Epigenetic Mechanisms in two primary immunodeficiencies: Hyper-IgM Syndrome and Common Variable Immunodeficiency

Virginia Carolina Rodríguez Cortez

EPIGENETIC MECHANISMS IN TWO PRIMARY IMMUNODEFICIENCIES:

HYPER-IgM SYNDROME AND COMMON VARIABLE IMMUNODEFICIENCY

Tesis Doctoral

Virginia Carolina Rodríguez Cortez

EPIGENETIC MECHANISMS IN TWO PRIMARY IMMUNODEFICIENCIES:

HYPER-IgM SYNDROME AND COMMON VARIABLE IMMUNODEFICIENCY

Memoria presentada por Virginia Carolina Rodríguez Cortez para optar al grado de Doctor en Biomedicina por la Universidad de Barcelona

UNIVERSIDAD DE BARCELONA – FACULTAD DE MEDICINA

PROGRAMA DE DOCTORADO EN BIOMEDICINA

Este trabajo ha sido realizado en el Grupo de Cromatina y Enfermedad
del Programa de Epigenética y Biología del Cáncer (PEBC)
del Institut d' Investigaciò Biomédica de Bellvitge (IDIBELL)

Dr. Esteban Ballestar Tarín
Director

Virginia Carolina Rodríguez Cortez
Doctoranda

Todo esfuerzo que va acompañado de sacrificio, perseverancia, disciplina y dedicación tiene una recompensa. Puede que no llegue cuando tú quieras, ni sea lo que esperas, pero sí, definitivamente tiene su recompensa...

AGRADECIMIENTOS

Hace tanto tiempo que deseo estar en ésta etapa, que no me creo que ahora mismo estoy aquí sentada haciendo lo último que dejé por hacer, los agradecimientos. Para mí algo muy especial, porque aprovechas de decir a todas las personas que han formado parte de éste recorrido tan largo que es la tesis, lo importantes que son y han sido para ti y lo mucho que les aprecias.

Pues comienzo contigo Esteban, estoy muy agradecida de que me hayas dado la oportunidad de formar parte de tu grupo todos estos años, y de apostar por proyectos que no han sido ideales, ni han dado sus frutos tan rápido como se espera. Pero más que eso, te doy infinitas gracias por tu comprensión y tu flexibilidad en los momentos en que más los necesitaba y tu interés por que estuviera bien. No olvidas que somos personas y que tenemos momentos en los que por diferentes razones no damos nuestro 100%, y creo que es una cualidad importante que todos los jefes deben tener. Eso sí, a veces puedes ser muy duro (muy duro!) pero bueno, ahí nos toca a nosotros tener paciencia. Yo, particularmente me llevo muchas experiencias divertidas contigo y me alegro mucho de que hayas sido mi jefe todos esto años.

Javi, para mi has sido un verdadero Maestro, sí, con M mayúscula. Ahora mismo sería incapaz de enumerar todas las cosas que he aprendido de ti. Tienes un don para enseñar! consérvalo siempre porque es un don maravilloso. De todos los "Javi tienes un momentín?" siempre, siempre, salía algo productivo. Ése experimento que creía que había sido una total pérdida de tiempo, siempre se convertía en algo útil. Gracias a ésa motivación, seguí adelante en muchos momentos en que ya estaba decidida a tirar la toalla. Así que miles de gracias para ti!... Henar Henar, mi super compi de ordenador, gracias por todos los momentos que compartimos, eres muy especial para mí, me divertí mucho enseñándote vídeos de Chino y Nacho y de animalitos, compartiendo nuestro día a día y las jornadas de fin de semana poniendo a punto anticuerpos imposibles! Espero que a pesar de la distancia sigamos cosechando la linda amistad que hemos tenido hasta ahora... Lauri! mi queridísima Lauri, mentora de maternidad, compañera y amiga especial. Gracias por todos tus consejos, por compartir tantas cosas conmigo y por brindarme tu amistad en todo momento. Eres una chica super trabajadora y excelente persona, así que solo puedo augurar cosas buenas para ti y por su puesto, espero compartirlas contigo... Roser, mi vecina de poyata, muchísimas gracias también por los momentos amenos de cháchara durante los experimentos, por estar siempre pendiente y ofrecer ayuda en los momentos complicados, y también gracias por las chácharas de los momentos doctorales

difíciles, porque te das cuenta de que no estás sólo y que es igual de difícil para los demás! te irá muy bien, eres muy trabajadora y muy buena...Loren, gracias por todos los momentos de risas con tus ocurrencias y habilidades de negociación, me divertí mucho!... Biola, Marina, también disfrute mucho de su compañía y aprendí mucho de ustedes, las recuerdo siempre con mucho cariño... Nuestras queridas vecinas, en su momento conocidas como "las apéndices", Lidia, Bruna y Olgui, gracias porque a pesar de que éramos dos grupos diferentes, nos sentíamos como si fuéramos uno. Sobre todo por aquellos intercambios de material (que solían ser unidireccionales) que nos hacían sentir como si fuéramos un solo laboratorio... Lidia, gracias por todos los momentos divertidos que compartimos, fueron geniales!... Bruna, muchas gracias por tu amistad y por estar allí también en momentos difíciles. Eres para mí un ejemplo de tenacidad y de fuerza!... Olgui, mi querida Olgui, que ahora es Olgui, Jordi y Arnau!, qué te puedo decir, estoy encantada de haber compartido con ustedes toda la experiencia de la maternidad. Hemos compartido muchos momentos divertidos y agradables y espero de corazón que sigamos compartiendo muchos momentos más porque los apreciamos mucho! (Julio, Lucía y yo)...

