D)J recombination. *Advances in immunology* 91, 45-109.
- Cobbe, N., and Heck, M.M. (2004). The evolution of SMC proteins: phylogenetic analysis and structural implications. *Mol Biol Evol* 21, 332-347.
- Compagnone-Post, P.A., and Osley, M.A. (1996). Mutations in the SPT4, SPT5, and SPT6 genes alter transcription of a subset of histone genes in *Saccharomyces cerevisiae*. *Genetics* 143, 1543-1554.
- Coticello, S.G., Ganesh, K., Xue, K., Lu, M., Rada, C., and Neuberger, M.S. (2008). Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNBL1. *Mol Cell* 31, 474-484.
- Coticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., and Neuberger, M.S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22, 367-377.
- Cook, A.J., Oganessian, L., Harumal, P., Basten, A., Brink, R., and Jolly, C.J. (2003). Reduced switching in SCID B cells is associated with altered somatic mutation of recombined S regions. *Journal of immunology* 171, 6556-6564.



- Corcoran, A.E. (2010). The epigenetic role of non-coding RNA transcription and nuclear organization in immunoglobulin repertoire generation. *Semin Immunol* 22, 353-361.
- Crouch, E.E., Li, Z., Takizawa, M., Fichtner-Feigl, S., Gourzi, P., Montano, C., Feigenbaum, L., Wilson, P., Janz, S., Papavasiliou, F.N., *et al.* (2007). Regulation of AID expression in the immune response. *The Journal of experimental medicine* 204, 1145-1156.
- Daniel, J.A., Santos, M.A., Wang, Z., Zang, C., Schwab, K.R., Jankovic, M., Filsuf, D., Chen, H.T., Gazumyan, A., Yamane, A., *et al.* (2010). PTIP promotes chromatin changes critical for immunoglobulin class switch recombination. *Science* 329, 917-923.
- Daniels, G.A., and Lieber, M.R. (1995a). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic acids research* 23, 5006-5011.
- Daniels, G.A., and Lieber, M.R. (1995b). Strand specificity in the transcriptional targeting of recombination at immunoglobulin switch sequences. *Proceedings of the National Academy of Sciences of the United States of America* 92, 5625-5629.
- de Miranda, N.F., Bjorkman, A., and Pan-Hammarstrom, Q. (2011). DNA repair: the link between primary immunodeficiency and cancer. *Annals of the New York Academy of Sciences* 1246, 50-63.
- De Vos, M., Hayward, B.E., Charlton, R., Taylor, G.R., Glaser, A.W., Picton, S., Cole, T.R., Maher, E.R., McKeown, C.M., Mann, J.R., *et al.* (2006). PMS2 mutations in childhood cancer. *J Natl Cancer Inst* 98, 358-361.
- de Yebenes, V.G., Bartolome-Izquierdo, N., and Ramiro, A.R. (2013). Regulation of B-cell development and function by microRNAs. *Immunol Rev* 253, 25-39.
- de Yebenes, V.G., Belver, L., Pisano, D.G., Gonzalez, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *The Journal of experimental medicine* 205, 2199-2206.
- Deardorff, M.A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodriguez, C., Arnedo, M., Loeys, B., Kline, A.D., *et al.* (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J Hum Genet* 80, 485-494.
- Deardorff, M.A., Wilde, J.J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Monnich, M., Yan, Y., Xu, W., Gil-Rodriguez, M.C., *et al.* (2012). RAD21 mutations cause a human cohesinopathy. *Am J Hum Genet* 90, 1014-1027.
- Dedeoglu, F., Horwitz, B., Chaudhuri, J., Alt, F.W., and Geha, R.S. (2004). Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. *Int Immunol* 16, 395-404.
- Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., *et al.* (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9566-9571.
- Degner, S.C., Wong, T.P., Jankevicius, G., and Feeney, A.J. (2009). Cutting edge: developmental stage-specific recruitment of cohesin to CTCF sites throughout immunoglobulin loci during B lymphocyte development. *Journal of immunology* 182, 44-48.
- Delbos, F., Aoufouchi, S., Faili, A., Weill, J.C., and Reynaud, C.A. (2007). DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. *The Journal of experimental medicine* 204, 17-23.
- Delker, R.K., Zhou, Y., Strikoudis, A., Stebbins, C.E., and Papavasiliou, F.N. (2013). Solubility-based genetic screen identifies RING finger protein 126 as an E3 ligase for activation-induced cytidine deaminase. *Proceedings of the National Academy of Sciences of the United States of America* 110, 1029-1034.
- Dengl, S., Mayer, A., Sun, M., and Cramer, P. (2009). Structure and in vivo requirement of the yeast Spt6 SH2 domain. *J Mol Biol* 389, 211-225.
- Di Noia, J.M., and Neuberger, M.S. (2007). Molecular Mechanisms of Antibody Somatic Hypermutation. *Annu Rev Biochem* 76, 1-22.
- Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A.D., Feldhahn, N., Resch, W., Oliveira, T.Y., Chait, B.T., *et al.* (2013). Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. *Science* 339, 711-715.
- Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (2003). AID mediates hypermutation by deaminating single stranded DNA. *The Journal of experimental medicine* 197, 1291-1296.

- Diebold, M.L., Koch, M., Loeliger, E., Cura, V., Winston, F., Cavarelli, J., and Romier, C. (2010a). The structure of an Iws1/Spt6 complex reveals an interaction domain conserved in TFIIS, Elongin A and Med26. *The EMBO journal* **29**, 3979-3991.
- Diebold, M.L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J., and Romier, C. (2010b). Noncanonical tandem SH2 enables interaction of elongation factor Spt6 with RNA polymerase II. *The Journal of biological chemistry* **285**, 38389-38398.
- Dinkelmann, M., Spehalski, E., Stoneham, T., Buis, J., Wu, Y., Sekiguchi, J.M., and Ferguson, D.O. (2009). Multiple functions of MRN in end-joining pathways during isotype class switching. *Nat Struct Mol Biol* **16**, 808-813.
- DiSanto, J.P., Bonnefoy, J.Y., Gauchat, J.F., Fischer, A., and de Saint Basile, G. (1993). CD40 ligand mutations in x-linked immunodeficiency with hyper-IgM. *Nature* **361**, 541-543.
- Doffinger, R., Smahi, A., Bessia, C., Geissmann, F., Feinberg, J., Durandy, A., Bodemer, C., Kenwick, S., Dupuis-Girod, S., Blanche, S., *et al.* (2001). X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF-kappaB signaling. *Nat Genet* **27**, 277-285.
- Doi, T., Kinoshita, K., Ikegawa, M., Muramatsu, M., and Honjo, T. (2003). De novo protein synthesis is required for the activation-induced cytidine deaminase function in class-switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 2634-2638.
- Doil, C., Mailand, N., Bekker-Jensen, S., Menard, P., Larsen, D.H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., *et al.* (2009). RNF168 binds and amplifies ubiquitin conjugates on damaged chromosomes to allow accumulation of repair proteins. *Cell* **136**, 435-446.
- Dorsett, D., and Strom, L. (2012). The ancient and evolving roles of cohesin in gene expression and DNA repair. *Curr Biol* **22**, R240-250.
- Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.H., Robbiani, D.F., Di Virgilio, M., San-Martin, B.R., Heidkamp, G., Schwickert, T.A., *et al.* (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* **28**, 630-638.
- Doyle, J.M., Gao, J., Wang, J., Yang, M., and Potts, P.R. (2010). MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. *Mol Cell* **39**, 963-974.
- Du, L., Peng, R., Bjorkman, A., Filipe de Miranda, N., Rosner, C., Kotnis, A., Berglund, M., Liu, C., Rosenquist, R., Enblad, G., *et al.* (2012). Cernunnos influences human immunoglobulin class switch recombination and may be associated with B cell lymphomagenesis. *The Journal of experimental medicine* **209**, 291-305.
- Du, L., van der Burg, M., Popov, S.W., Kotnis, A., van Dongen, J.J., Gennery, A.R., and Pan-Hammarstrom, Q. (2008). Involvement of Artemis in nonhomologous end-joining during immunoglobulin class switch recombination. *The Journal of experimental medicine* **205**, 3031-3040.
- Durandy, A., Schiff, C., Bonnefoy, J.Y., Forveille, M., Rousset, F., Mazzei, G., Milili, M., and Fischer, A. (1993). Induction by anti-CD40 antibody or soluble CD40 ligand and cytokines of IgG, IgA and IgE production by B cells from patients with X-linked hyper IgM syndrome. *Eur J Immunol* **23**, 2294-2299.
- Durandy, A., Taubenheim, N., Peron, S., and Fischer, A. (2007). Pathophysiology of B-cell intrinsic immunoglobulin class switch recombination deficiencies. *Advances in immunology* **94**, 275-306.
- Ehrenstein, M.R., and Neuberger, M.S. (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. *The EMBO journal* **18**, 3484-3490.
- Ehrenstein, M.R., Rada, C., Jones, A.M., Milstein, C., and Neuberger, M.S. (2001). Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14553-14558.
- Eitoku, M., Sato, L., Senda, T., and Horikoshi, M. (2008). Histone chaperones: 30 years from isolation to elucidation of the mechanisms of nucleosome assembly and disassembly. *Cell Mol Life Sci* **65**, 414-444.
- Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., *et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Molecular and cellular biology* **24**, 3324-3336.
- Estruch, F., Peiro-Chova, L., Gomez-Navarro, N., Durban, J., Hodge, C., Del Olmo, M., and Cole, C.N. (2009). A genetic screen in *Saccharomyces cerevisiae* identifies new genes that interact with

- mex67-5, a temperature-sensitive allele of the gene encoding the mRNA export receptor. *Mol Genet Genomics* **281**, 125-134.
- Falck, J., Coates, J., and Jackson, S.P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature* **434**, 605-611.
- Fay, A., Misulovin, Z., Li, J., Schaaf, C.A., Gause, M., Gilmour, D.S., and Dorsett, D. (2011). Cohesin selectively binds and regulates genes with paused RNA polymerase. *Curr Biol* **21**, 1624-1634.
- Feeney, K.M., Wasson, C.W., and Parish, J.L. (2010). Cohesin: a regulator of genome integrity and gene expression. *Biochem J* **428**, 147-161.
- Ferrari, S., Giliani, S., Insalaco, A., Al-Ghonaium, A., Soresina, A.R., Loubser, M., Avanzini, M.A., Marconi, M., Badolato, R., Ugazio, A.G., *et al.* (2001). Mutations of CD40 gene cause an autosomal recessive form of immunodeficiency with hyper IgM. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12614-12619.
- Fournier, M.L., Gilmore, J.M., Martin-Brown, S.A., and Washburn, M.P. (2007). Multidimensional separations-based shotgun proteomics. *Chem Rev* **107**, 3654-3686.
- Fousteri, M.I., and Lehmann, A.R. (2000). A novel SMC protein complex in *Schizosaccharomyces pombe* contains the Rad18 DNA repair protein. *The EMBO journal* **19**, 1691-1702.
- Franco, S., Gostissa, M., Zha, S., Lombard, D.B., Murphy, M.M., Zarrin, A.A., Yan, C., Tepsuporn, S., Morales, J.C., Adams, M.M., *et al.* (2006). H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol Cell* **21**, 201-214.
- Frank, K.M., Sekiguchi, J.M., Seidl, K.J., Swat, W., Rathbun, G.A., Cheng, H.L., Davidson, L., Kangaloo, L., and Alt, F.W. (1998). Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173-177.
- Frey, S., Bertocci, B., Delbos, F., Quint, L., Weill, J.C., and Reynaud, C.A. (1998). Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. *Immunity* **9**, 127-134.
- Fritz, E.L., Rosenberg, B.R., Lay, K., Mihailovic, A., Tuschl, T., and Papavasiliou, F.N. (2013). A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat Immunol* **14**, 749-755.
- Fuleihan, R., Ramesh, N., Loh, R., Jabara, H., Rosen, R.S., Chatila, T., Fu, S.M., Stamenkovic, I., and Geha, R.S. (1993). Defective expression of the CD40 ligand in X chromosome-linked immunoglobulin deficiency with normal or elevated IgM. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 2170-2173.
- Fuxa, M., Skok, J., Souabni, A., Salvaggio, G., Roldan, E., and Busslinger, M. (2004). Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev* **18**, 411-422.
- Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fujiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J., *et al.* (1998). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**, 891-902.
- Gazumyan, A., Timachova, K., Yuen, G., Siden, E., Di Virgilio, M., Woo, E.M., Chait, B.T., Reina San-Martin, B., Nussenzweig, M.C., and McBride, K.M. (2011). Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. *Molecular and cellular biology* **31**, 442-449.
- Geisberger, R., Rada, C., and Neuberger, M.S. (2009). The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 6736-6741.
- Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* **71**, 101-132.
- Goldfarb, A.N., Flores, J.P., and Lewandowska, K. (1996). Involvement of the E2A basic helix-loop-helix protein in immunoglobulin heavy chain class switching. *Mol Immunol* **33**, 947-956.
- Gonda, H., Sugai, M., Nambu, Y., Katakai, T., Agata, Y., Mori, K.J., Yokota, Y., and Shimizu, A. (2003). The balance between Pax5 and Id2 activities is the key to AID gene expression. *The Journal of experimental medicine* **198**, 1427-1437.
- Gordon, M.S., Kanegai, C.M., Doerr, J.R., and Wall, R. (2003). Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalpha, CD79a). *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4126-4131.
- Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., and Lieber, M.R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**, 492-495.

- Guacci, V., Koshland, D., and Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. *Cell* **91**, 47-57.
- Haering, C.H., Lowe, J., Hochwagen, A., and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* **9**, 773-788.
- Hagstrom, K.A., and Meyer, B.J. (2003). Condensin and cohesin: more than chromosome compactor and glue. *Nat Rev Genet* **4**, 520-534.
- Han, L., Mao, W., and Yu, K. (2012). X-ray repair cross-complementing protein 1 (XRCC1) deficiency enhances class switch recombination and is permissive for alternative end joining. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 4604-4608.
- Han, L., Masani, S., and Yu, K. (2010). Cutting edge: CTNNB1 is dispensable for Ig class switch recombination. *Journal of immunology* **185**, 1379-1381.
- Hanson, E.P., Monaco-Shawver, L., Solt, L.A., Madge, L.A., Banerjee, P.P., May, M.J., and Orange, J.S. (2008). Hypomorphic nuclear factor-kappaB essential modulator mutation database and reconstitution system identifies phenotypic and immunologic diversity. *J Allergy Clin Immunol* **122**, 1169-1177 e1116.
- Harper, J.W., and Elledge, S.J. (2007). The DNA damage response: ten years after. *Mol Cell* **28**, 739-745.
- Harriman, G.R., Bradley, A., Das, S., Rogers-Fani, P., and Davis, A.C. (1996). IgA class switch in I alpha exon-deficient mice. Role of germline transcription in class switch recombination. *J Clin Invest* **97**, 477-485.
- Harris, R.S., Sale, J.E., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). AID is essential for immunoglobulin V gene conversion in a cultured B cell line. *Curr Biol* **12**, 435-438.
- Hasler, J., Rada, C., and Neuberger, M.S. (2011). Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A). *Proceedings of the National Academy of Sciences of the United States of America*.
- He, B., Xu, W., Santini, P.A., Polydorides, A.D., Chiu, A., Estrella, J., Shan, M., Chadburn, A., Villanacci, V., Plebani, A., *et al.* (2007). Intestinal bacteria trigger T cell-independent immunoglobulin A(2) class switching by inducing epithelial-cell secretion of the cytokine APRIL. *Immunity* **26**, 812-826.
- Hein, K., Lorenz, M.G., Siebenkotten, G., Petry, K., Christine, R., and Radbruch, A. (1998). Processing of switch transcripts is required for targeting of antibody class switch recombination. *The Journal of experimental medicine* **188**, 2369-2374.
- Hesslein, D.G., and Schatz, D.G. (2001). Factors and forces controlling V(D)J recombination. *Advances in immunology* **78**, 169-232.
- Hirano, M., Anderson, D.E., Erickson, H.P., and Hirano, T. (2001). Bimodal activation of SMC ATPase by intra- and inter-molecular interactions. *The EMBO journal* **20**, 3238-3250.
- Hirano, M., and Hirano, T. (1998). ATP-dependent aggregation of single-stranded DNA by a bacterial SMC homodimer. *The EMBO journal* **17**, 7139-7148.
- Hirano, M., and Hirano, T. (2002). Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. *The EMBO journal* **21**, 5733-5744.
- Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nat Rev Mol Cell Biol* **7**, 311-322.
- Hombach, J., Tsubata, T., Leclercq, L., Stappert, H., and Reth, M. (1990). Molecular components of the B-cell antigen receptor complex of the IgM class. *Nature* **343**, 760-762.
- Horsfield, J.A., Print, C.G., and Monnich, M. (2012). Diverse developmental disorders from the one ring: distinct molecular pathways underlie the cohesinopathies. *Front Genet* **3**, 171.
- Huen, M.S., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M.B., and Chen, J. (2007). RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell* **131**, 901-914.
- Imai, K., Catalan, N., Plebani, A., Marodi, L., Sanal, O., Kumaki, S., Nagendran, V., Wood, P., Glastre, C., Sarrot-Reynauld, F., *et al.* (2003a). Hyper-IgM syndrome type 4 with a B lymphocyte-intrinsic selective deficiency in Ig class-switch recombination. *J Clin Invest* **112**, 136-142.
- Imai, K., Slupphaug, G., Lee, W.I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H.E., *et al.* (2003b). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nat Immunol* **4**, 1023-1028.
- Imai, K., Zhu, Y., Revy, P., Morio, T., Mizutani, S., Fischer, A., Nonoyama, S., and Durandy, A. (2005). Analysis of class switch recombination and somatic hypermutation in patients affected with autosomal dominant hyper-IgM syndrome type 2. *Clin Immunol* **115**, 277-285.

- Ise, W., Kohyama, M., Schraml, B.U., Zhang, T., Schwer, B., Basu, U., Alt, F.W., Tang, J., Oltz, E.M., Murphy, T.L., *et al.* (2011). The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nat Immunol* *12*, 536-543.
- Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 1975-1980.
- Ivanovska, I., Jacques, P.E., Rando, O.J., Robert, F., and Winston, F. (2011). Control of chromatin structure by spt6: different consequences in coding and regulatory regions. *Molecular and cellular biology* *31*, 531-541.
- Jaehning, J.A. (2010). The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta* *1799*, 379-388.
- Jain, A., Ma, C.A., Liu, S., Brown, M., Cohen, J., and Strober, W. (2001). Specific missense mutations in NEMO result in hyper-IgM syndrome with hypohydrotic ectodermal dysplasia. *Nat Immunol* *2*, 223-228.
- Jain, A., Ma, C.A., Lopez-Granados, E., Means, G., Brady, W., Orange, J.S., Liu, S., Holland, S., and Derry, J.M. (2004). Specific NEMO mutations impair CD40-mediated c-Rel activation and B cell terminal differentiation. *J Clin Invest* *114*, 1593-1602.
- 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. *The Journal of experimental medicine* *203*, 319-323.
- Jeevan-Raj, B.P., Robert, I., Heyer, V., Page, A., Wang, J.H., Cammas, F., Alt, F.W., Losson, R., and Reina-San-Martin, B. (2011). Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *The Journal of experimental medicine* *208*, 1649-1660.
- Jensen, M.M., Christensen, M.S., Bonven, B., and Jensen, T.H. (2008). Requirements for chromatin reassembly during transcriptional downregulation of a heat shock gene in *Saccharomyces cerevisiae*. *FEBS J* *275*, 2956-2964.
- Johnson, K., Angelin-Duclos, C., Park, S., and Calame, K.L. (2003). Changes in histone acetylation are associated with differences in accessibility of V(H) gene segments to V-DJ recombination during B-cell ontogeny and development. *Molecular and cellular biology* *23*, 2438-2450.
- 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. *Journal of immunology* *176*, 4221-4234.
- Ju, L., Wing, J., Taylor, E., Brandt, R., Slijepcevic, P., Horsch, M., Rathkolb, B., Racz, I., Becker, L., Hans, W., *et al.* (2013). SMC6 is an essential gene in mice, but a hypomorphic mutant in the ATPase domain has a mild phenotype with a range of subtle abnormalities. *DNA Repair (Amst)* *12*, 356-366.
- Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F.W. (2006). Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* *24*, 541-570.
- Jung, S., Rajewsky, K., and Radbruch, A. (1993). Shutdown of class switch recombination by deletion of a switch region control element. *Science* *259*, 984-987.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* *467*, 430-435.
- Kaplan, C.D., Holland, M.J., and Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3'-end formation at the *Saccharomyces cerevisiae* GAL10-GAL7 locus. *The Journal of biological chemistry* *280*, 913-922.
- Kaplan, C.D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science* *301*, 1096-1099.
- Kaplan, C.D., Morris, J.R., Wu, C., and Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in *D. melanogaster*. *Genes Dev* *14*, 2623-2634.
- Kasahara, Y., Kaneko, H., Fukao, T., Terada, T., Asano, T., Kasahara, K., and Kondo, N. (2003). Hyper-IgM syndrome with putative dominant negative mutation in activation-induced cytidine deaminase. *J Allergy Clin Immunol* *112*, 755-760.
- Keegan, B.R., Feldman, J.L., Lee, D.H., Koos, D.S., Ho, R.K., Stainier, D.Y., and Yelon, D. (2002). The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo. *Development* *129*, 1623-1632.

- Kegel, A., and Sjogren, C. (2010). The Smc5/6 complex: more than repair? *Cold Spring Harb Symp Quant Biol* 75, 179-187.
- Kenter, A.L., Feldman, S., Wuerffel, R., Achour, I., Wang, L., and Kumar, S. (2012). Three-dimensional architecture of the IgH locus facilitates class switch recombination. *Annals of the New York Academy of Sciences* 1267, 86-94.
- Kim, J., Guermah, M., McGinty, R.K., Lee, J.S., Tang, Z., Milne, T.A., Shilatifard, A., Muir, T.W., and Roeder, R.G. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* 137, 459-471.
- Kim, J., Guermah, M., and Roeder, R.G. (2010). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* 140, 491-503.
- Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *The EMBO journal* 23, 354-364.
- Kim, S.T., Xu, B., and Kastan, M.B. (2002). Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev* 16, 560-570.
- Kinoshita, K., Tashiro, J., Tomita, S., Lee, C.G., and Honjo, T. (1998). Target specificity of immunoglobulin class switch recombination is not determined by nucleotide sequences of S regions. *Immunity* 9, 849-858.
- Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., *et al.* (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147, 95-106.
- Kliszczak, M., Stephan, A.K., Flanagan, A.M., and Morrison, C.G. (2012). SUMO ligase activity of vertebrate Mms21/Nse2 is required for efficient DNA repair but not for Smc5/6 complex stability. *DNA Repair (Amst)* 11, 799-810.
- Kobayashi, J., Tauchi, H., Sakamoto, S., Nakamura, A., Morishima, K., Matsuura, S., Kobayashi, T., Tamai, K., Tanimoto, K., and Komatsu, K. (2002). NBS1 localizes to gamma-H2AX foci through interaction with the FHA/BRCT domain. *Curr Biol* 12, 1846-1851.
- Kobayashi, M., Aida, M., Nagaoka, H., Begum, N.A., Kitawaki, Y., Nakata, M., Stanlie, A., Doi, T., Kato, L., Okazaki, I.M., *et al.* (2009). AID-induced decrease in topoisomerase 1 induces DNA structural alteration and DNA cleavage for class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* 106, 22375-22380.
- Kobayashi, M., Sabouri, Z., Sabouri, S., Kitawaki, Y., Pommier, Y., Abe, T., Kiyonari, H., and Honjo, T. (2011). Decrease in topoisomerase I is responsible for activation-induced cytidine deaminase (AID)-dependent somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America* 108, 19305-19310.
- Kok, F.O., Oster, E., Mentzer, L., Hsieh, J.C., Henry, C.A., and Sirotkin, H.I. (2007). The role of the SPT6 chromatin remodeling factor in zebrafish embryogenesis. *Dev Biol* 307, 214-226.
- Kolas, N.K., Chapman, J.R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F.D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T.M., *et al.* (2007). Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science* 318, 1637-1640.
- Komissarova, N., and Kashlev, M. (1997). RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *The Journal of biological chemistry* 272, 15329-15338.
- Korthauer, U., Graf, D., Mages, H.W., Briere, F., Padayachee, M., Malcolm, S., Ugazio, A.G., Notarangelo, L.D., Levinsky, R.J., and Kroczek, R.A. (1993). Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM. *Nature* 361, 539-541.
- Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296, 158-162.
- Kotnis, A., Du, L., Liu, C., Popov, S.W., and Pan-Hammarstrom, Q. (2009). Non-homologous end joining in class switch recombination: the beginning of the end. *Philos Trans R Soc Lond B Biol Sci* 364, 653-665.
- Kracker, S., Bergmann, Y., Demuth, I., Frappart, P.O., Hildebrand, G., Christine, R., Wang, Z.Q., Sperling, K., Digweed, M., and Radbruch, A. (2005). Nibrin functions in Ig class-switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* 102, 1584-1589.
- Kracker, S., Gardes, P., Mazerolles, F., and Durandy, A. (2010a). Immunoglobulin class switch recombination deficiencies. *Clin Immunol* 135, 193-203.
- Kracker, S., Imai, K., Gardes, P., Ochs, H.D., Fischer, A., and Durandy, A.H. (2010b). Impaired induction of DNA lesions during immunoglobulin class-switch recombination in humans

- influences end-joining repair. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 22225-22230.
- Kratz, C.P., Niemeyer, C.M., Juttner, E., Kartal, M., Weninger, A., Schmitt-Graeff, A., Kontny, U., Lauten, M., Utzolino, S., Radecke, J., *et al.* (2008). Childhood T-cell non-Hodgkin's lymphoma, colorectal carcinoma and brain tumor in association with cafe-au-lait spots caused by a novel homozygous PMS2 mutation. *Leukemia* *22*, 1078-1080.
- Kroczyk, R.A., Graf, D., Brugnani, D., Giliani, S., Korthuer, U., Ugazio, A., Senger, G., Mages, H.W., Villa, A., and Notarangelo, L.D. (1994). Defective expression of CD40 ligand on T cells causes "X-linked immunodeficiency with hyper-IgM (HIGM1)". *Immunol Rev* *138*, 39-59.
- Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., *et al.* (2003a). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* *11*, 721-729.
- Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J.F. (2002). RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Molecular and cellular biology* *22*, 6979-6992.
- Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., *et al.* (2003b). Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Molecular and cellular biology* *23*, 4207-4218.
- Kuang, F.L., Luo, Z., and Scharff, M.D. (2009). H3 trimethyl K9 and H3 acetyl K9 chromatin modifications are associated with class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 5288-5293.
- Kueng, S., Hegemann, B., Peters, B.H., Lipp, J.J., Schleiffer, A., Mechtler, K., and Peters, J.M. (2006). Wapl controls the dynamic association of cohesin with chromatin. *Cell* *127*, 955-967.
- Kumar, R., Dimenna, L., Schrode, N., Liu, T.C., Franck, P., Munoz-Descalzo, S., Hadjantonakis, A.K., Zarrin, A.A., Chaudhuri, J., Elemento, O., *et al.* (2013). AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature*.
- Kuppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* *5*, 251-262.
- Kuppers, R., and Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* *20*, 5580-5594.
- Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y., and Tonegawa, S. (1989). Junctional sequences of T cell receptor gamma delta genes: implications for gamma delta T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell* *59*, 859-870.
- Lebecque, S.G., and Gearhart, P.J. (1990). Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *The Journal of experimental medicine* *172*, 1717-1727.
- LeBien, T.W., and Tedder, T.F. (2008). B lymphocytes: how they develop and function. *Blood* *112*, 1570-1580.
- Lee, J.H., and Paull, T.T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science* *308*, 551-554.
- Lee-Theilen, M., Matthews, A.J., Kelly, D., Zheng, S., and Chaudhuri, J. (2011). CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nat Struct Mol Biol* *18*, 75-79.
- Lepse, C.L., Kumar, R., and Ganea, D. (1994). Extrachromosomal eukaryotic DNA substrates for switch recombination: analysis of isotype and cell specificity. *DNA Cell Biol* *13*, 1151-1161.
- Leung, H., and Maizels, N. (1992). Transcriptional regulatory elements stimulate recombination in extrachromosomal substrates carrying immunoglobulin switch-region sequences. *Proceedings of the National Academy of Sciences of the United States of America* *89*, 4154-4158.
- Li, G., Alt, F.W., Cheng, H.L., Brush, J.W., Goff, P.H., Murphy, M.M., Franco, S., Zhang, Y., and Zha, S. (2008). Lymphocyte-specific compensation for XLF/cernunnos end-joining functions in V(D)J recombination. *Mol Cell* *31*, 631-640.
- Li, G., Zan, H., Xu, Z., and Casali, P. (2013). Epigenetics of the antibody response. *Trends Immunol.*
- Li, L., Halaby, M.J., Hakem, A., Cardoso, R., El Ghamrasni, S., Harding, S., Chan, N., Bristow, R., Sanchez, O., Durocher, D., *et al.* (2010). Rnf8 deficiency impairs class switch recombination, spermatogenesis, and genomic integrity and predisposes for cancer. *The Journal of experimental medicine* *207*, 983-997.

- 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.
- Li, Z., Scherer, S.J., Ronai, D., Iglesias-Ussel, M.D., Peled, J.U., Bardwell, P.D., Zhuang, M., Lee, K., Martin, A., Edelmann, W., *et al.* (2004). Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. *The Journal of experimental medicine* 200, 47-59.
- Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. *The Journal of biological chemistry* 283, 1-5.
- 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.
- Lieber, M.R., Ma, Y., Pannicke, U., and Schwarz, K. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol* 4, 712-720.
- Lindroos, H.B., Strom, L., Itoh, T., Katou, Y., Shirahige, K., and Sjogren, C. (2006). Chromosomal association of the Smc5/6 complex reveals that it functions in differently regulated pathways. *Mol Cell* 22, 755-767.
- Lis, J.T. (2007). Imaging Drosophila gene activation and polymerase pausing in vivo. *Nature* 450, 198-202.
- Liu, J., Zhang, J., Gong, Q., Xiong, P., Huang, H., Wu, B., Lu, G., Wu, J., and Shi, Y. (2011). Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain. *The Journal of biological chemistry* 286, 29218-29226.
- Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451, 841-845.
- Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., and Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* 27, 561-571.
- Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W.H., Ranish, J., Shokat, K.M., and Hahn, S. (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Molecular and cellular biology* 29, 4852-4863.
- Lorenz, M., Jung, S., and Radbruch, A. (1995). Switch transcripts in immunoglobulin class switching. *Science* 267, 1825-1828.
- Losada, A., Hirano, M., and Hirano, T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes Dev* 12, 1986-1997.
- Losada, A., and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev* 19, 1269-1287.
- Lou, Z., Minter-Dykhouse, K., Franco, S., Gostissa, M., Rivera, M.A., Celeste, A., Manis, J.P., van Deursen, J., Nussenzweig, A., Paull, T.T., *et al.* (2006). MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol Cell* 21, 187-200.
- Lougaris, V., Badolato, R., Ferrari, S., and Plebani, A. (2005). Hyper immunoglobulin M syndrome due to CD40 deficiency: clinical, molecular, and immunological features. *Immunol Rev* 203, 48-66.
- 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.
- MacDuff, D.A., Neuberger, M.S., and Harris, R.S. (2006). MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. *Mol Immunol* 43, 1099-1108.
- Maeda, K., Singh, S.K., Eda, K., Kitabatake, M., Pham, P., Goodman, M.F., and Sakaguchi, N. (2010). GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *The Journal of biological chemistry* 285, 23945-23953.
- Maes, J., Chappaz, S., Cavelier, P., O'Neill, L., Turner, B., Rougeon, F., and Goodhardt, M. (2006). Activation of V(D)J recombination at the IgH chain JH locus occurs within a 6-kilobase chromatin domain and is associated with nucleosomal remodeling. *Journal of immunology* 176, 5409-5417.
- Mai, T., Zan, H., Zhang, J., Hawkins, J.S., Xu, Z., and Casali, P. (2010). Estrogen receptors bind to and activate the HOXC4/HoxC4 promoter to potentiate HoxC4-mediated activation-induced cytosine deaminase induction, immunoglobulin class switch DNA recombination, and somatic hypermutation. *The Journal of biological chemistry* 285, 37797-37810.