Me llevo con especial recuerdo éste PEBC7, las comidas para hablar de todo y de nada, momentos super super divertidos, cumpleaños, fiestas de disfraces, montaña, burritos, baby shower, nacimientos... oh! tantas cosas!, dolores de barriga y lágrimas de tantas risas luego de analizar los beneficios de comer piña, o los curiosos efectos de la primavera y otras cosas que no puedo contar por aquí! han sido y serán momentos muy especiales para mí. Siempre me sentí como en casa... De ésta experiencia forman parte también por su puesto, las allegadas del PEBC7, Vicky, Laura Barberá y Palomita. Chicas, son geniales! les tengo un cariño muy muy especial. Vicky gracias por todos los asesoramientos tesísticos, me ayudaste mucho! muchas gracias también por los momentos que hemos compartido, han sido muy agradables!... Laura, aquí en casa te tenemos especial cariño también, lo pasamos genial los días de patines y todos los otros días que hemos compartido, esperamos que sean muchos más!... Palomita, Palomita, bah! eres una chica genial y super especial, siempre con una sonrisa, siempre con buena vibra, contagiando alegría. Gracias y mil gracias porque tu ayuda ha sido determinante también. He aprendido muchísimas cosas de ti y muchas cosas que he podido hacer durante mi tesis has sido gracias a ti! Deseo de todo corazón que te vaya excelente en cualquier cosa que decidas hacer (que puede ser cualquier cosa, literalmente!). Estoy muy agradecida contigo y con toda la gente del PEBC4, son unos excelentes compañeros! Álex, Anna,

Laia, Helena, Irene, Nuria, todos, todos, siempre dispuestos a compartir su experiencia y a echar una mano cuando sea. Mil gracias a todos!

A los nuevos miembros del PEBC7 también les agradezco mucho! Damiana, Carlos, Antonio, Lluis, Natalia, Alba, Francesc muchas gracias. Lluis, mi nuevo compañero de ordenador, gracias por ser tan listo siempre y ayudarme con mis rollos con la tecnología, siempre una solución a cualquier inconveniente informático! (darle a la tecla de escape y esas cosas)... Damiana, gracias por compartir tus anécdotas con Olivia, son siempre muy divertidas... Carlos, gracias por tu trabajo, seguro te irá muy bien... Nono, gracias por tu excelente disponibilidad y buena vibra con mis ChIPs infernales! (obvia ésta parte Francesc!)... Gracias de antemano a Francesc, porque ha heredado las tareas pendientes del proyecto AID. Aprenderás un montón, resolviendo problemas es cuando más aprendes!... Alba, eres una chica muy simpática, espero que te vaya muy bien en la tesis y que sigamos en contacto! que tenemos un ensayo pendiente eh? no se me olvida... Natalia, muchas muchas gracias, en éste poco tiempo que nos hemos ido conociendo has sido una excelente compañera! espero que te vaya muy bien en la tesis y que sigamos en contacto! hay unos sushis y varias cosas pendiente por allí. Para ti y para todos, estoy disponible para cualquier cosa que necesiten! los aprecio mucho!!!

A otras personas del PEBC también les estoy muy agradecida, Fer, Catia, siempre dispuestos a echar una mano... Carmen de microscopía, un ejemplo de fuerza y alegría!!!, todos los compañeros de pasillo que con 2 minutos de conversación y una sonrisa también han formado parte de éste recorrido... Mil Gracias!

Ya salgo del PEBC para agradecer a otras personas también, porque aunque no sean del ambiente laboral, su ayuda y su compañia han sido esenciales para mi. Mi querida amiga Patricia, una persona de referencia para mí en todos los sentidos, tus consejos, tu ayuda y tu apoyo para mi han sido vitales. Da igual dónde estemos y si pasamos muchos días sin hablar, la amistad que tenemos va mucho más allá de todo eso... He aprendido muchíiiiiisimo de ti, te quiero mucho y espero que nuestra amistad siga creciendo. Aprovecho para dejar plasmado aquí mi más sincero agradecimiento a tus papás, el Sr. Antonio y la Sra. Rosalía, sin su ayuda hubiese sido imposible venir aquí... eso no lo olvidaré nunca... Gracias por hacerme sentir como en casa, son muy especiales para mí!

Al team de Venezuela en Barcelona, María Cecilia, Manuel (los famosos tíos guays), gracias por tanto cariño para mi Lucía y por su apoyo en todo momento, los queremos mucho!. Arnoldo, Irma, Luisda, Sofi, Luis, Yenni, Ainhoa, Yaiza, Carlos,

Marc, son todos nuestra familia aquí, nuestro pedazo de Venezuela en España! esperamos seguir compartiendo momentos especiales y brindándonos apoyo... los quiero mucho!

Mi familia política, mis suegros, mis cuñados y sobrinos, todos siempre pendientes de nosotros, brindándonos amor, apoyo para cualquier cosa que necesitemos. Sra. Carmen, un especial agradecimiento para usted, siempre disponible para ayudarnos en todo, siempre atenta y buscando soluciones a cualquier cosa que se nos presente, la quiero mucho. Los quiero mucho a todos!

Mi familia, mi Papá, mi mamá, mi hermana, mis pilares, mi motivación. Lo que soy ahora es gracias a ustedes, gracias por apoyarme siempre, por estar pendientes de mi, por quererme tanto. No ha pasado un día, ni un solo día en que no los extrañe y desee estar cerca de ustedes. Uno de mis deseos más grandes es poder estar para ustedes y poder llegar a ser ése pilar que ustedes han sido para mi, los amo!.

Julio, mi kiki (si, kiki, tiene su historia...). Nada de lo que pueda escribir aquí llegará a expresarte todo mi agradecimiento. Han sido tantos años... Tenemos toda una vida juntos. Gracias por todo tu cariño, tu amor, tu apoyo, tu PACIENCIA, tu confianza. Has sacrificado tantas tantas cosas por mi, que lo que más deseo es hacerte feliz y pasar toooda mi vida contigo. Gracias por aguantar tanto y sobre todo por no tirar la toalla cuando te encontrabas en los momentos más dificiles. TE AMO con todo mi corazón y espero que sigamos juntos compartiendo la vida con nuestros retoñitos...