- Mailand, N., Bekker-Jensen, S., Faustrup, H., Melander, F., Bartek, J., Lukas, C., and Lukas, J. (2007). RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell* *131*, 887-900.
- Maizels, N. (2005). Immunoglobulin Gene Diversification. *Annu Rev Genet*.
- Malagon, F., and Aguilera, A. (2001). Yeast spt6-140 mutation, affecting chromatin and transcription, preferentially increases recombination in which Rad51p-mediated strand exchange is dispensable. *Genetics* *158*, 597-611.
- Manis, J.P., Dudley, D., Kaylor, L., and Alt, F.W. (2002). IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. *Immunity* *16*, 607-617.
- Manis, J.P., Gu, Y., Lansford, R., Sonoda, E., Ferrini, R., Davidson, L., Rajewsky, K., and Alt, F.W. (1998a). Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. *The Journal of experimental medicine* *187*, 2081-2089.
- 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.
- Manis, J.P., van der Stoep, N., Tian, M., Ferrini, R., Davidson, L., Bottaro, A., and Alt, F.W. (1998b). Class switching in B cells lacking 3' immunoglobulin heavy chain enhancers. *The Journal of experimental medicine* *188*, 1421-1431.
- Marculescu, R., Vanura, K., Montpellier, B., Roulland, S., Le, T., Navarro, J.M., Jager, U., McBlane, F., and Nadel, B. (2006). Recombinase, chromosomal translocations and lymphoid neoplasia: targeting mistakes and repair failures. *DNA Repair (Amst)* *5*, 1246-1258.
- Martin, A., Li, Z., Lin, D.P., Bardwell, P.D., Iglesias-Ussel, M.D., Edelman, W., and Scharff, M.D. (2003). Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. *The Journal of experimental medicine* *198*, 1171-1178.
- Martomo, S.A., Yang, W.W., and Gearhart, P.J. (2004). A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. *The Journal of experimental medicine* *200*, 61-68.
- Matthews, A.G., Kuo, A.J., Ramon-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.
- Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* *17*, 1272-1278.
- McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M., and Oettinger, M.A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* *83*, 387-395.
- McBride, K.M., Barreto, V., Ramiro, A.R., Stavropoulos, P., and Nussenzweig, M.C. (2004). Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *The Journal of experimental medicine* *199*, 1235-1244.
- McBride, K.M., Gazumyan, A., Woo, E.M., Barreto, V.M., Robbiani, D.F., Chait, B.T., and Nussenzweig, M.C. (2006). Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* *103*, 8798-8803.
- McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. *The Journal of experimental medicine* *205*, 2585-2594.
- McCormack, W.T., Tjoelker, L.W., and Thompson, C.B. (1991). Avian B-cell development: generation of an immunoglobulin repertoire by gene conversion. *Annu Rev Immunol* *9*, 219-241.
- Mehta, A., Kinter, M.T., Sherman, N.E., and Driscoll, D.M. (2000). Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Molecular and cellular biology* *20*, 1846-1854.
- Melby, T.E., Ciampaglio, C.N., Briscoe, G., and Erickson, H.P. (1998). The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. *J Cell Biol* *142*, 1595-1604.
- Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* *91*, 35-45.
- Migliazza, A., Martinotti, S., Chen, W., Fusco, C., Ye, B.H., Knowles, D.M., Offit, K., Chaganti, R.S., and Dalla-Favera, R. (1995). Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* *92*, 12520-12524.

- Minegishi, Y., Lavoie, A., Cunningham-Rundles, C., Bedard, P.M., Hebert, J., Cote, L., Dan, K., Sedlak, D., Buckley, R.H., Fischer, A., *et al.* (2000). Mutations in activation-induced cytidine deaminase in patients with hyper IgM syndrome. *Clin Immunol* 97, 203-210.
- Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *The Journal of biological chemistry* 279, 52353-52360.
- Morshead, K.B., Ciccone, D.N., Taverna, S.D., Allis, C.D., and Oettinger, M.A. (2003). Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. *Proceedings of the National Academy of Sciences of the United States of America* 100, 11577-11582.
- Mueller, C.L., Porter, S.E., Hoffman, M.G., and Jaehning, J.A. (2004). The Paf1 complex has functions independent of actively transcribing RNA polymerase II. *Mol Cell* 14, 447-456.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553-563.
- Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N.A., and Honjo, T. (2007). Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Advances in immunology* 94, 1-36.
- Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of biological chemistry* 274, 18470-18476.
- Nakamura, M., Kondo, S., Sugai, M., Nazarea, M., Imamura, S., and Honjo, T. (1996). High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int Immunol* 8, 193-201.
- Nakanishi, Y., Kondo, S., Wakisaka, N., Tsuji, A., Endo, K., Murono, S., Ito, M., Kitamura, K., Muramatsu, M., and Yoshizaki, T. (2013). Role of activation-induced cytidine deaminase in the development of oral squamous cell carcinoma. *PLoS one* 8, e62066.
- Nambu, Y., Sugai, M., Gonda, H., Lee, C.G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* 302, 2137-2140.
- Navaratnam, N., Morrison, J.R., Bhattacharya, S., Patel, D., Funahashi, T., Giannoni, F., Teng, B.B., Davidson, N.O., and Scott, J. (1993). The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *The Journal of biological chemistry* 268, 20709-20712.
- Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochim Biophys Acta* 1809, 34-45.
- Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* 11, 709-719.
- Nijnik, A., Dawson, S., Crockford, T.L., Woodbine, L., Visetnoi, S., Bennett, S., Jones, M., Turner, G.D., Jeggo, P.A., Goodnow, C.C., *et al.* (2009). Impaired lymphocyte development and antibody class switching and increased malignancy in a murine model of DNA ligase IV syndrome. *J Clin Invest* 119, 1696-1705.
- Nilsen, H., Stamp, G., Andersen, S., Hrivnak, G., Krokan, H.E., Lindahl, T., and Barnes, D.E. (2003). Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene* 22, 5381-5386.
- Nishiwaki, K., Sano, T., and Miwa, J. (1993). *emb-5*, a gene required for the correct timing of gut precursor cell division during gastrulation in *Caenorhabditis elegans*, encodes a protein similar to the yeast nuclear protein SPT6. *Mol Gen Genet* 239, 313-322.
- Nonoyama, S., Hollenbaugh, D., Aruffo, A., Ledbetter, J.A., and Ochs, H.D. (1993). B cell activation via CD40 is required for specific antibody production by antigen-stimulated human B cells. *The Journal of experimental medicine* 178, 1097-1102.
- Nordick, K., Hoffman, M.G., Betz, J.L., and Jaehning, J.A. (2008). Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. *Eukaryot Cell* 7, 1158-1167.
- Notarangelo, L.D., Duse, M., and Ugazio, A.G. (1992). Immunodeficiency with hyper-IgM (HIM). *Immunodeficiency Rev* 3, 101-121.

- Nowak, U., Matthews, A.J., Zheng, S., and Chaudhuri, J. (2011). The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nat Immunol* 12, 160-166.
- 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.
- Okazaki, I.M., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. *The Journal of experimental medicine* 197, 1173-1181.
- Okazaki, I.M., Kotani, A., and Honjo, T. (2007). Role of AID in tumorigenesis. *Advances in immunology* 94, 245-273.
- Okazaki, I.M., Okawa, K., Kobayashi, M., Yoshikawa, K., Kawamoto, S., Nagaoka, H., Shinkura, R., Kitawaki, Y., Taniguchi, H., Natsume, T., *et al.* (2011). Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America* 108, 7920-7925.
- Orthwein, A., and Di Noia, J.M. (2012). Activation induced deaminase: how much and where? *Semin Immunol* 24, 246-254.
- Orthwein, A., Patenaude, A.M., Affar el, B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *The Journal of experimental medicine* 207, 2751-2765.
- Orthwein, A., Zahn, A., Methot, S.P., Godin, D., Conticello, S.G., Terada, K., and Di Noia, J.M. (2012). Optimal functional levels of activation-induced deaminase specifically require the Hsp40 DnaJa1. *The EMBO journal* 31, 679-691.
- Ott, D.E., and Marcu, K.B. (1989). Molecular requirements for immunoglobulin heavy chain constant region gene switch-recombination revealed with switch-substrate retroviruses. *Int Immunol* 1, 582-591.
- Outwin, E.A., Irmisch, A., Murray, J.M., and O'Connell, M.J. (2009). Smc5-Smc6-dependent removal of cohesin from mitotic chromosomes. *Molecular and cellular biology* 29, 4363-4375.
- Palecek, J., Vidot, S., Feng, M., Doherty, A.J., and Lehmann, A.R. (2006). The Smc5-Smc6 DNA repair complex. bridging of the Smc5-Smc6 heads by the KLEISIN, Nse4, and non-Kleisin subunits. *The Journal of biological chemistry* 281, 36952-36959.
- Pan-Hammarstrom, Q., Jones, A.M., Lahdesmaki, A., Zhou, W., Gatti, R.A., Hammarstrom, L., Gennery, A.R., and Ehrenstein, M.R. (2005). Impact of DNA ligase IV on nonhomologous end joining pathways during class switch recombination in human cells. *The Journal of experimental medicine* 201, 189-194.
- Panigrahi, A.K., and Pati, D. (2012). Higher-order orchestration of hematopoiesis: is cohesin a new player? *Exp Hematol* 40, 967-973.
- Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., *et al.* (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422-433.
- Park, S.R., Zan, H., Pal, Z., Zhang, J., Al-Qahtani, A., Pone, E.J., Xu, Z., Mai, T., and Casali, P. (2009). HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation. *Nat Immunol* 10, 540-550.
- Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006). PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proceedings of the National Academy of Sciences of the United States of America* 103, 395-400.
- Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Kuppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412, 341-346.
- Patenaude, A.M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzi, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* 16, 517-527.
- Pauklin, S., and Petersen-Mahrt, S.K. (2009). Progesterone inhibits activation-induced deaminase by binding to the promoter. *Journal of immunology* 183, 1238-1244.
- Pauklin, S., Hernandez, I.V., Bachmann, G., Ramiro, A.R., and Petersen-Mahrt, S.K. (2009). Estrogen directly activates AID transcription and function. *The Journal of experimental medicine* 206, 99-111.
- Paull, T.T., Rogakou, E.P., Yamazaki, V., Kirchgessner, C.U., Gellert, M., and Bonner, W.M. (2000). A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr Biol* 10, 886-895.

- Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina-San-Martin, B., Barreto, V., *et al.* (2010). Activation-Induced Cytidine Deaminase Targets DNA at Sites of RNA Polymerase II Stalling by Interaction with Spt5. *Cell* **143**, 122-133.
- Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125**, 703-717.
- Pei, H., Wu, X., Liu, T., Yu, K., Jelinek, D.F., and Lou, Z. (2013). The histone methyltransferase MMSET regulates class switch recombination. *Journal of immunology* **190**, 756-763.
- Peled, J.U., Kuang, F.L., Iglesias-Ussel, M.D., Roa, S., Kalis, S.L., Goodman, M.F., and Scharff, M.D. (2008). The biochemistry of somatic hypermutation. *Annu Rev Immunol* **26**, 481-511.
- Penheiter, K.L., Washburn, T.M., Porter, S.E., Hoffman, M.G., and Jaehning, J.A. (2005). A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets. *Mol Cell* **20**, 213-223.
- Perlot, T., Li, G., and Alt, F.W. (2008). Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 3843-3848.
- Peron, S., Metin, A., Gardes, P., Alyanakian, M.A., Sheridan, E., Kratz, C.P., Fischer, A., and Durandy, A. (2008). Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination. *The Journal of experimental medicine* **205**, 2465-2472.
- Peron, S., Pan-Hammarstrom, Q., Imai, K., Du, L., Taubenheim, N., Sanal, O., Marodi, L., Bergelin-Besancon, A., Benkerrou, M., de Villartay, J.P., *et al.* (2007). A primary immunodeficiency characterized by defective immunoglobulin class switch recombination and impaired DNA repair. *The Journal of experimental medicine* **204**, 1207-1216.
- Peters, A., and Storb, U. (1996). Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* **4**, 57-65.
- Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R., *et al.* (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* **414**, 660-665.
- Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99-103.
- Petry, K., Siebenkotten, G., Christine, R., Hein, K., and Radbruch, A. (1999). An extrachromosomal switch recombination substrate reveals kinetics and substrate requirements of switch recombination in primary murine B cells. *Int Immunol* **11**, 753-763.
- Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* **424**, 103-107.
- Pham, P., Smolka, M.B., Calabrese, P., Landolph, A., Zhang, K., Zhou, H., and Goodman, M.F. (2008). Impact of phosphorylation and phosphorylation-null mutants on the activity and deamination specificity of activation-induced cytidine deaminase. *The Journal of biological chemistry*.
- Phung, Q.H., Winter, D.B., Cranston, A., Tarone, R.E., Bohr, V.A., Fishel, R., and Gearhart, P.J. (1998). Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. *The Journal of experimental medicine* **187**, 1745-1751.
- Pinaud, E., Khamlichi, A.A., Le Morvan, C., Drouet, M., Nalesso, V., Le Bert, M., and Cogne, M. (2001). Localization of the 3' IgH locus elements that effect long-distance regulation of class switch recombination. *Immunity* **15**, 187-199.
- Pokholok, D.K., Hannett, N.M., and Young, R.A. (2002). Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Mol Cell* **9**, 799-809.
- Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101-1105.
- Potter, M., and Wiener, F. (1992). Plasmacytomagenesis in mice: model of neoplastic development dependent upon chromosomal translocations. *Carcinogenesis* **13**, 1681-1697.
- Potts, P.R. (2009). The Yin and Yang of the MMS21-SMC5/6 SUMO ligase complex in homologous recombination. *DNA Repair (Amst)* **8**, 499-506.