Mi Lucía... mi luz, mi inspiración, ahora mi centro y mi mayor motivo para dar lo mejor de mí. Llegaste en un momento clave de mi vida para enseñarme a priorizar, algo que necesitaba muchísimo... Sólo espero darte todo el amor, cariño y paciencia para hacerte una mujer de bien.. Te amo! y tú, mi Noelia (nombre provisional), que aún estás en mi vientre... ya te estoy esperando para igual que a Lucía, darte lo mejor de mí...Te amo también!, perdoname por las sesiones intensas con la tesis y por el estrés de mami que seguro has sentido en algún momento! ya te mimaré mucho para compensarte!

Pues termino aquí, me llevo excelentes experiencias, compañeros y amigos y por su puesto, lo que más me ha enseñado la tesis en todo éste tiempo: PACIENCIA y PERSEVERANCIA... creo que son un par de dones vitales para salir airosos de este largo y duro camino... con clínex en mano y al mejor estilo de Gustavo Cerati en concierto, me despido con un: GRACIAS....TOTALES!!!!!

INDEX

ABBREVIATIONS

5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymethyluracil
AID	Activation-induced Cytidine Deaminase
AUMA	Amplification of UnMethylated Alu's
BCL-6	B-cell lymphoma 6
BCR	B cell receptor
BER	Base excision repair
BLNK	B-cell linker
BTK	Bruton's tyrosine kinase
CGI	CpG islands
CLP	Common lymphoid progenitor
C-NHEJ	Non-homologous end joining pathway
CSR	Class Switch Recombination
CVID	Common Variable Immunodeficiency
DDR	DNA damage response
dn	double-null
DNMT	DNA methyltransferase
DSBs	Double strand breaks
EBF1	Early B-cell factor 1
EXO1	Exonuclease 1
GANP	GC associated nuclear protein
GC	Germinal center
GO	Gene Ontology
HA	Haemagglutinin
HBV	Hepatitis B virus
HC	Heavy chain
HDAC	Histone deacetylase
HIGM	Hyper-IgM Syndrome
HIV	Human immunodeficiency virus
HMTase	Histone methyltransferase
HP1	Heterochromatin protein 1
ICF	Centromeric instability and facial anomalies
IGH	Immunoglobulin heavy chain locus
iPS	Induced pluripotent stem cells
KAP-1	Krüppel-associated protein 1

KDM	Lysine demethylase
KMT	Lysine methyltransferase
LC	Light chains
LMB	Leptomycin B
LMPP	Lymphoid-primed multipotent progenitor
MMR	Mismatch repair pathway
MPP	Multipotent Progenitor
NES	Nuclear export signal
NK	Natural Killer
NLS	Nuclear localization signal
PAD	Primary Antibody deficiency
PBMCs	Peripheral blood mononuclear cells
PCH	Pericentromeric heterochromatin
PEI	Polyethylemine
PGCs	Primordial germ cells
PID	Primary Immunodeficiency
PKA	Protein kinase A
PKC	Protein kinase C
Pol η	DNA polymerase η
pre-BCR	pre-B cell receptor
PTBP2	Polypyrimidine-tract binding protein-2
RAG	Recombination-activating gen 1/2
S	Switch repeat regions
SAM	S-adenosyl methionine
SCL	Surrogate light chains
SHM	Somatic Hypermutation
Spt6	Suppressor of Ty6
SSBs	Single strand breaks
ssDNA	Single strand DNA
TCR	T cell receptor
TDG	Thymine DNA glycosylase
TdT	Terminal deoxynucleotidyl Transferase
TET	Ten-eleven translocation
TF	Transcription factor
UTR	Untranslated Region
V(D)J	Variable, Diversity and Joining

WT	Wild type
XLA	X-linked Agammaglobulinemia
μGLT	μ Germline transcripts

INTRODUCTION

1. INTRODUCTION

1.1. B cell Development

The immune response is the result of a highly coordinated action of different cells and molecules that constitute the immune system. It provides protection against internal or external aggressors through two types of defense lines: the innate and the adaptive immunity. Innate immunity, mediated by macrophages, neutrophils, dendritic cells and natural killer lymphocytes; represents the first line of defense against microorganisms through rapid but nonspecific mechanisms. Meanwhile, adaptive immunity, mediated by T and B lymphocytes, confers an extraordinarily specific response, that can be cellular or humoral, with the additional ability to "remember" the recognized aggressors, thus offering stronger and more efficient responses after repetitive exposures to antigens (Abbas and Janeway, 2000).

B lymphocytes are considered the main mediators of the humoral response. By its capacity to produce antibodies with high affinity and specificity, B cells lead to the recognition and neutralization of virtually any potential pathogen by the immune system. These cells are developed through a multi-stage process that begins in primary lymphoid tissues (bone marrow and liver in the fetal life and bone marrow in the adult life), and ends with the functional maturation in secondary lymphoid organs, mainly constituted by lymph nodes and spleen (LeBien and Tedder, 2008).

The commitment to differentiation towards the B cell lineage is achieved through important changes in gene expression including the acquisition of different cell surface markers and the expression of specific proteins. These changes are tightly regulated, not only by successive networks of transcription factors that ensure the activation of lineage-specific genetic programs and the inactivation of non-required genes, but also by epigenetic mechanisms (Parra, 2009).

A central event during B cell differentiation is the assembly of the genes that encode the B cell receptor (BCR), a pivotal molecule in B cell biology. BCR is an immunoglobulin (Ig) receptor formed by the covalent association of two identical heavy chains (HC) and two identical light chains (LC), that additionally requires the heterodimer formed by Igα/Igβ for the signal transduction process (Casola et al., 2004).

HC and LC are composed by a variable region, which confers the diversity and specificity against different antigens, and a constant region related to the different effectors functions of the antibodies. V(D)J recombination, is the process of somatic

rearrangements that leads to the assembly of variable regions of the HC by combining different V (variable), D (diverse) and J (joining) gene segments. In the LC only V and J segments are recombined. The gene assembly of HC and LC and the expression of the BCR and its accessory molecules is achieved by highly ordered events throughout B cell differentiation.

B cells are the result of successive differentiation steps that start with hematopoietic stem cells (HSCs) in adult bone marrow. HSCs initially give rise, by asymmetric division, one stem cell and one multipotent progenitor (MPP). MPPs further differentiate to committed myeloid and lymphoid progenitors. The first lymphoid-committed progenitor is the lymphoid-primed multipotent progenitor (LMPP) followed by common lymphoid progenitor (CLP), which can potentially differentiate into B cells, T cells or Natural Killer (NK) cells (Traver and Akashi, 2004). In the bone marrow the commitment of CLPs towards the B cell lineage comprises the differentiation throughout the following stages: early B ($CD34^+$ $CD19^-$ $CD10^+$), pro-B ($CD34^+$ $CD19^+$ $CD10^+$), large pre-BI ($CD34^+$ $CD19^+$ $CD10^+$), large pre-BII ($CD34^-$ $CD19^+$ $CD10^+$), small pre-BII ($CD34^-$ $CD19^+$ $CD10^+$) and immature B cells (Blom and Spits, 2006).

The differentiation of LMPPs into CLPs is mainly promoted by the transcription factors Ikaros, and PU.1. Then, the commitment and differentiation of CLP towards the B cell lineage is initiated by the transcription factors E2A (encoded by *TCF3*) and early B-cell factor 1 (EBF1) which activate the expression of key B cell genes such as *CD79A*, *CD79B*, *CD179A*, *CD179B* (encoding Igα, Igβ and the surrogate light chains VpreB and λ5 respectively), and the transcription factors *PAX5* and *FOXO1* (György et al., 2012). The concerted action of these transcription factors, mainly PAX5 which is considered the master regulator of B cell differentiation, allows the establishment of the proB cell stage by activating the expression of *CD19* and *BLNK*, which encode proteins necessary for BCR signaling, as well as the expression of Recombination-activating gen (RAG) 1/2 and terminal deoxynucleotidyl transferase (TdT) (Cobaleda et al., 2007). These enzymes mediate the rearrangement of V(D)J segments, therefore they are considered the main regulators of V(D)J recombination. Thus, the expression of BCR components and V(D)J recombination allows the transition to the pre-B cell stage.

The pre-B cell stage is characterized by the occurrence of two main events, the pre-BCR expression and the rearrangement of variable region of LC. The pre-BCR is composed by the rearranged μHC, the surrogate light chains (SCL) VpreB and λ5, and the signal transducing subunits Igα/Igβ. Signaling through pre-BCR promotes a

proliferation phase and a maturation phase in which the LC of BCR, λ and κ, are rearranged (Reth and Nielsen, 2014).

The last stage of differentiation in the bone marrow is reached with expression of BCR by substituting the surrogate LC of the pre-BCR, for the recombined κ and λ LC. This generates immature B cells which enter to the periphery to continue the differentiation process in secondary lymphoid organs (Reth and Nielsen, 2014) (Figure I1).

CLP, pro-B and pre-B cells constitute the main differentiation stages during B cell development in the bone marrow. Nevertheless, differentiation is achieved by passing through several transitional stages, such as early B stage (between CLP and pro-B), and the transitional pre-B cell stages pre-BI, Large pre-BII and small pre-BII before reaching the immature B cell stage. During these transitional steps, B cells acquire the changes necessaries for the proper expression of the key proteins mentioned above.

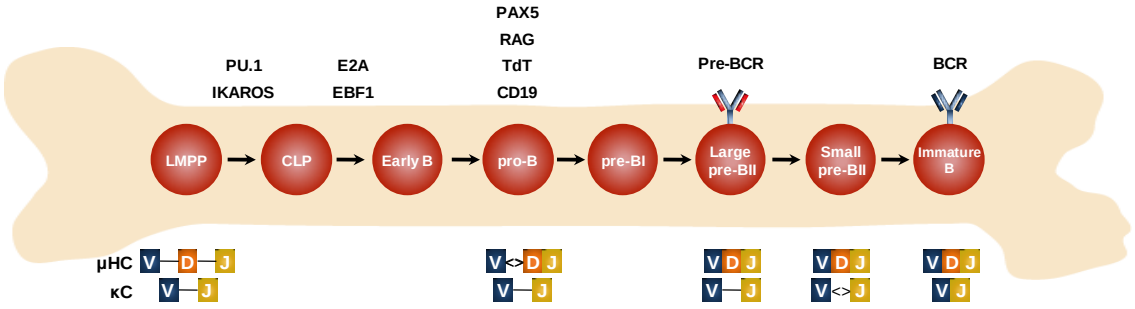


Figure I1. Early stages of human B cell development in bone marrow. LMPP indicates lymphoid-primed multipotent progenitor; CLP, common lymphoid progenitor; BCR, B cell receptor. Adapted from Blom & Spits, 2006 and Reth & Nielsen, 2014.

1.2. B cell Activation

Immature B cells expressing functional BCR migrate from the bone marrow to the secondary lymphoid organs (spleen, lymph nodes, tonsils, Peyer patches, and mucosal tissues). This migratory process is mainly mediated by adhesion molecules and cytokine gradients. Immature B cells acquire the cell surface markers IgD, CD21 and CD22, and pass through two transitional stages (T1 and T2) before becoming naïve mature B cells, marginal zone B cells or follicular B cells in the spleen (LeBien and Tedder, 2008).