- Potts, P.R., Porteus, M.H., and Yu, H. (2006). Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *The EMBO journal* **25**, 3377-3388.
- Potts, P.R., and Yu, H. (2005). Human MMS21/NSE2 is a SUMO ligase required for DNA repair. *Molecular and cellular biology* **25**, 7021-7032.
- Potts, P.R., and Yu, H. (2007). The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nat Struct Mol Biol* **14**, 581-590.
- Qiu, H., Hu, C., Wong, C.M., and Hinnebusch, A.G. (2006). The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II. *Molecular and cellular biology* **26**, 3135-3148.
- Quartier, P., Bustamante, J., Sanal, O., Plebani, A., Debre, M., Deville, A., Litzman, J., Levy, J., Fermand, J.P., Lane, P., *et al.* (2004). Clinical, immunologic and genetic analysis of 29 patients with autosomal recessive hyper-IgM syndrome due to Activation-Induced Cytidine Deaminase deficiency. *Clin Immunol* **110**, 22-29.
- Quong, M.W., Harris, D.P., Swain, S.L., and Murre, C. (1999). E2A activity is induced during B-cell activation to promote immunoglobulin class switch recombination. *The EMBO journal* **18**, 6307-6318.
- 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.
- Rada, C., Ehrenstein, M.R., Neuberger, M.S., and Milstein, C. (1998). Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity* **9**, 135-141.
- Rada, C., Jarvis, J.M., and Milstein, C. (2002a). AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7003-7008.
- Rada, C., and Milstein, C. (2001). The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. *The EMBO journal* **20**, 4570-4576.
- Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002b). Immunoglobulin Isotype Switching Is Inhibited and Somatic Hypermutation Perturbed in UNG-Deficient Mice. *Curr Biol* **12**, 1748-1755.
- Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. *Cell* **141**, 432-445.
- Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201-1212.
- Rajagopal, D., Maul, R.W., Ghosh, A., Chakraborty, T., Khamlichi, A.A., Sen, R., and Gearhart, P.J. (2009). Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *The Journal of experimental medicine* **206**, 1237-1244.
- Ramachandran, S., Chahwan, R., Nepal, R.M., Frieder, D., Panier, S., Roa, S., Zaheen, A., Durocher, D., Scharff, M.D., and Martin, A. (2010). The RNF8/RNF168 ubiquitin ligase cascade facilitates class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 809-814.
- Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., *et al.* (2006). Role of genomic instability and p53 in AID-induced c-myc-IgH translocations. *Nature* **440**, 105-109.
- Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., and Nussenzweig, M.C. (2004). AID Is Required for c-myc/IgH Chromosome Translocations In Vivo. *Cell* **118**, 431-438.
- Ranjit, S., Khair, L., Linehan, E.K., Ucher, A.J., Chakrabarti, M., Schrader, C.E., and Stavnezer, J. (2011). AID binds cooperatively with UNG and Msh2-Msh6 to Ig switch regions dependent upon the AID C terminus. *Journal of immunology* **187**, 2464-2475.
- Rappold, I., Iwabuchi, K., Date, T., and Chen, J. (2001). Tumor suppressor p53 binding protein 1 (53BP1) is involved in DNA damage-signaling pathways. *J Cell Biol* **153**, 613-620.
- Reina-San-Martin, B., Chen, H.T., Nussenzweig, A., and Nussenzweig, M.C. (2004). ATM is required for efficient recombination between immunoglobulin switch regions. *The Journal of experimental medicine* **200**, 1103-1110.
- Reina-San-Martin, B., Chen, J., Nussenzweig, A., and Nussenzweig, M.C. (2007). Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1-/- B cells. *Eur J Immunol* **37**, 235-239.

- Reina-San-Martin, B., Difilippantonio, S., Hanitsch, L., Masilamani, R.F., Nussenzweig, A., and Nussenzweig, M.C. (2003). H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. *The Journal of experimental medicine* *197*, 1767-1778.
- Reina-San-Martin, B., Nussenzweig, M.C., Nussenzweig, A., and Difilippantonio, S. (2005). Genomic instability, endoreduplication, and diminished Ig class-switch recombination in B cells lacking Nbs1. *Proceedings of the National Academy of Sciences of the United States of America* *102*, 1590-1595.
- Remeseiro, S., and Losada, A. (2013). Cohesin, a chromatin engagement ring. *Curr Opin Cell Biol* *25*, 63-71.
- Retter, I., Chevillard, C., Scharfe, M., Conrad, A., Hafner, M., Im, T.H., Ludewig, M., Nordsiek, G., Severitt, S., Thies, S., *et al.* (2007). Sequence and characterization of the Ig heavy chain constant and partial variable region of the mouse strain 129S1. *Journal of immunology* *179*, 2419-2427.
- Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labeolouse, R., Gennery, A., *et al.* (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* *102*, 565-575.
- Reynaud, C.A., Anquez, V., Dahan, A., and Weill, J.C. (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell* *40*, 283-291.
- Reynaud, C.A., Anquez, V., Grimal, H., and Weill, J.C. (1987). A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* *48*, 379-388.
- Reynaud, C.A., Anquez, V., and Weill, J.C. (1991). The chicken D locus and its contribution to the immunoglobulin heavy chain repertoire. *Eur J Immunol* *21*, 2661-2670.
- Reynaud, C.A., Dahan, A., Anquez, V., and Weill, J.C. (1989). Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* *59*, 171-183.
- Rhodes, J.M., Bentley, F.K., Print, C.G., Dorsett, D., Misulovin, Z., Dickinson, E.J., Crosier, K.E., Crosier, P.S., and Horsfield, J.A. (2010). Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. *Dev Biol* *344*, 637-649.
- Rivera-Munoz, P., Soulas-Sprauel, P., Le Guyader, G., Abramowski, V., Bruneau, S., Fischer, A., Paques, F., and de Villartay, J.P. (2009). Reduced immunoglobulin class switch recombination in the absence of Artemis. *Blood* *114*, 3601-3609.
- Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., *et al.* (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol Cell* *36*, 631-641.
- Robert, I., Dantzer, F., and Reina-San-Martin, B. (2009). Parp1 facilitates alternative NHEJ, whereas Parp2 suppresses IgH/c-myc translocations during immunoglobulin class switch recombination. *The Journal of experimental medicine* *206*, 1047-1056.
- Rocha, P.P., Micsinai, M., Kim, J.R., Hewitt, S.L., Souza, P.P., Trimarchi, T., Strino, F., Parisi, F., Kluger, Y., and Skok, J.A. (2012). Close proximity to Igh is a contributing factor to AID-mediated translocations. *Mol Cell* *47*, 873-885.
- Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S., and Bonner, W.M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *The Journal of biological chemistry* *273*, 5858-5868.
- Rogozin, I.B., and Kolchanov, N.A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta* *1171*, 11-18.
- Rolef Ben-Shahar, T., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., and Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science* *321*, 563-566.
- Rollins, R.A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* *152*, 577-593.
- Ronai, D., Iglesias-Ussel, M.D., Fan, M., Li, Z., Martin, A., and Scharff, M.D. (2007). Detection of chromatin-associated single-stranded DNA in regions targeted for somatic hypermutation. *The Journal of experimental medicine* *204*, 181-190.
- Rooney, S., Alt, F.W., Sekiguchi, J., and Manis, J.P. (2005). Artemis-independent functions of DNA-dependent protein kinase in Ig heavy chain class switch recombination and development.

- Proceedings of the National Academy of Sciences of the United States of America *102*, 2471-2475.
- Rosenberg, B.R., and Papavasiliou, F.N. (2007). Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. *Advances in immunology* *94*, 215-244.
- Roth, D.B., Zhu, C., and Gellert, M. (1993). Characterization of broken DNA molecules associated with V(D)J recombination. *Proceedings of the National Academy of Sciences of the United States of America* *90*, 10788-10792.
- Roulland, S., Faroudi, M., Mamessier, E., Sungalee, S., Salles, G., and Nadel, B. (2011). Early steps of follicular lymphoma pathogenesis. *Advances in immunology* *111*, 1-46.
- Roulland, S., Navarro, J.M., Grenot, P., Milili, M., Agopian, J., Montpellier, B., Gauduchon, P., Lebailly, P., Schiff, C., and Nadel, B. (2006). Follicular lymphoma-like B cells in healthy individuals: a novel intermediate step in early lymphomagenesis. *The Journal of experimental medicine* *203*, 2425-2431.
- Rozenblatt-Rosen, O., Hughes, C.M., Nannepaga, S.J., Shanmugam, K.S., Copeland, T.D., Guszczynski, T., Resau, J.H., and Meyerson, M. (2005). The parafibromin tumor suppressor protein is part of a human Paf1 complex. *Molecular and cellular biology* *25*, 612-620.
- Rozenblatt-Rosen, O., Nagaike, T., Francis, J.M., Kaneko, S., Glatt, K.A., Hughes, C.M., LaFramboise, T., Manley, J.L., and Meyerson, M. (2009). The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proceedings of the National Academy of Sciences of the United States of America* *106*, 755-760.
- Rulten, S.L., Fisher, A.E., Robert, I., Zuma, M.C., Rouleau, M., Ju, L., Poirier, G., Reina-San-Martin, B., and Caldecott, K.W. (2011). PARP-3 and APLF function together to accelerate nonhomologous end-joining. *Mol Cell* *41*, 33-45.
- Sakai, E., Bottaro, A., and Alt, F.W. (1999). The Ig heavy chain intronic enhancer core region is necessary and sufficient to promote efficient class switch recombination. *Int Immunol* *11*, 1709-1713.
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature* *286*, 676-683.
- Santos, M.A., Huen, M.S., Jankovic, M., Chen, H.T., Lopez-Contreras, A.J., Klein, I.A., Wong, N., Barbancho, J.L., Fernandez-Capetillo, O., Nussenzweig, M.C., *et al.* (2010). Class switching and meiotic defects in mice lacking the E3 ubiquitin ligase RNF8. *The Journal of experimental medicine* *207*, 973-981.
- Saribasak, H., and Gearhart, P.J. (2012). Does DNA repair occur during somatic hypermutation? *Semin Immunol* *24*, 287-292.
- Saribasak, H., Maul, R.W., Cao, Z., McClure, R.L., Yang, W., McNeill, D.R., Wilson, D.M., 3rd, and Gearhart, P.J. (2011). XRCC1 suppresses somatic hypermutation and promotes alternative nonhomologous end joining in Igh genes. *The Journal of experimental medicine* *208*, 2209-2216.
- Sayegh, C.E., Quong, M.W., Agata, Y., and Murre, C. (2003). E-proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat Immunol* *4*, 586-593.
- Schar, P., Fasi, M., and Jessberger, R. (2004). SMC1 coordinates DNA double-strand break repair pathways. *Nucleic acids research* *32*, 3921-3929.
- Schatz, D.G. (2004). Antigen receptor genes and the evolution of a recombinase. *Semin Immunol* *16*, 245-256.
- Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. *Cell* *59*, 1035-1048.
- Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. *Annu Rev Genet* *45*, 167-202.
- Scheeren, F.A., Nagasawa, M., Weijer, K., Cupedo, T., Kirberg, J., Legrand, N., and Spits, H. (2008). T cell-independent development and induction of somatic hypermutation in human IgM+ IgD+ CD27+ B cells. *The Journal of experimental medicine* *205*, 2033-2042.
- Schlissel, M., Constantinescu, A., Morrow, T., Baxter, M., and Peng, A. (1993). Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. *Genes Dev* *7*, 2520-2532.
- Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P., and Odom, D.T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res* *20*, 578-588.

- Schrader, C.E., Edelmann, W., Kucherlapati, R., and Stavnezer, J. (1999). Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. *The Journal of experimental medicine* *190*, 323-330.
- Schrader, C.E., Guikema, J.E., Linehan, E.K., Selsing, E., and Stavnezer, J. (2007). Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. *Journal of immunology* *179*, 6064-6071.
- Schrader, C.E., Linehan, E.K., Mochegova, S.N., Woodland, R.T., and Stavnezer, J. (2005). Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *The Journal of experimental medicine* *202*, 561-568.
- Seidl, K.J., Bottaro, A., Vo, A., Zhang, J., Davidson, L., and Alt, F.W. (1998). An expressed neo(r) cassette provides required functions of the 1gamma2b exon for class switching. *Int Immunol* *10*, 1683-1692.
- Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnoli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K., *et al.* (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* *476*, 467-471.
- Sergeant, J., Taylor, E., Palecek, J., Fousteri, M., Andrews, E.A., Sweeney, S., Shinagawa, H., Watts, F.Z., and Lehmann, A.R. (2005). Composition and architecture of the Schizosaccharomyces pombe Rad18 (Smc5-6) complex. *Molecular and cellular biology* *25*, 172-184.
- Sharpe, M.J., Milstein, C., Jarvis, J.M., and Neuberger, M.S. (1991). Somatic hypermutation of immunoglobulin kappa may depend on sequences 3' of C kappa and occurs on passenger transgenes. *The EMBO journal* *10*, 2139-2145.
- Shen, H.M. (2007). Activation-induced cytidine deaminase acts on double-strand breaks in vitro. *Mol Immunol* *44*, 974-983.
- Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* *280*, 1750-1752.
- Shen, X., Xi, G., Radhakrishnan, Y., and Clemmons, D.R. (2009). Identification of novel SHPS-1-associated proteins and their roles in regulation of insulin-like growth factor-dependent responses in vascular smooth muscle cells. *Mol Cell Proteomics* *8*, 1539-1551.
- Shi, X., Chang, M., Wolf, A.J., Chang, C.H., Frazer-Abel, A.A., Wade, P.A., Burton, Z.F., and Jaehning, J.A. (1997). Cdc73p and Paf1p are found in a novel RNA polymerase II-containing complex distinct from the Srbp-containing holoenzyme. *Molecular and cellular biology* *17*, 1160-1169.
- Shiloh, Y. (2003). ATM and related protein kinases: safeguarding genome integrity. *Nat Rev Cancer* *3*, 155-168.
- Shinkura, R., Ito, S., Begum, N.A., Nagaoka, H., Muramatsu, M., Kinoshita, K., Sakakibara, Y., Hijikata, H., and Honjo, T. (2004). Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol* *5*, 707-712.
- Singh, S.K., Maeda, K., Eid, M.M., Almofty, S.A., Ono, M., Pham, P., Goodman, M.F., and Sakaguchi, N. (2013). GANP regulates recruitment of AID to immunoglobulin variable regions by modulating transcription and nucleosome occupancy. *Nat Commun* *4*, 1830.
- Soulas-Sprauel, P., Le Guyader, G., Rivera-Munoz, P., Abramowski, V., Olivier-Martin, C., Goujet-Zalc, C., Charneau, P., and de Villartay, J.P. (2007). Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. *The Journal of experimental medicine* *204*, 1717-1727.
- Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., and Hartzog, G.A. (2002). The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *The EMBO journal* *21*, 1764-1774.
- Stanlie, A., Aida, M., Muramatsu, M., Honjo, T., and Begum, N.A. (2010). Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 22190-22195.
- Stanlie, A., Begum, N.A., Akiyama, H., and Honjo, T. (2012). The DSIF subunits Spt4 and Spt5 have distinct roles at various phases of immunoglobulin class switch recombination. *PLoS Genet* *8*, e1002675.
- Stavnezer, J., Bjorkman, A., Du, L., Cagigi, A., and Pan-Hammarstrom, Q. (2010). Mapping of switch recombination junctions, a tool for studying DNA repair pathways during immunoglobulin class switching. *Advances in immunology* *108*, 45-109.
- Stavnezer, J., Bradley, S.P., Rousseau, N., Pearson, T., Shanmugam, A., Waite, D.J., Rogers, P.R., and Kenter, A.L. (1999). Switch recombination in a transfected plasmid occurs preferentially in