Marginal zone B cells constitute a particular B cell subpopulation that shares properties of both, innate and adaptive immunity. These B cell subtype receives its

name from its localization at the marginal zone, the interface of circulation (red pulp) and lymphoid tissue (white pulp) of the spleen. The main function of marginal zone B cells, by their strategic localization, is the protection against blood-borne antigens. This function is mediated by the "innate side" and the "adaptive side" of these B cell subtype. The "innate side" of marginal zone B cells is conferred by their ability to recognize highly conserved microbial determinants through less specific BCR and Toll-like receptors (Pone et al., 2012). This recognition triggers the adaptive side of marginal zone B cells, which leads to the rapid generation of low-affinity antibodies by producing short-lived plasma cells. The quick responses of marginal zone B cells constitute a compensation mechanism for the time required by follicular B cells to generate high affinity antibodies. Thus, marginal zone B cells provide a first line of defense until the formation of plasma cells (Cerutti et al., 2013).

Moreover, follicular B cells are the precursors of the final effectors of B cell differentiation: plasma cells, which produce antibodies, and memory B cells, which provide long-term protection against repetitive exposure to antigens. The differentiation of follicular B cells occurs in specific microscopic structures of secondary lymphoid organs known as germinal centers (GC) where the GC reaction is produced. This reaction is triggered by the recognition of antigens by BCR and implies important changes in B cell expression, as well as the communication between B cells and other immune cell types (Figure I2).

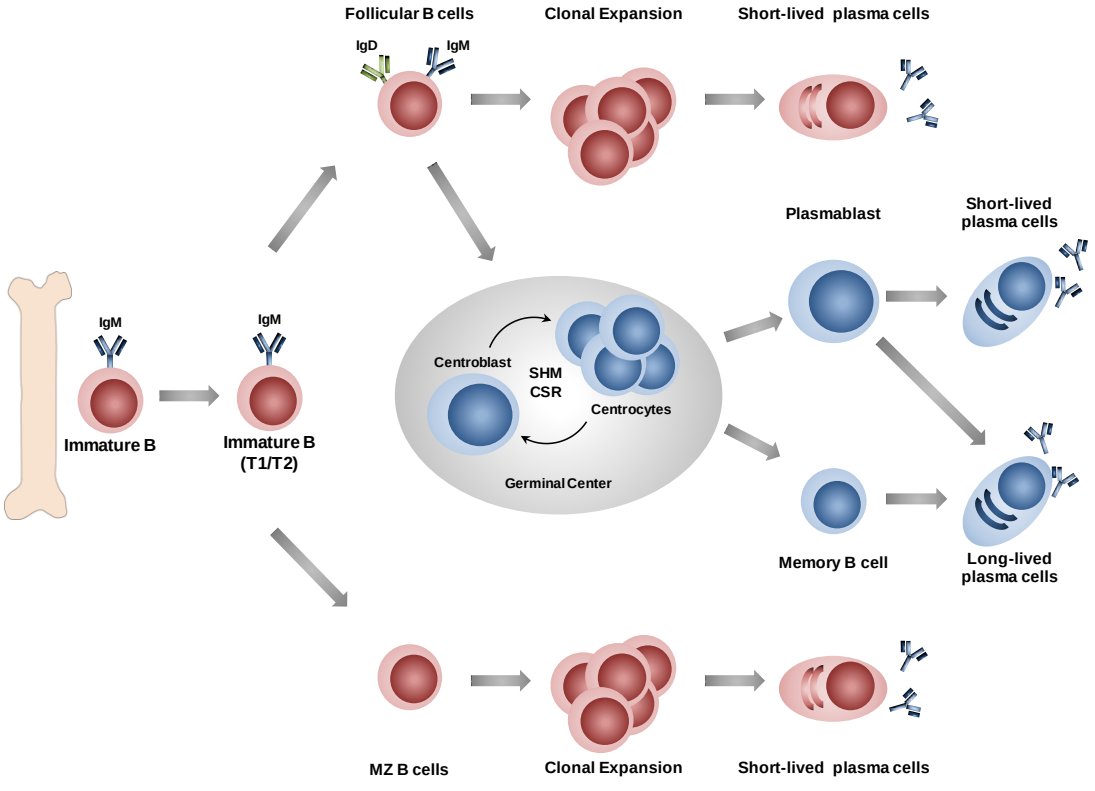


Figure I2. B cell development in the periphery. Immature B cells can differentiate into Follicular B cells or marginal zone B cells. Follicular B cells can undergo SHM and CSR at the Germinal Centers of secondary lymphoid organs to give rise plasma cells or memory B cells which provide the long-term protection of the humoral response. Additionally marginal zone B cells can give rise plasma cells that produce low-affinity antibodies, and follicular B cells can also give rise short-lived extrafollicular plasma cells that secrete antigen-specific germ line–encoded antibodies. Adapted from LeBien & Tedder, 2008.

The GC formation is mediated by the intense proliferation that B cells undergo after their activation. The highly proliferating cells are grouped in specific areas leading to the generation of two microscopic distinguishable zones, the dark zone and light zone. Dark zone concentrates the proliferating B cells, known as centroblasts while, light zone is composed by follicular dendritic cells, T cells, macrophages and centrocytes, which are smaller non-dividing B cells (Nieuwenhuis and Opstelten, 1984; Victora and Nussenzweig, 2012).

The main goal of the reaction at the GC is the generation of B cells able to produce antibodies with high affinity and specificity. This is achieved by two main processes: Class Switch Recombination (CSR) and Somatic Hypermutation (SHM). Both of them are mediated by the enzyme called Activation-Induced Cytidine Deaminase (AID), which is considered the master regulator of B cell activation.

SHM mediates the antibodies affinity maturation process in the dark zone of GC by the introduction of point mutations in the rearranged variable regions of immunoglobulin encoding genes. This is started when follicular B cells adapt their genetic program to initiate the formation of centroblasts. These cells are characterized by a gene expression profile that promotes high proliferation rates, apoptosis and the blocking of DNA damage response. The increased proliferation of centroblasts provides a wide repertoire of modified immunoglobulins, whereas increased apoptosis allows the rapid elimination of cells that produce defective antibodies. Meanwhile, the blocking of DNA damage impedes the cell cycle arrest that can be triggered by the genotoxic stress induced by high proliferation and by the DNA modifications produced during SHM (Klein and Dalla-Favera, 2008). The gene expression changes associated with this phase of GC reaction are driven by the master transcriptional regulator of centroblasts, the B-cell lymphoma 6 (BCL-6) transcriptional repressor (Shaffer et al., 2000).

In the light zone B cells cooperate with other immune cell types to finalize the differentiation process. The selection of improved antibodies is mediated by FDCs by challenging centroblasts that have undergone SHM with the antigens that triggered B cell activation, whereas macrophages eliminate B cells that produce less efficient

antibodies. Another important event in the light zone is the cooperation between B and T cells. B cells can present antigens to T cells and receive co-stimulatory signals from them to undergo CSR. Through this process B cells change the immunoglobulin isotype from IgM to IgG, IgA or IgE, generating antibodies with different effector functions. The B and T cell interaction is mediated by CD40 receptor, expressed by B cells, and its ligand CD154, expressed by T cells. Although CD40-CD154 is the main interaction that induces CSR, it has also been reported the participation of other molecules such as inducible T-cell co-stimulator (ICOS), transmembrane activator and calcium-modulating cyclophilin-ligand interactor (TACI) and B-cell-activating factor receptor (BAFFR) (Klein and Dalla-Favera, 2008) (Figure I3).

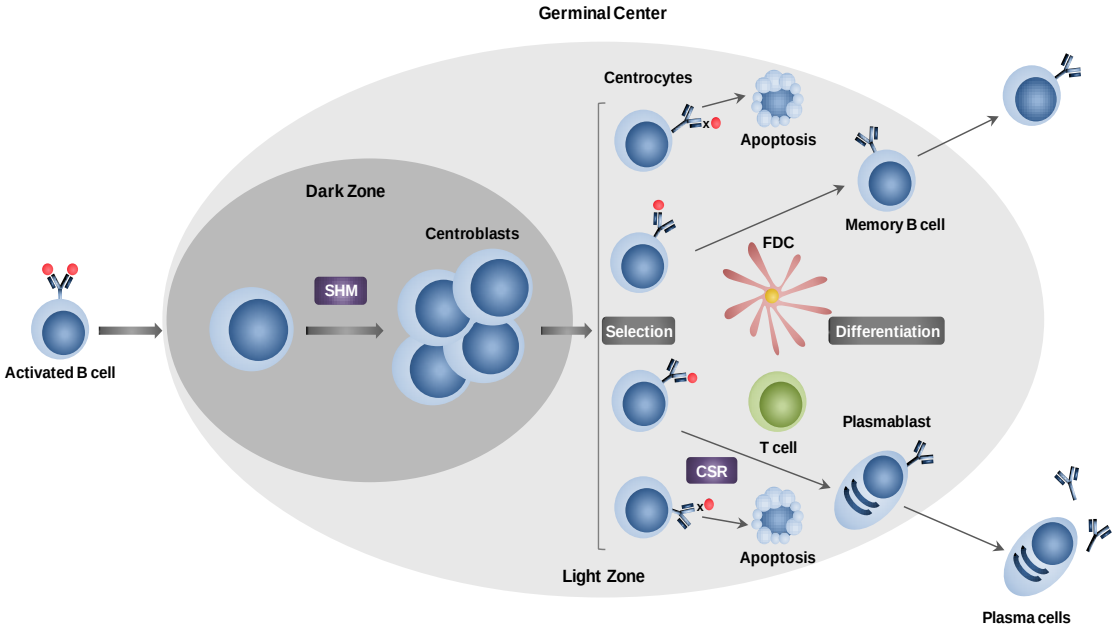


Figure I3. Germinal Center Reaction. Once mature B cells are activated by their encounter with antigens starts the proliferation and the formation of germinal centers. In the dark zone the highly proliferating centroblasts undergo SHM and express mutated receptors. Then in the light zone, mutated receptors are tested with help of FDC and T cells. Centrocytes that express low-affinity receptors are eliminated by apoptosis, while cells that express high affinity antibodies can undergo CSR and be differentiated into memory B cells or plasma cells to confer long-term protection against specific antigens. FDC indicates Follicular dendritic cells; SHM, somatic hypermutation; CSR, class switch recombination. Adapted from Klein & Dalla-Favera, 2008.

Initially it was thought that the GC reaction was a unidirectional process going from dark zone to light zone. However, it is now known that GC reaction is a highly dynamic process in which cells are in constant movement throughout dark zone and light zone and even class switched cells, near the completion of the differentiation process, can go back to the dark zone to have another round of SHM (Allen et al., 2007; Schwickert et al., 2007).

1.2.1. Class Switch Recombination

Class Switch Recombination (CSR) is the process through which B cells are able to change the expression of IgM to a different immunoglobulin isotype. CSR occurs in response to B cell activation and is a central event in the adaptive immunity as mentioned before. Regions that encode the constant heavy chains of the different immunoglobulin isotypes are organized in tandem in the immunoglobulin heavy chain locus (*IGH*) at chromosome 14. The coding regions are grouped in specific segments known as germline transcription units. Each unit includes a cytokine-inducible promoter followed by a non-coding exon (I-exon), a switch intronic region (S) and the C_H exon cluster (Matthews et al., 2014) (Figure I4).