- a B cell line that undergoes switch recombination of its chromosomal Ig heavy chain genes. *Journal of immunology* 163, 2028-2040.
- Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008a). Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 26, 261-292.
- Stavnezer, J., Guikema, J.E., and Schrader, C.E. (2008b). Mechanism and Regulation of Class Switch Recombination. *Annu Rev Immunol* 26, 261-292.
- Stavnezer, J., and Schrader, C.E. (2006). Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. *Trends Genet* 22, 23-28.
- Stavnezer-Nordgren, J., and Sirlin, S. (1986). Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *The EMBO journal* 5, 95-102.
- Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proceedings of the National Academy of Sciences of the United States of America* 79, 3418-3422.
- Stephan, A.K., Kliszczak, M., Dodson, H., Cooley, C., and Morrison, C.G. (2011a). Roles of vertebrate Smc5 in sister chromatid cohesion and homologous recombinational repair. *Molecular and cellular biology* 31, 1369-1381.
- Stephan, A.K., Kliszczak, M., and Morrison, C.G. (2011b). The Nse2/Mms21 SUMO ligase of the Smc5/6 complex in the maintenance of genome stability. *FEBS Lett* 585, 2907-2913.
- Stewart, G.S., Panier, S., Townsend, K., Al-Hakim, A.K., Kolas, N.K., Miller, E.S., Nakada, S., Ylanko, J., Olivarius, S., Mendez, M., *et al.* (2009). The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. *Cell* 136, 420-434.
- Stewart, G.S., Wang, B., Bignell, C.R., Taylor, A.M., and Elledge, S.J. (2003). MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature* 421, 961-966.
- Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., and Jackson, S.P. (2005). MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* 123, 1213-1226.
- Suematsu, S., Matsusaka, T., Matsuda, T., Ohno, S., Miyazaki, J., Yamamura, K., Hirano, T., and Kishimoto, T. (1992). Generation of plasmacytomas with the chromosomal translocation t(12;15) in interleukin 6 transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* 89, 232-235.
- Sun, M., Lariviere, L., Dengl, S., Mayer, A., and Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *The Journal of biological chemistry* 285, 41597-41603.
- Swanson, M.S., Malone, E.A., and Winston, F. (1991). SPT5, an essential gene important for normal transcription in *Saccharomyces cerevisiae*, encodes an acidic nuclear protein with a carboxy-terminal repeat. *Molecular and cellular biology* 11, 4286.
- Swanson, M.S., and Winston, F. (1992). SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in *Saccharomyces cerevisiae*. *Genetics* 132, 325-336.
- Swanson, P.C., and Desiderio, S. (1998). V(D)J recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. *Immunity* 9, 115-125.
- Swanson, P.C., and Desiderio, S. (1999). RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of a V(D)J initiation complex. *Molecular and cellular biology* 19, 3674-3683.
- Ta, V.T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., *et al.* (2003). AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat Immunol* 4, 843-848.
- Taccioli, G.E., Amatucci, A.G., Beamish, H.J., Gell, D., Xiang, X.H., Torres Arzayus, M.I., Priestley, A., Jackson, S.P., Marshak Rothstein, A., Jeggo, P.A., *et al.* (1998). Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. *Immunity* 9, 355-366.
- Taccioli, G.E., Gottlieb, T.M., Blunt, T., Priestley, A., Demengeot, J., Mizuta, R., Lehmann, A.R., Alt, F.W., Jackson, S.P., and Jeggo, P.A. (1994). Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. *Science* 265, 1442-1445.
- Takeda, Y., Yashima, K., Hayashi, A., Sasaki, S., Kawaguchi, K., Harada, K., Murawaki, Y., and Ito, H. (2012). Expression of AID, P53, and Mlh1 proteins in endoscopically resected differentiated-type early gastric cancer. *World J Gastrointest Oncol* 4, 131-137.
- Tashiro, J., Kinoshita, K., and Honjo, T. (2001). Palindromic but not G-rich sequences are targets of class switch recombination. *Int Immunol* 13, 495-505.

- Teng, B., Burant, C.F., and Davidson, N.O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* 260, 1816-1819.
- Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 Is a Negative Regulator of Activation-Induced Cytidine Deaminase. *Immunity*.
- Thompson, C.B., and Neiman, P.E. (1987). Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. *Cell* 48, 369-378.
- Tian, M., and Alt, F.W. (2000). Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. *The Journal of biological chemistry* 275, 24163-24172.
- Tonegawa, S. (1983). Somatic generation of antibody diversity. *Nature* 302, 575-581.
- Tran, T.H., Nakata, M., Suzuki, K., Begum, N.A., Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress *Aicda* by overcoming the effects of silencers. *Nature immunology* 11, 148-154.
- Uchimura, Y., Barton, L.F., Rada, C., and Neuberger, M.S. (2011). REG-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. *The Journal of experimental medicine*.
- Unniraman, S., Zhou, S., and Schatz, D.G. (2004). Identification of an AID-independent pathway for chromosomal translocations between the Igh switch region and *Myc*. *Nat Immunol* 5, 1117-1123.
- Vanti, M., Gallastegui, E., Respaldiza, I., Rodriguez-Gil, A., Gomez-Herreros, F., Jimeno-Gonzalez, S., Jordan, A., and Chavez, S. (2009). Yeast genetic analysis reveals the involvement of chromatin reassembly factors in repressing HIV-1 basal transcription. *PLoS Genet* 5, e1000339.
- Vardimon, L., Kressmann, A., Cedar, H., Maechler, M., and Doerfler, W. (1982). Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proceedings of the National Academy of Sciences of the United States of America* 79, 1073-1077.
- Vega, H., Trainer, A.H., Gordillo, M., Crosier, M., Kayserili, H., Skovby, F., Uzielli, M.L., Schnur, R.E., Manouvrier, S., Blair, E., *et al.* (2010). Phenotypic variability in 49 cases of ESCO2 mutations, including novel missense and codon deletion in the acetyltransferase domain, correlates with ESCO2 expression and establishes the clinical criteria for Roberts syndrome. *J Med Genet* 47, 30-37.
- Verkade, H.M., Bugg, S.J., Lindsay, H.D., Carr, A.M., and O'Connell, M.J. (1999). Rad18 is required for DNA repair and checkpoint responses in fission yeast. *Mol Biol Cell* 10, 2905-2918.
- Vrouwe, M.G., Elghalbzouri-Maghrani, E., Meijers, M., Schouten, P., Godthelp, B.C., Bhuiyan, Z.A., Redeker, E.J., Mannens, M.M., Mullenders, L.H., Pastink, A., *et al.* (2007). Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. *Hum Mol Genet* 16, 1478-1487.
- Vuong, B.Q., Lee, M., Kabir, S., Irimia, C., Macchiarulo, S., McKnight, G.S., and Chaudhuri, J. (2009). Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat Immunol* 10, 420-426.
- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., *et al.* (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* 12, 343-356.
- Wade, P.A., Werel, W., Fentzke, R.C., Thompson, N.E., Leykam, J.F., Burgess, R.R., Jaehning, J.A., and Burton, Z.F. (1996). A novel collection of accessory factors associated with yeast RNA polymerase II. *Protein Expr Purif* 8, 85-90.
- Waizenegger, I.C., Hauf, S., Meinke, A., and Peters, J.M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* 103, 399-410.
- Wang, H., Rosidi, B., Perrault, R., Wang, M., Zhang, L., Windhofer, F., and Iliakis, G. (2005). DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. *Cancer Res* 65, 4020-4030.
- Wang, L., Whang, N., Wuerffel, R., and Kenter, A.L. (2006a). AID-dependent histone acetylation is detected in immunoglobulin S regions. *The Journal of experimental medicine* 203, 215-226.
- Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *The Journal of experimental medicine* 206, 1817-1830.

- Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H., and Iliakis, G. (2006b). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic acids research* 34, 6170-6182.
- Ward, I.M., Minn, K., Jorda, K.G., and Chen, J. (2003). Accumulation of checkpoint protein 53BP1 at DNA breaks involves its binding to phosphorylated histone H2AX. *The Journal of biological chemistry* 278, 19579-19582.
- Wardemann, H., Yurasov, S., Schaefer, A., Young, J.W., Meffre, E., and Nussenzweig, M.C. (2003). Predominant autoantibody production by early human B cell precursors. *Science* 301, 1374-1377.
- Watrin, E., and Peters, J.M. (2009). The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *The EMBO journal* 28, 2625-2635.
- Weller, S., Faili, A., Aoufouchi, S., Gueranger, Q., Braun, M., Reynaud, C.A., and Weill, J.C. (2003). Hypermutation in human B cells in vivo and in vitro. *Annals of the New York Academy of Sciences* 987, 158-165.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., *et al.* (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796-801.
- White, D.E., Negorev, D., Peng, H., Ivanov, A.V., Maul, G.G., and Rauscher, F.J., 3rd (2006). KAP1, a novel substrate for PIKK family members, colocalizes with numerous damage response factors at DNA lesions. *Cancer Res* 66, 11594-11599.
- Wiesendanger, M., Kneitz, B., Edelman, W., and Scharff, M.D. (2000). Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. *The Journal of experimental medicine* 191, 579-584.
- Willmann, K.L., Milosevic, S., Pauklin, S., Schmitz, K.M., Rangam, G., Simon, M.T., Maslen, S., Skehel, M., Robert, I., Heyer, V., *et al.* (2012). A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *The Journal of experimental medicine* 209, 2099-2111.
- Winston, F., Chaleff, D.T., Valent, B., and Fink, G.R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*. *Genetics* 107, 179-197.
- Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *The Journal of biological chemistry* 278, 34739-34742.
- Wu, N., and Yu, H. (2012). The Smc complexes in DNA damage response. *Cell Biosci* 2, 5.
- Wu, X., Geraldes, P., Platt, J.L., and Cascalho, M. (2005). The double-edged sword of activation-induced cytidine deaminase. *Journal of immunology* 174, 934-941.
- Wu, X., and Stavnezer, J. (2007). DNA polymerase beta is able to repair breaks in switch regions and plays an inhibitory role during immunoglobulin class switch recombination. *The Journal of experimental medicine* 204, 1677-1689.
- Wuerffel, R., Wang, L., Grigera, F., Manis, J., Selsing, E., Perlot, T., Alt, F.W., Cogne, M., Pinaud, E., and Kenter, A.L. (2007). S-S synapsis during class switch recombination is promoted by distantly located transcriptional elements and activation-induced deaminase. *Immunity* 27, 711-722.
- Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev* 17, 654-663.
- Xiao, T., Kao, C.F., Krogan, N.J., Sun, Z.W., Greenblatt, J.F., Osley, M.A., and Strahl, B.D. (2005). Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Molecular and cellular biology* 25, 637-651.
- Xie, A., Kwok, A., and Scully, R. (2009). Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat Struct Mol Biol* 16, 814-818.
- Xiong, B., Lu, S., and Gerton, J.L. (2010). Hos1 is a lysine deacetylase for the Smc3 subunit of cohesin. *Curr Biol* 20, 1660-1665.
- Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qahtani, A., White, C.A., Park, S.R., *et al.* (2010). 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat Struct Mol Biol* 17, 1124-1135.
- Xu, Z., Zan, H., Pone, E.J., Mai, T., and Casali, P. (2012). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat Rev Immunol* 12, 517-531.

- Xue, K., Rada, C., and Neuberger, M.S. (2006). The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*<sup>-/-</sup> *ung*<sup>-/-</sup> mice. *The Journal of experimental medicine* 203, 2085-2094.
- Yadav, A., Olaru, A., Saltis, M., Setren, A., Cerny, J., and Livak, F. (2006). Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. *Mol Immunol* 43, 529-541.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999a). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* 97, 41-51.
- Yamaguchi, Y., Wada, T., Watanabe, D., Takagi, T., Hasegawa, J., and Handa, H. (1999b). Structure and function of the human transcription elongation factor DSIF. *The Journal of biological chemistry* 274, 8085-8092.
- Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2010). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* 12, 62-69.
- Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat Immunol* 12, 62-69.
- Yamane, A., Robbiani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C., *et al.* (2013). RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. *Cell Rep* 3, 138-147.
- Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P., *et al.* (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature* 449, 478-482.
- Yancopoulos, G.D., and Alt, F.W. (1985). Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* 40, 271-281.
- Yancopoulos, G.D., DePinho, R.A., Zimmerman, K.A., Lutzker, S.G., Rosenberg, N., and Alt, F.W. (1986). Secondary genomic rearrangement events in pre-B cells: VHDJH replacement by a LINE-1 sequence and directed class switching. *The EMBO journal* 5, 3259-3266.
- Yart, A., Gstaiger, M., Wirbelauer, C., Pecnik, M., Anastasiou, D., Hess, D., and Krek, W. (2005). The HRPT2 tumor suppressor gene product parafibromin associates with human PAF1 and RNA polymerase II. *Molecular and cellular biology* 25, 5052-5060.
- Yazdi, P.T., Wang, Y., Zhao, S., Patel, N., Lee, E.Y., and Qin, J. (2002). SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. *Genes Dev* 16, 571-582.
- Ye, J. (2004). The immunoglobulin IGHD gene locus in C57BL/6 mice. *Immunogenetics* 56, 399-404.
- Yoh, S.M., Cho, H., Pickle, L., Evans, R.M., and Jones, K.A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct *lws1*-dependent mRNA splicing and export. *Genes Dev* 21, 160-174.
- Yoh, S.M., Lucas, J.S., and Jones, K.A. (2008). The *lws1*:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev* 22, 3422-3434.
- Youdell, M.L., Kizer, K.O., Kisseleva-Romanova, E., Fuchs, S.M., Duro, E., Strahl, B.D., and Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Molecular and cellular biology* 28, 4915-4926.
- 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.
- Yu, K., Huang, F.T., and Lieber, M.R. (2004). DNA substrate length and surrounding sequence affect the activation-induced deaminase activity at cytidine. *The Journal of biological chemistry* 279, 6496-6500.
- Zan, H., White, C.A., Thomas, L.M., Mai, T., Li, G., Xu, Z., Zhang, J., and Casali, P. (2012). Rev1 recruits *ung* to switch regions and enhances *du* glycosylation for immunoglobulin class switch DNA recombination. *Cell Rep* 2, 1220-1232.
- Zaprazna, K., and Atchison, M.L. (2012). YY1 controls immunoglobulin class switch recombination and nuclear activation-induced deaminase levels. *Molecular and cellular biology* 32, 1542-1554.
- Zha, S., Alt, F.W., Cheng, H.L., Brush, J.W., and Li, G. (2007). Defective DNA repair and increased genomic instability in *Cernunnos-XLF*-deficient murine ES cells. *Proceedings of the National Academy of Sciences of the United States of America* 104, 4518-4523.