CSR is a multistep process that is initiated by the recruitment of CSR machinery to the S regions for the generation of double-strand breaks (DSBs). This is followed by DNA deletional recombination of the S intervening regions and the activation of DNA damage response leading to DNA repair. As a result, the IgM coding region is excised and the default Cμ exons are exchanged for an alternative set of (Ch) exons, that could be from Cα, Cγ,or Cε, leading to the production of IgA, IgG or IgE respectively (Kracker and Durandy, 2011).

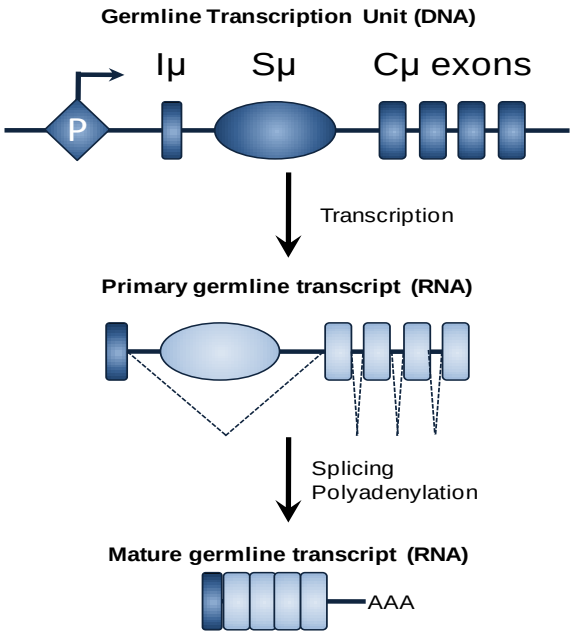


Figure I4. Germline transcription through S regions. The transcription unit includes a cytokine-inducible promoter (P), an intervening (I)-exon, S region, and Ch exons. The primary transcript is spliced and polyadenylated to generate a noncoding mature transcript. Adapted from Matthews et al., 2014.

Initiation of transcription is a crucial event for the recruitment of CSR machinery by two major roles. First, it has been described that transcription across the germline transcription units does not produce functional proteins, but instead allows the generation of DNA:RNA hybrids known as R-loops. These structures lead to the exposition of the non-transcribed single strand DNA (ssDNA), which is the substrate for AID action (Chaudhuri et al., 2003; Han et al., 2011). Although AID only acts on ssDNA, it has been demonstrated that AID is able to mutate both template and non-template strands, which can be explained by the interaction between AID and components of the RNA exosome complex that degrades DNA:RNA hybrids and exposes the template strand to AID deamination (Basu et al., 2011). The second role of R-loops generated during germline transcription, is to promote the stalling of RNA Polymerase II (PolII) that recruits AID through the splicing factor Spt5 and the recruitment of another AID interactors (Pavri et al., 2010).

Once AID is recruited to the S regions, it starts the generation of DSBs. AID deaminates cytosines and converts them into uracils generating dU:dG mismatches. These mismatches can be recognized and processed by different pathways resulting in either their repair, generation of mutations or DSBs. It has been reported that the main pathway in processing dU:dG mismatches during CSR is the Base excision repair (BER) pathway. In BER, AID-introduced uracils are removed by uracil-DNA-glycosylase generating abasic sites, which are subsequently recognized and nicked of by apurinic/apyrimidic endonucleases, mainly by APE1. Staggered DSBs are generated by this pathway when the single strand breaks (SSBs) produced are closely spaced in the opposite strands (Stavnezer et al., 2008).

It has also been reported that the mismatch repair pathway (MMR) plays a relevant role during CSR. MSH2 and MSH6, which are MMR proteins, can recognize dU:dG mismatches and recruit other effector proteins such as exonuclease 1 (EXO1), to produce SSBs in mismatch adjacent regions. Proximal SSBs produced by both, MMR and BER pathway, can generate DSBs, thus MMR supplements BER pathway in DSBs generation in S regions (Chaudhuri and Alt, 2004; Rada et al., 2004).

DNA damage response (DDR) is triggered once DSBs generated in S regions are detected by the DNA damage sensors NBS1 and ATM. It starts with the phosphorylation of histone H2AX (γ-H2AX) and the recruitment of 53BP1 protein, both of which has been recently implicated in the formation of the required synapsis of the intervening S regions (Manis et al., 2004; Reina-San-Martin et al., 2007). DNA repair of the generated DSBs is the final step of CSR. It is mainly mediated by the classical non-

homologous end joining pathway (C-NHEJ) and is complemented by the action of the alternative end-joining (A-EJ) and homologous recombination pathways (Boboila et al., 2010; Hasham et al., 2010; Yan et al., 2007).

CSR is the result of the orchestrated action of several enzymes and scaffold proteins. Expression of different immunoglobulin isotypes allows the adaptation of the effector immune mechanisms to provide a proper and efficient response against the different pathogens.

1.2.2. Somatic Hypermutation

SHM shares several mechanistic features with CSR, however it is a less characterized process. As well as CSR, SHM is initiated by transcription. In this case, the targeted transcribed region is the rearranged variable region of HC and LC at the Ig locus. The mutated region is 100–200 bp downstream of the transcription initiation site and extends for 1.5–2.0 kb (Lebecque and Gearhart, 1990).

AID is the key enzyme that initiates SHM. Its catalytic activity generates mutations at high rates (10^{-2}, 10^{-3} mutations per base pair per generation) in the WGCW (W = A/T) hotspot sequence motif. As in CSR, AID deaminates cytosines to generate dU:dG mismatches. The next step to generate mutations is the recruitment and action of DNA repair machinery. Proteins of this pathway also act during CSR mainly to produce SSBs and DSBs as mentioned above; however, during SHM the damage initiated by AID, rather than be repaired in an error-free manner (which is the canonical function of DNA repair machinery), it is transformed into a mutation (Peled et al., 2008).