- Zha, S., Guo, C., Boboila, C., Oksenyich, V., Cheng, H.L., Zhang, Y., Wesemann, D.R., Yuen, G., Patel, H., Goff, P.H., *et al.* (2011). ATM damage response and XLF repair factor are functionally redundant in joining DNA breaks. *Nature* **469**, 250-254.
- Zhang, J., Bottaro, A., Li, S., Stewart, V., and Alt, F.W. (1993). A selective defect in IgG2b switching as a result of targeted mutation of the I gamma 2b promoter and exon. *The EMBO journal* **12**, 3529-3537.
- Zhang, J., Shi, X., Li, Y., Kim, B.J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S.Y., Wang, Y., *et al.* (2008). Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. *Mol Cell* **31**, 143-151.
- Zhao, X., and Blobel, G. (2005). A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4777-4782.
- Zhu, B., Mandal, S.S., Pham, A.D., Zheng, Y., Erdjument-Bromage, H., Batra, S.K., Tempst, P., and Reinberg, D. (2005). The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev* **19**, 1668-1673.
- Zhu, Y., Nonoyama, S., Morio, T., Muramatsu, M., Honjo, T., and Mizutani, S. (2003). Type two hyper-IgM syndrome caused by mutation in activation-induced cytidine deaminase. *J Med Dent Sci* **50**, 41-46.
- Ziv, Y., Bielopolski, D., Galanty, Y., Lukas, C., Taya, Y., Schultz, D.C., Lukas, J., Bekker-Jensen, S., Bartek, J., and Shiloh, Y. (2006). Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nat Cell Biol* **8**, 870-876.
- Zonana, J., Elder, M.E., Schneider, L.C., Orlow, S.J., Moss, C., Golabi, M., Shapira, S.K., Farndon, P.A., Wara, D.W., Emmal, S.A., *et al.* (2000). A novel X-linked disorder of immune deficiency and hypohidrotic ectodermal dysplasia is allelic to incontinentia pigmenti and due to mutations in IKK-gamma (NEMO). *Am J Hum Genet* **67**, 1555-1562.

# **ANNEX I**

# Materials and methods

## Cell lines

EBV-immortalized B cells were obtained from the laboratory of Lymphocyte interactions and lymphocytes B terminal maturation headed by Dr. Anne Durandy (Necker Hospital, Paris) and isolated from peripheral blood mononuclear cells of eight healthy donors (Ctr), one patient affected by CSR-ID due to a loss of AID ( $AID^{-/-}$ ) and four patients affected by CSR-ID due to a specific defect in class switch recombination (CSR-ID). Human B cells and mouse CH12 B cells were cultured in RPMI 1640 medium with 10% FCS, 10 mM Hepes, 1 mM sodium pyruvate, penicillin, streptomycin and 50  $\mu$ M 2-mercaptoethanol.

## Transcriptome analysis

Total RNA was extracted from human EBV-immortalized B cells using RNeasy Mini Kit (Qiagen), according to the manufacturer's protocol. RNA quality was verified through Agilent 2100 Bioanalyzer and the DGE cDNA library prepared according to the Illumina pipeline. Sequencing was performed through the Illumina Genome Analyzer and data obtained have been annotated through Illumina software. Data analysis was performed with DESeq (see references). Briefly, sample variance was estimated according to the samples mean and differentially expressed genes were obtained. Correlation between samples belonging to the same group was evaluated by calculating Pearson correlation coefficients, which indicate whether there is any dependence on two individual data sets and whose values are ranging from 1 (high correlation) to -1 (low correlation). Correlation between data obtained from controls and patients was calculated by applying the variance stabilizing transformation (VST) function, and data were plotted as heat map to evaluate the homogeneity between each group of samples. Differentially expressed genes were filtered according to the p value adjusted (padj), namely p value adjusted for multiple testing according to the Benjamini-Hochberg procedure which controls the false discovery rate. The filters applied were  $padj < 0.05$  for controls vs. patients and controls vs.  $AID^{-/-}$  and  $padj < 0.01$  for patients vs.  $AID^{-/-}$  as, in this latter case, no replicates for the AID-deficient cell line were included in the analysis and thus a more stringent condition allowed us to reduce the number of false discoveries. In order to obtain the list of differentially expressed genes for each condition, data were filtered according to the  $\log_2FC < 1$  (downregulated) and  $\log_2FC > 1$  (upregulated) and plotted as scatter plot. The differentially expressed genes obtained for each condition were analyzed through the Ingenuity Pathway Analysis (IPA) software and further compared to identify common deregulated genes within the conditions analyzed.

## Retroviral transduction

shRNA sequences targeting *Aicda* and *Spt6* (Table I) were cloned into the LMP vector (Open Biosystems) according to the methods described by Paddison et al. (Paddison et al., 2004). BOSC23 cells were plated ( $1.5 \times 10^5$  cells/ml) and, after 24h, transfected with 1  $\mu$ g of shRNA vector and 1  $\mu$ g of pCL-Ampho helper plasmid (Imgenex) by using FuGENE (Promega). After 48h,  $1 \times 10^6$  CH12 B cells were infected with the viral supernatant supplemented with Hepes (20 mM) and polybrene (10  $\mu$ g/ml) and spinned at 1150 x *g* for 90 min at RT; BOSC23 cells were harvested and analyzed for GFP expression, to estimate the transfection efficiency. Transduced CH12 B cells were harvested 24h post-infection and selected with 0.5  $\mu$ g/ml puromycin for 3-5 days before performing CSR assay; an aliquot was analyzed for GFP expression, to estimate the infection efficiency.

## Lentiviral transduction

Lentivirus shRNAs targeting *Spt6* and the non-target control were obtained from Sigma; forward and reverse oligos harboring the shRNA sequence targeting *Aicda*, *Smc5* and *Smc6* were annealed and cloned into the pLKO.1-turboGFP vector harboring an adapter by using *AgeI/EcoRI* restriction sites (see Table I for list of shRNAs). Lenti-X 293T cells (Clontech) were plated ( $1.5 \times 10^5$  cells/ml) and transfected after 24h with 1  $\mu$ g of shRNA vector, 0.9  $\mu$ g of pCMVdR8.91 and 0.1  $\mu$ g of VSV-G helper plasmids (Addgene) by using FuGENE (Promega). After 48h,  $1 \times 10^6$  CH12 B cells were infected with the viral supernatant supplemented with Hepes (20 mM) and polybrene (10  $\mu$ g/ml), spinned at 1150 x *g* for 90 min at RT and incubated for 4h at 37°C. The culture was then diluted with complete RPMI medium by adding 2.5X of the initial volume. Lenti-X 293T cells were harvested and analyzed for GFP expression, to estimate the transfection efficiency. Transduced CH12 B cells were harvested 48h post-infection and selected with 1  $\mu$ g/ml puromycin for 2-5 days before performing CSR assay, by splitting cells every day in order to improve the selection efficiency. An aliquot was analyzed for GFP expression, to estimate the infection efficiency.

## Class switch recombination assay

CH12 B cells have been plated ( $1 \times 10^5$  cells/ml) with RPMI medium supplemented with 5 ng/ml IL-4 (Sigma-Aldrich), 1 ng/ml TGF $\beta$  (R&D Biosystems), 200 ng/ml anti-CD40 (eBiosciences) and 1  $\mu$ g/ml puromycin. Cells have been harvested after 48h or 72h and stained with PE anti-IgA antibody (SouthernBiotech). Dead cells have been excluded from the analysis by staining with 50 nM ToPro-3 (Invitrogen). Data were collected on a FACSCalibur (BD) and analyzed with FlowJo software (Tree Star, Inc.).



## Real time quantitative RT-PCR

Total RNA was extracted from EBV-immortalized human B cells or CH12 B cells using TRIzol reagent (Life Technologies) and cDNA was prepared according to the SuperScript II (Invitrogen) manufacturer's protocol. qPCR was performed in triplicates using the Universal Probe Library (UPL) system (Roche) and a LightCycler 480 (Roche). Transcript quantities were calculated relative to standard curves and normalized to *GAPDH* or *HPRT* mRNA using the  $\Delta\Delta C_t$  method (see Table II for list of primers and probes).

**Table I. List of shRNAs used in this study**

### Retrovirus-mediated knockdown

shRNA	Targeted sequence (5'-3')	Reference
non-target		Open Biosystems
sh-AID	ACCAGTCGCCATTATAATGCAA	
sh-Spt6 #1	CCCGATGCTATTTATTCAGTTT	TRCN00000931
sh-Spt6 #2	CGCTGAGAATCCGCAAAGAGAA	
sh-Spt6 #3	CGCTGACTGGATCTATAGAAAT	
sh-Spt6 #4	CACTATGACTTTGATGCGGAAG	

### Lentivirus-mediated knockdown

shRNA	Targeted sequence (5'-3')	Reference
non-target	CGTGATCTTCACCGACAAGAT	
sh-AID	GCGAGATGCATTTTCGTATGTT	TRCN0000112031
sh-Spt6 #1	CCGATGCTATTTATTCAGTTT	TRCN0000093118
sh-Spt6 #2	CGTATCCAAGACCTCTGATA	TRCN0000093121
sh-Spt6 #3	GTCCATAAAGTGGCGTCAAAT	TRCN0000306284
sh-Spt6 #4	AGAGCTCAGTTGTAGGTATAA	TRCN0000306343
sh-Spt6 #5	CGGATCATGAAGATCGATATT	TRCN0000332144
sh-Smc5 #1	GGAACCTTCAGCAGGCATTAAC	TRCN0000241747
sh-Smc5 #2	TCATATCTTCGGGAGTTATTT	TRCN0000241749
sh-Smc5 #3	CCCATAATGCTCACGATTAAT	TRCN0000241750
sh-Smc5 #4	ACGGAGTGTGAGTGATCATAT	TRCN0000241751
sh-Smc5 #5	TCAGGGTATGGACCCAATTAA	TRCN0000241748
sh-Smc6 #1	CCTACCTTGATCTGGATAATA	TRCN0000113215
sh-Smc6 #2	GCCTTTAATGACGCTGAGGTT	TRCN0000113216
sh-Smc6 #3	CCGAGTTAGGAAAGAAGATAT	TRCN0000113217
sh-Smc6 #4	CGTGGATGGAAGTCGATCTTA	TRCN0000113218
sh-Smc6 #5	CGGAGACAATTTACCCATAAA	TRCN0000113219

**Table II. List of RT-qPCR primers and probes used in this study****DGE validation**

<b>Primer</b>	<b>Sequence (5'-3')</b>	<b>UPL probe</b>
GAPDH Fwd	AGCCACATCGCTCAGACAC	60
GAPDH Rev	GCCCAATACGACCAAATCC	60
AID Fwd	CTCTGGACACCACTATGGACAG	69
AID Rev	GCGGACATTTTTGAATTGGT	69
PSMA4 Fwd	TGGGAATTTTAGCAAATGATGG	84
PSMA4 Rev	CACTGCAAGCCATGTCCTC	84
MRPS6 Fwd	CGAGCTGGCTTTAATCCTGA	82
MRPS6 Rev	AGGTTTTCCAAGTCCCTCACT	82
SPARC Fwd	TTCCCTGTACACTGGCAGTTC	36
SPARC Rev	AATGCTCCATGGGGATGA	36
IL-1 $\alpha$ Fwd	GGTTGAGTTTAAGCCAATCCA	6
IL-1 $\alpha$ Rev	TGCTGACCTAGGCTTGATGA	6
P2RX1 Fwd	TACGTGGTGCAAGAGTCAGG	9
P2RX1 Rev	CCAGGTCACAGTGCCAGTC	9
PFN2 Fwd	AGGTGGGGAGCCAACATAC	5
PFN2 Rev	CCCCTTCTTTTCCATTACAA	5
SERPING1 Fwd	CATCGCCAGCCTCCTTAC	15
SERPING1 Rev	GAGGATGCTCTCCAGGTTTG	15
LAD1 Fwd	CTCCCACCCGTCACACTC	34
LAD1 Rev	CTGCTGTAGGTTTCGCTGTGT	34
TESC Fwd	CCCTCACATCGAGAAGGAGT	43
TESC Rev	GTGATCCCCCTCGTACACCTG	43
PRF1 Fwd	CCGCTTCTCTATACGGGATTC	79
PRF1 Rev	GCAGCAGCAGGAGAAGGAT	79

**Smc6 knockdown**

<b>Primer</b>	<b>Sequence (5'-3')</b>	<b>UPL probe</b>
HPRT Fwd	GTCAACGGGGGACATAAAAAG	22
HPRT Rev	CAACAATCAAGACATTCTTTCCA	22
Smc6 Fwd	GACGAGAGCTTGACATGAAGG	2
Smc6 Rev	GGCAGATCTGTCTGCCTGT	2

**Table III. List of antibodies used in this study**

<b>Antibody</b>	<b>Clone</b>	<b>Source</b>	<b>Use</b>
AID	Strasbg 9 (AID-2E11)	IGBMC (Jeevan-Raj et al., 2011)	WB, IP
Flag	M2	Sigma	WB, IP
$\beta$ -Actin	A1978	Sigma	WB
Nbs1	gift from M. Nussenzweig		
KAP1	1TB 1A9	IGBMC (Jeevan-Raj et al., 2011)	WB
Spt5	sc-28678	Santa Cruz	WB
Spt6	NB100-2582	Novus Biologicals	WB
Smc5	ab18038	AbCam	WB

WB: western blot; IP: immunoprecipitation

## References

Simon Anders (2010), Analysing RNA-Seq data with the “DESeq” package, Bioconductor packages

Paddison, P.J., Cleary, M., Silva, J.M., Chang, K., Sheth, N., Sachidanandam, R., Hannon, G.J. (2004). Cloning of short hairpin RNAs for gene knockdown in mammalian cells. *Nature methods* 1(2):163-7

# **ANNEX II**

# Résumé de thèse

## Introduction

Lors des réponses immunitaires, le répertoire des lymphocytes B est diversifié par les mécanismes d'hypermutation somatique (HMS) et de commutation isotypique (CI) (Di Noia and Neuberger, 2007; Xu et al., 2012). L'HMS modifie l'affinité des anticorps pour l'antigène, en favorisant une production d'anticorps hautement spécifiques pour la reconnaissance des agents pathogènes par introduction de mutations dans la région variable des gènes codant pour les chaînes lourdes (IgH) et légères (IgL) des immunoglobulines (Ig) (Di Noia and Neuberger, 2007). La CI modifie l'isotype des anticorps pour adapter la réponse au type d'antigène et à la voie d'accès dans l'organisme. Elle constitue un événement de recombinaison qui a lieu dans les gènes IgH. L'HMS ainsi que la CI sont dépendantes de l'expression d'« activation-induced cytidine deaminase » (AID), exprimée par les lymphocytes B matures après leur rencontre avec l'antigène (Muramatsu et al., 2000; Revy et al., 2000). AID est une déaminase qui convertit les cytosines en uraciles au niveau de l'ADN, en introduisant des lésions (mésappariements dU :dG) dans les gènes des Ig (Petersen-Mahrt et al., 2002). Ces lésions, au cours de la CI, sont localisées au niveau de régions S (de « switch ») qui sont hautement répétitives et activement transcrites avant recombinaison (Stavnezer-Nordgren and Sirlin, 1986). Les mésappariements dU :dG sont réparés directement par excision de base ou par la voie de réparation des mésappariements, et donnent lieu à des cassures double-brins ou à des mutations dans le locus IgH (Petersen-Mahrt et al., 2002). Ces cassures vont être réparées par jonction d'extrémités non homologues (« non-homologous end-joining », NHEJ), et résultent en l'expression d'anticorps d'un isotype différent (Ramiro et al., 2007). L'action d'AID peut être très dangereuse pour les cellules : la surexpression d'AID et les dommages produits dans l'ADN peuvent causer des maladies auto-immunes, de même les cassures double-brins générées au cours de la déamination par AID peuvent favoriser des translocations avec des oncogènes et conduisent au développement de cancers (Okazaki et al., 2007). C'est pourquoi l'activité d'AID est finement régulée au niveau de son expression (Muramatsu et al., 1999) mais aussi au niveau de sa localisation grâce à la présence de signaux de localisation nucléaire (NLS) dans le domaine N-terminal ainsi que d'exportation nucléaire (NES) dans le domaine C-terminal (Ito et al., 2004; Patenaude et al., 2009). De plus, de nombreux facteurs sont décrits comme impliqués dans sa régulation: la protéine kinase A (PKA) (Vuong et al., 2009), la protéine de réplication A (RPA) (Chaudhuri et al., 2004), CTNBL1 (Conticello et al., 2008), Hsp90 (Orthwein et al., 2010), KAP1 (Jeevan-Raj et al., 2011) sont des exemples. Egalement, l'étude et la caractérisation de patients atteints d'une immunodéficience des lymphocytes B (« hyper-IgM syndrome », HIGM) ont été cruciales pour définir son rôle dans l'HMS et la CI ainsi que dans les voies de réparation de l'ADN impliquées dans la résolution des lésions induites (Durandy et al., 2007; Kracker et al., 2010). En effet, les mutations dans le gène codant pour AID causent un défaut d'HMS

et CI (Durandy et al., 2007; Kracker et al., 2010; Revy et al., 2000) ; par contre, une délétion du domaine C-terminal d'AID cause un défaut spécifique de la CI (Barreto et al., 2003; Doi et al., 2009; Durandy et al., 2007; Shinkura et al., 2004; Ta et al., 2003). De plus, des patients atteints d'une nouvelle forme d'immunodéficiences présentent un défaut spécifique de la CI (ID-CI) qui n'est pas lié à la carence d'AID ni d'autres facteurs notamment impliqués dans ce processus (Imai et al., 2003). Malgré l'expression d'AID, les cassures double-brins ne sont pas détectées au niveau des gènes IgH, suggérant qu'AID n'est pas ciblée de manière appropriée aux régions S (Durandy et al., 2007; Imai et al., 2003; Kracker et al., 2010).

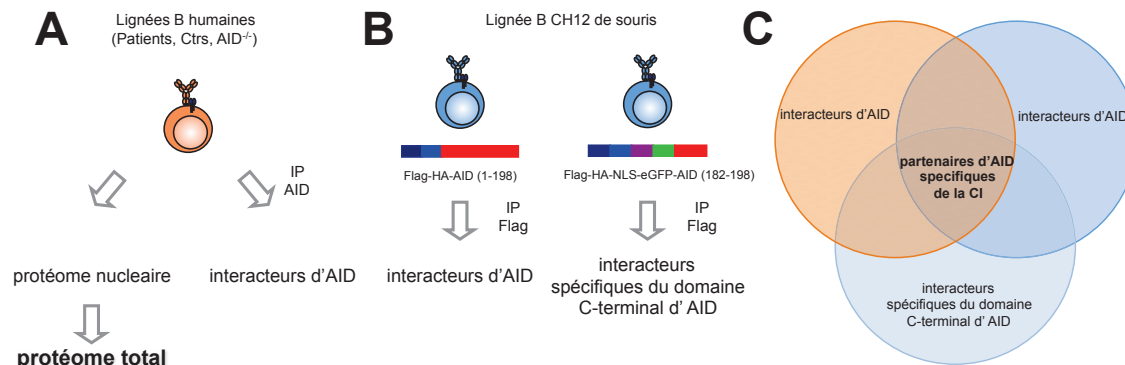
En dépit de nombreuses avancées récentes, les mécanismes moléculaires de la régulation d'AID aux loci Ig ne sont pas totalement définis. Afin de comprendre quels sont les facteurs nécessaires pour cibler AID au locus IgH pendant la CI, nous avons entrepris l'étude des lymphocytes B isolés chez les patients ID-CI en analysant leur transcriptome et protéome ; ainsi, nous avons intégré ces études avec un modèle *in vitro* de CI en utilisant la lignée de lymphome B murin CH12.

## Résultats

### 1. Identification de nouveaux partenaires d'AID impliqués dans la CI

Dans un premier temps, afin d'identifier les cofacteurs d'AID spécifiques de la CI, nous avons établi une collaboration avec l'Hôpital Necker (Paris) et nous nous sommes focalisés sur les patients ID-CI, en analysant leur transcriptome par « Digital Gene Expression-tag profiling » (DGE). Notre hypothèse était que, si chez ces patients la cause du défaut de CI était un manque d'un cofacteur d'AID causé par une mutation, nous pourrions être capable d'identifier des gènes sous exprimés ou surexprimés par rapport aux contrôles. De plus, nos collaborateurs étaient en charge de l'analyse génétique des patients et de leurs familles, en cherchant des mutations. En comparant les patients et leurs contrôles - des individus sains et ainsi un patient atteint d'une déficience d'AID (AID<sup>-/-</sup>) et en conséquence ayant un défaut d'HMS et de CI - nous avons identifié 280 gènes surexprimés et 90 sous-exprimés chez les patients ID-CI par rapport aux contrôles. Des gènes montrant une dérégulation chez les patients ID-CI ont été séquencés par nos collaborateurs afin d'identifier des mutations. Cependant, aucune mutation n'a été identifiée parmi les gènes candidats. Afin d'identifier les mécanismes de la déficience ID-CI, nous avons entrepris une approche protéomique. Nous avons analysé le protéome nucléaire par spectrométrie de masse, afin d'identifier les protéines qui sont exprimées de façon différente chez les patients ID-CI par rapport aux contrôles (Figure 1). Pour limiter notre recherche aux cofacteurs d'AID spécifiques de la CI, nous avons comparé les données obtenues avec l'étude du protéome des cellules B humaines avec l'ensemble des partenaires d'AID identifiés chez une lignée de souris de lymphome B, les cellules CH12. Nous avons généré des lignées CH12 qui expriment la protéine AID murine comportant les épitopes Flag et HA en N-terminal (AID<sup>Flag-HA</sup>) ou exprimant seulement les 17 derniers résidus du domaine C-terminal d'AID fusionnés à

la protéine eGFP (eGFP-AID<sup>182-198</sup>). Après immunoprécipitation d'AID, nous avons identifié ses partenaires nucléaires par spectrométrie de masse (Figure 1B). De cette façon, nous avons été capables de comparer les partenaires d'AID dans les cellules B humaines et murines avec les protéines qui interagissent avec le domaine C-terminal d'AID, qui a été montré nécessaire pour la CI (Barrete et al., 2003; Dei et al., 2009; Durandy et al., 2007; Shinkura et al., 2004; Ta et al., 2003).



**Figure 1. Stratégie d'identification des partenaires d'AID spécifiques de la CI par analyse protéomique des cellules B humaines et murines**

Schéma expliquant l'analyse protéomique conduite sur les cellules B humaines et murines. (A) Les protéines nucléaires totales ont été isolées des lignées de cellules B humaines immortalisées avec le virus Epstein-Barr (EBV) et obtenues de patients ID-CI (Patients), des individus sains (Ctrs) et d'un patient AID<sup>-/-</sup> (AID<sup>-/-</sup>) après ou en absence d'immunoprécipitation d'AID et identifiées par spectrométrie de masse. Les données obtenues ont été croisées avec les données résultant de l'analyse par spectrométrie de masse de (B) cellules B murines CH12 exprimant la protéine AID murine comportant les épitopes Flag et HA en N-terminal (Flag-HA-AID<sup>(1-198)</sup>) ou exprimant seulement les 17 derniers résidus du domaine C-terminal d'AID fusionnés à la protéine eGFP et à un signal de localisation nucléaire (NLS) pour une localisation appropriée dans les cellules (Flag-HA-NLS-eGFP-AID<sup>(182-198)</sup>). (C) Diagramme de Venn montrant comment l'intégration des données obtenues par l'analyse des cellules B humaines et murines permet l'identification des facteurs qui sont dans le même complexe qu'AID et qui sont spécifiquement nécessaires pour la régulation de la CI.

Nous avons identifié une liste de facteurs connus pour être importants dans la régulation de la CI et/ou notamment pour leur interaction avec AID, mais aussi d'autres facteurs avec un rôle non encore décrit lors de la diversification des immunoglobulines. De façon intéressante, la protéine Spt6 est identifiée dans les cellules B humaines isolées d'individus sains mais est absente chez les cellules B des patients ID-CI et le patient AID<sup>-/-</sup>. Spt6 est une chaperonne d'histones, impliquée dans la transcription et capable d'interagir avec Spt5 et l'ARN polymérase II en pause (Andrulis et al., 2002; Endeh et al., 2004; Kregan et al., 2002) et, de plus, Spt5 a été décrit comme impliqué dans la CI (Pavri et al., 2010). Ces données suggèrent donc que Spt6 pourrait être le facteur responsable de la déficience de CI observée chez les patients ID-CI.

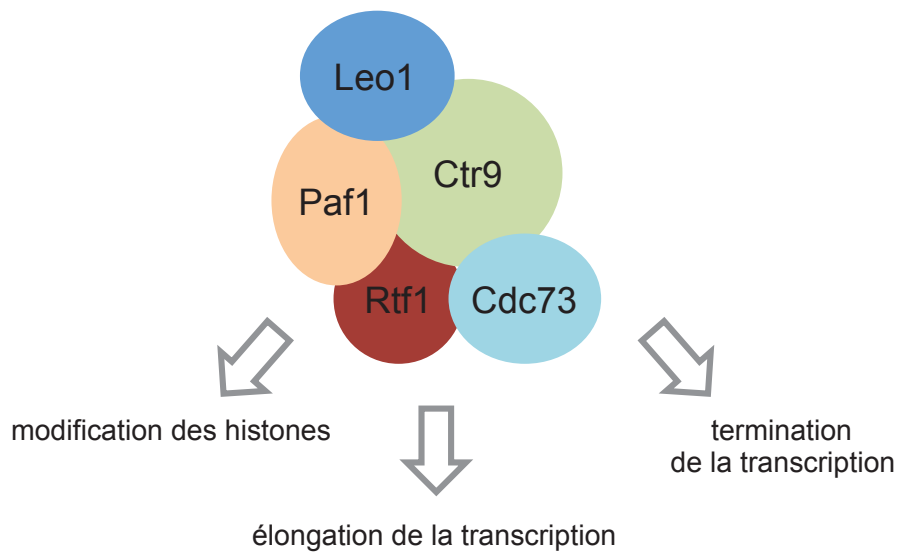
## 2. Spt6 est un nouveau régulateur de la commutation isotypique des immunoglobulines

Nous avons vérifié, d'abord, l'expression de Spt6 dans les cellules B isolées des patients ID-CI et les contrôles et nous avons observé une expression très variable. Malgré tout, il apparaît que Spt6 est sous exprimé dans les patients ID-CI. Par immunoprécipitation d'AID, nous avons montré une interaction entre Spt6 et AID dans la lignée de cellules murines CH12 et nous avons vérifié si la réduction de l'expression de Spt6 pouvait causer des conséquences au niveau de la CI. La lignée CH12 est un système très efficace pour l'étude de la CI car après stimulation avec des cytokines (notamment IL-4, CD40L et TGF $\beta$ ), il est possible de quantifier l'efficacité de CI par marquage des IgA de surface et analyse par cytométrie en flux (Nakamura et al., 1996). Donc, nous avons utilisé des vecteurs retroviraux et lentiviraux pour générer des lignées CH12 infectées avec des shARN qui ciblent Spt6, un shARN qui cible AID et un « non-target » shARN comme contrôle. Nos résultats, encourageants, montrent une réduction d'efficacité de la CI après inactivation partielle de Spt6, mais une publication du laboratoire de Tasuku Honjo a décrit l'implication spécifique de Spt6 lors de la CI et pas dans l'HMS (Okazaki et al., 2011). Dans l'impossibilité de continuer avec la caractérisation de Spt6 lors de la diversification des immunoglobulines, nos données ont été néanmoins importantes pour un autre projet focalisé sur le complexe PAF.

## 3. Le complexe PAF dans la régulation d'AID

La transcription des régions S du locus des IgH est une des étapes nécessaires de la CI et HMS pour rendre l'ADNss accessible à AID pour la déamination. De nombreux facteurs sont impliqués dans ce processus et ont été décrits comme régulateurs d'AID, notamment le facteur de splicing CTNNBL1 (Conticello et al., 2008), l'ARN exosome (Basu et al., 2011), Spt5 (Pavri et al., 2010), Spt4 (Stanlie et al., 2012) et le complexe FACT (Stanlie et al., 2010). En collaboration avec le laboratoire de Svend Petersen-Mahrt, nous avons contribué à la compréhension du rôle du complexe PAF dans la CI (Willmann et al., 2012). En effet, le complexe PAF a été isolé à partir des protéines associées à la chromatine dans les cellules B de lymphome de poulet DT40, capables de diversifier leurs gènes Ig par conversion génique, et dans la lignée CH12 surexprimant AID<sup>Flag-HA</sup>. L'association avec AID a été confirmée par co-immunoprécipitation d'AID et des sous unités du complexe PAF (PAF1, Leo1 et Ctr9) ainsi que l'association entre AID et Spt5, Spt6 et l'ARN polymérase II. De plus, en absence de Leo1, on observe un défaut de CI dans les cellules CH12 et un recrutement réduit d'AID aux loci Ig (Willmann et al., 2012). Ces données suggèrent que le complexe PAF, avec son rôle de modificateur d'histones au niveau du promoteur des gènes activement transcrits et sa présence pendant l'élongation de la transcription avec Spt5 et Spt6 (Jaehning, 2010), peut réguler la présence d'AID aux régions S transcrites avant la recombinaison.





**Figure 2. Rôles principaux du complexe PAF**

Représentation schématique du complexe PAF humain (hPAF) et ses rôles principaux. Adaptée de (Jaehning, 2010).

#### 4. Caractérisation du complexe Smc5/6 lors de la diversification des immunoglobulines

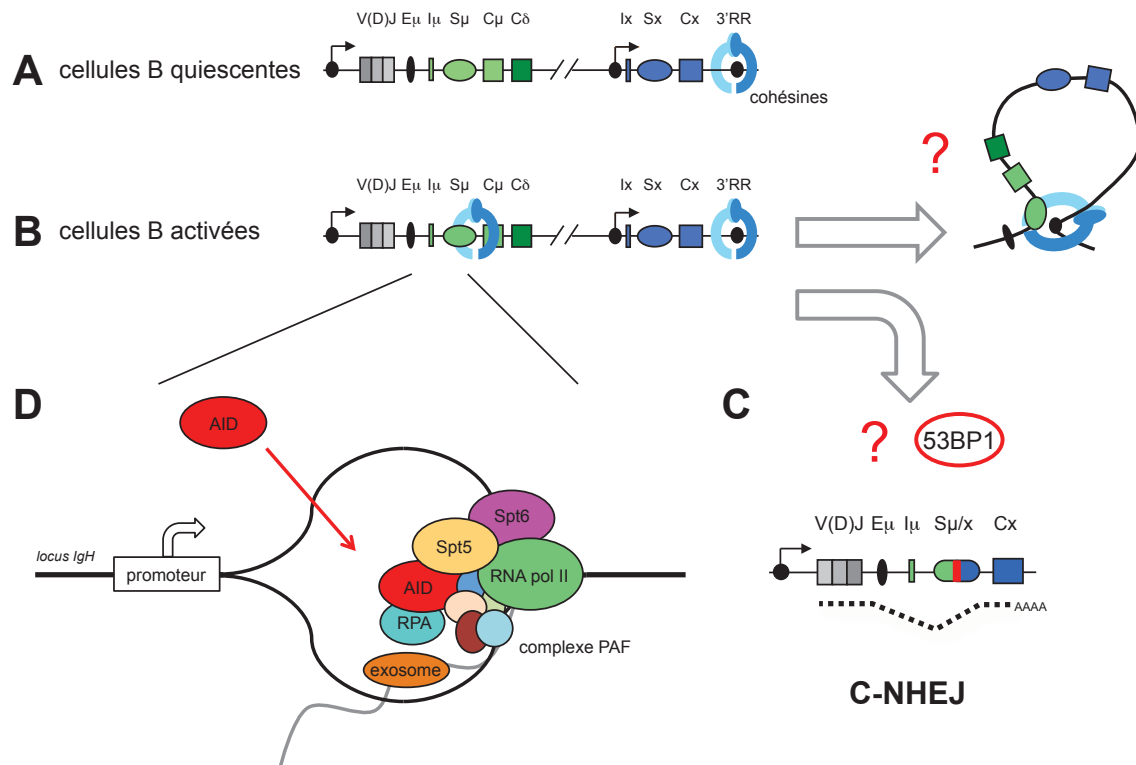
Nos collaborateurs à l'Hôpital Necker ayant découvert une mutation chez un patient ID-CI sur le gène codant pour Smc5, nous avons dirigé notre recherche sur cette protéine, en essayant de comprendre si elle pouvait jouer un rôle dans la CI. Smc5 est un membre de la famille des protéines Smc (Structural maintenance of chromosomes) principalement impliquées dans la régulation de la division cellulaire et la dynamique des chromosomes (Losada and Hirano, 2005). Smc5 forme un hétérodimère avec la protéine Smc6 et ce complexe, parmi différentes fonctions, a été impliqué dans la réparation des cassures double-brins de l'ADN par recombinaison homologue en recrutant les cohésines (Smc1/Smc3) et dans le contrôle du cycle cellulaire (De Piccoli et al., 2006; Harvey et al., 2004; Potts et al., 2006). Nous avons observé une expression variable de Smc5 dans les patients par rapport aux contrôles et nous avons poursuivi des expériences de knockdown pour vérifier si une expression réduite de Smc5 dans la lignée murine CH12 avait un effet sur la CI. Nous avons observé un défaut de la CI d'environ 20%, en fonction de l'efficacité de l'inactivation de Smc5. De plus, pour avoir une vision globale du rôle du complexe Smc5/6, nous avons aussi vérifié si Smc6 était nécessaire pour la CI des immunoglobulines. Nous avons obtenu des résultats identiques à l'inactivation de Smc5. D'autres investigations sont en cours pour clarifier le rôle du complexe Smc5/6 lors de la diversification des anticorps.

## 5. Les cohésines sont impliquées dans la régulation de la CI des immunoglobulines

L'analyse du protéome des cellules B humaines et murines surexprimant AID nous a permis d'obtenir une liste de partenaires potentiels d'AID. Parmi eux, nous avons identifié les cohésines (Smc1/Smc3), qui régulent la cohésion des chromatides pendant la division cellulaire et sont impliquées dans la réparation de l'ADN par recombinaison homologue (Nasmyth and Haering, 2009). De plus, elles régulent l'expression génique chez la *Drosophila* (Dorsett, 2009) et elles ont un rôle structural en favorisant des boucles (loops) entre deux régions très distantes du génome pendant la transcription, dans les premières étapes du développement des lymphocytes B et T (Degner et al., 2011; Guo et al., 2011; Kagey et al., 2010; Seitan et al., 2011). Nous avons donc supposé qu'elles peuvent également être impliquées dans la régulation de la CI des Ig. Nous avons observé que AID est dans un complexe avec les cohésines pendant la CI et que les cohésines sont recrutées de façon active au locus IgH. De plus, le knockdown de Smc1 et Smc3 et des facteurs régulateurs du complexe, Nipbl et Wapal, réduit l'efficacité de CI des cellules CH12 infectées avec lentivirus. En effet, nous avons ainsi observé - par analyse de la séquence génomique des régions S hybrides produites après CI - que la voie classique de réparation privilégiée pendant ce processus de recombinaison (C-NHEJ) était altérée en favorisant une voie alternative (A-NHEJ). Nos résultats suggèrent que les cohésines peuvent jouer un rôle lors de la CI par régulation de la voie de réparation des cassures double brins générées pendant ce processus (manuscrit publié).

## Conclusion et perspectives

Au cours de ce travail de thèse, nous nous sommes focalisés sur les nouveaux régulateurs d'AID lors de la CI. Nous avons identifié Spt6 comme nouveau candidat et nous avons montré que Spt6, Spt5 et le complexe PAF sont associés avec AID dans les cellules B murines CH12. Nous avons aussi investigué le rôle du complexe Smc5/Smc6 et nos études sont encore en cours. Enfin, nous avons également identifié le complexe cohésine comme partenaire d'AID et nous avons observé qu'une carence de cohésines dans les cellules CH12 a un impact sur la CI des cellules B et que leur présence influence la voie de réparation des cassures double-brins de l'ADN pendant la recombinaison (Figure 3). En conclusion, nos études ont aidé à définir un nouveau scénario de régulation d'AID dans sa localisation physiologique du locus IgH. Comprendre comment AID est ciblée au locus IgH et analyser les régulateurs moléculaires qui limitent les dommages collatéraux à l'ADN pourra fournir de nouvelles pistes de recherche pour le développement des thérapies contre le cancer.



### Figure 3. Modèle proposé de CI

Au cours de ces études, nous avons montré que, **(A)** dans les cellules B quiescentes, le complexe cohésine est localisé à la région régulatrice localisée au 3' du locus IgH (3'RR) et que **(B)** dans les cellules B activées, les cohésines sont activement recrutées au niveau de Sμ-Cμ et cette interaction dynamique peut réguler les interactions à longue distance entre les régions S et la région 3'RR; par contre, des expériences additionnelles vont clarifier ce mécanisme. De plus, **(C)** les cohésines sont impliquées dans la réparation des dommages de l'ADN provoqués par l'activité d'AID, qui grâce à la voie classique du NHEJ, résultant en des microhomologies courtes au niveau des régions S. **(D)** L'identification de Spt6 comme facteur spécifique de la CI et notre contribution à la caractérisation du complexe PAF montrent que AID est présente au niveau des régions S en complexe avec l'ARN polymérase II, les facteurs d'élongation de la transcription Spt5 et Spt6 et le complexe PAF; notre résultats suggèrent que la transcription et les modifications de la chromatine peuvent réguler l'activité d'AID comme désaminase.

## References

- Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P., and Lis, J.T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* **420**, 837-841.
- Barreto, V., Reina-San-Martin, B., Ramiro, A.R., McBride, K.M., and Nussenzweig, M.C. (2003). C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* **12**, 501-508.
- 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.
- Chaudhuri, J., Khuong, C., and Alt, F.W. (2004). Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* **430**, 992-998.
- Coticello, S.G., Ganesh, K., Xue, K., Lu, M., Rada, C., and Neuberger, M.S. (2008). Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNBL1. *Mol Cell* **31**, 474-484.
- De Piccoli, G., Cortes-Ledesma, F., Ira, G., Torres-Rosell, J., Uhle, S., Farmer, S., Hwang, J.Y., Machin, F., Ceschia, A., McAleenan, A., *et al.* (2006). Smc5-Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination. *Nat Cell Biol* **8**, 1032-1034.
- Degner, S.C., Verma-Gaur, J., Wong, T.P., Bossen, C., Iverson, G.M., Torkamani, A., Vettermann, C., Lin, Y.C., Ju, Z., Schulz, D., *et al.* (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 9566-9571.
- Di Noia, J.M., and Neuberger, M.S. (2007). Molecular Mechanisms of Antibody Somatic Hypermutation. *Annu Rev Biochem* **76**, 1-22.
- Doi, T., Kato, L., Ito, S., Shinkura, R., Wei, M., Nagaoka, H., Wang, J., and Honjo, T. (2009). The C-terminal region of activation-induced cytidine deaminase is responsible for a recombination function other than DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2758-2763.
- Dorsett, D. (2009). Cohesin, gene expression and development: lessons from *Drosophila*. *Chromosome Res* **17**, 185-200.
- Durandy, A., Taubenheim, N., Peron, S., and Fischer, A. (2007). Pathophysiology of B-cell intrinsic immunoglobulin class switch recombination deficiencies. *Advances in immunology* **94**, 275-306.
- Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D.K., Aida, M., Inukai, N., Narita, T., Yamada, T., Furuya, A., *et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Molecular and cellular biology* **24**, 3324-3336.

- Guo, C., Yoon, H.S., Franklin, A., Jain, S., Ebert, A., Cheng, H.L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., *et al.* (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424-430.
- Harvey, S.H., Sheedy, D.M., Cuddihy, A.R., and O'Connell, M.J. (2004). Coordination of DNA damage responses via the Smc5/Smc6 complex. *Molecular and cellular biology* **24**, 662-674.
- Imai, K., Catalan, N., Plebani, A., Marodi, L., Sanal, O., Kumaki, S., Nagendran, V., Wood, P., Glastre, C., Sarrot-Reynauld, F., *et al.* (2003). Hyper-IgM syndrome type 4 with a B lymphocyte-intrinsic selective deficiency in Ig class-switch recombination. *J Clin Invest* **112**, 136-142.
- Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1975-1980.
- Jaehning, J.A. (2010). The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta* **1799**, 379-388.
- Jeevan-Raj, B.P., Robert, I., Heyer, V., Page, A., Wang, J.H., Cammas, F., Alt, F.W., Losson, R., and Reina-San-Martin, B. (2011). Epigenetic tethering of AID to the donor switch region during immunoglobulin class switch recombination. *The Journal of experimental medicine*.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., *et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435.
- Kracker, S., Gardes, P., Mazerolles, F., and Durandy, A. (2010). Immunoglobulin class switch recombination deficiencies. *Clin Immunol* **135**, 193-203.
- Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J.F. (2002). RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Molecular and cellular biology* **22**, 6979-6992.
- Losada, A., and Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev* **19**, 1269-1287.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553-563.
- Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of biological chemistry* **274**, 18470-18476.
- Nakamura, M., Kondo, S., Sugai, M., Nazarea, M., Imamura, S., and Honjo, T. (1996). High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int Immunol* **8**, 193-201.
- Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. *Annu Rev Genet* **43**, 525-558.

- Okazaki, I.M., Kotani, A., and Honjo, T. (2007). Role of AID in tumorigenesis. *Advances in immunology* **94**, 245-273.
- Okazaki, I.M., Okawa, K., Kobayashi, M., Yoshikawa, K., Kawamoto, S., Nagaoka, H., Shinkura, R., Kitawaki, Y., Taniguchi, H., Natsume, T., *et al.* (2011). Histone chaperone Spt6 is required for class switch recombination but not somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7920-7925.
- Orthwein, A., Patenaude, A.M., Affar el, B., Lamarre, A., Young, J.C., and Di Noia, J.M. (2010). Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *The Journal of experimental medicine* **207**, 2751-2765.
- Patenaude, A.M., Orthwein, A., Hu, Y., Campo, V.A., Kavli, B., Buschiazzo, A., and Di Noia, J.M. (2009). Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* **16**, 517-527.
- Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina-San-Martin, B., Barreto, V., *et al.* (2010). Activation-Induced Cytidine Deaminase Targets DNA at Sites of RNA Polymerase II Stalling by Interaction with Spt5. *Cell* **143**, 122-133.
- Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99-103.
- Potts, P.R., Porteus, M.H., and Yu, H. (2006). Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *The EMBO journal* **25**, 3377-3388.
- Ramiro, A., Reina-San-Martin, B., McBride, K., Jankovic, M., Barreto, V., Nussenzweig, A., and Nussenzweig, M.C. (2007). The role of activation-induced deaminase in antibody diversification and chromosome translocations. *Advances in immunology* **94**, 75-107.
- Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, 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.
- Seitan, V.C., Hao, B., Tachibana-Konwalski, K., Lavagnoli, T., Mira-Bontenbal, H., Brown, K.E., Teng, G., Carroll, T., Terry, A., Horan, K., *et al.* (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* **476**, 467-471.
- Shinkura, R., Ito, S., Begum, N.A., Nagaoka, H., Muramatsu, M., Kinoshita, K., Sakakibara, Y., Hijikata, H., and Honjo, T. (2004). Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol* **5**, 707-712.
- Stanlie, A., Aida, M., Muramatsu, M., Honjo, T., and Begum, N.A. (2010). Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 22190-22195.

- Stanlie, A., Begum, N.A., Akiyama, H., and Honjo, T. (2012). The DSIF subunits Spt4 and Spt5 have distinct roles at various phases of immunoglobulin class switch recombination. *PLoS Genet* 8, e1002675.
- Stavnezer-Nordgren, J., and Sirlin, S. (1986). Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *The EMBO journal* 5, 95-102.
- Ta, V.T., Nagaoka, H., Catalan, N., Durandy, A., Fischer, A., Imai, K., Nonoyama, S., Tashiro, J., Ikegawa, M., Ito, S., *et al.* (2003). AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat Immunol* 4, 843-848.
- Vuong, B.Q., Lee, M., Kabir, S., Irimia, C., Macchiarulo, S., McKnight, G.S., and Chaudhuri, J. (2009). Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat Immunol* 10, 420-426.
- Willmann, K.L., Milosevic, S., Pauklin, S., Schmitz, K.M., Rangam, G., Simon, M.T., Maslen, S., Skehel, M., Robert, I., Heyer, V., *et al.* (2012). A role for the RNA pol II-associated PAF complex in AID-induced immune diversification. *The Journal of experimental medicine* 209, 2099-2111.
- Xu, Z., Zan, H., Pone, E.J., Mai, T., and Casali, P. (2012). Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat Rev Immunol* 12, 517-531.

## Ebe SCHIAVO

### Molecular mechanisms controlling immunoglobulin class switch recombination

#### **Résumé français**

Lors des réponses immunitaires, le répertoire des lymphocytes B est diversifié par l'hypermutation somatique (HMS) et la commutation isotypique (CI), qui dépendent d'« activation-induced cytidine deaminase » (AID), qui introduit des lésions dans les gènes Ig. Une déficience d'AID cause un défaut d'HMS et CI; par contre, une délétion de son domaine C-terminal cause un défaut spécifique de la CI, suggère que ce domaine d'AID interagit avec des facteurs spécifiques de la CI. Pour identifier ces facteurs nous avons étudié une immunodéficiência présentant un défaut spécifique de la CI qui n'est pas lié à la carence d'AID ni à un défaut d'HMS. En effet, pas de cassure d'ADN ont été détectées au niveau des gènes IgH en suggérant qu'AID n'est pas proprement ciblé dans ces loci. Nous avons identifié des candidats : Spt6, le complexe des cohésines et le complexe Smc5/6. Dans les cellules B activées, AID interagit avec Spt6, Spt5, l'ARN polymérase II et le complexe PAF. Par contre, les cohésines peuvent réguler la structure du locus IgH pendant la CI et ainsi la voie de réparation des cassures de l'ADN générées pendant la CI. Ces résultats contribuent à une meilleure compréhension des étapes de la CI.

#### **Summary**

During immune responses, B cell repertoire is diversified through somatic hypermutation (SHM) and class switch recombination (CSR). SHM and CSR require activation-induced cytidine deaminase (AID), which induces DNA damage. While AID deficiency abrogates SHM and CSR, C-terminal truncations impair CSR without affecting SHM and it has been proposed that the C-terminal domain of AID associates with CSR-specific factor(s). In order to identify these factors we studied a human CSR-specific immunodeficiency, characterized by normal SHM and AID expression. B cells from these patients do not display DSBs at S regions, suggesting that they might lack an AID-binding factor(s) required to target AID to S regions during CSR. Through a multi-approach strategy we identified candidate factors, including Spt6, the cohesin complex and the Smc5/6 complex. We show that, in B cells poised to undergo CSR, AID is in a complex with Spt6, Spt5, the RNA polymerase II and the PAF complex while cohesins might regulate the 3D structure of the IgH locus and the pathway of DSBs repair at the Ig S regions. Our work thus contributes to a better understanding of the CSR reaction.