The proposed mechanism to explain this paradox is that during SHM occurs a DNA repair process in which error-prone polymerases participate rather than the high-fidelity polymerases that normally avoid DNA mutations. It has been reported that DNA polymerase η (Pol η) and REV1 are mainly the low-fidelity polymerases involved in SHM (Casali et al., 2006).

After AID action, proteins of BER and MMR are recruited. Similarly to CSR, uracil-DNA-glycosylase excises the uracil produced by AID and generates an abasic site that is processed by APE1, which generate SSBs. Moreover, proteins of the MMR, MSH2-MSH6 can also recognize the uracils and recruit EXO1 to create a gap. After the participation of BER and MMR proteins, error-prone polymerases are recruited through mechanisms that remain to be clarified (Di Noia and Neuberger, 2007).

One of the proposed mechanisms is that REV1 and other error-prone polymerases could be recruited by the monoubiquitylated form of PCNA, which is a heterotrimer with relevant functions during replication that is monoubiquitylated when the replication fork is stalled by DNA lesions. As an alternative, but not excluding mechanism, it has been postulated that error-prone and error-free polymerases are differentially recruited during cell cycle (Langerak et al., 2009).

SHM and CSR are very particular processes in which the machinery that normally prevents and repairs DNA damage is recruited to facilitate the random introduction of mutations and the generation of SSBs or DSBs to recombine different immunoglobulin gene segments. There are similarities between SHM and CSR, even though the final outcome is different. In any case, the hallmark of both events, as mentioned above, is the expression and action of the same enzyme, AID (Figure I5).

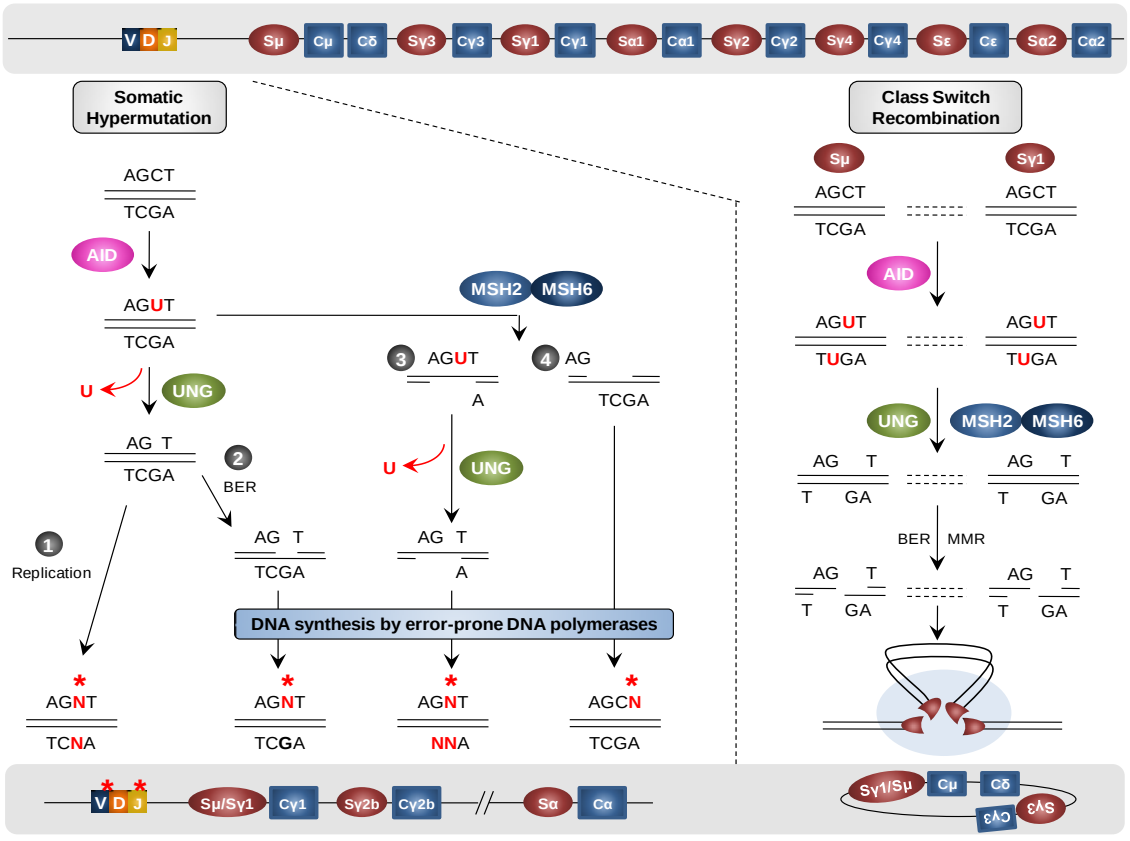


Figure I5. Models of Somatic Hypermutations and Class Switch Recombination. The upper gray square shows a non-mutated, non-switched, rearranged human Ig gene. The lower gray square show the Ig gene after SHM and CSR. During SHM (left part) AID deaminates cytosines to generate uracils that are recognized and processed by uracil-DNA-glycosylase. The generated abasic sites might be a template during replication and generate any mutation (1), or abasic sites can be processed by BER machinery which generates SSBs that are filled by recruitment of error prone polymerases (2). The MMR machinery can be also recruited and acts before (4) or after (3) uracil-DNA-glycosylase. MMR machinery produces gaps that are repaired by error-prone polymerases. During CSR (right part) AID targets S regions and deaminates cytosines to generate uracils. This uracils are processed by uracil-DNA-glycosylase and the action of BER and MMR machineries introduce SSBs and DSBs. In this case is exemplified the recombination between Sμ and Sγ1 regions. Adapted from Longerich, Basu, Alt, & Storb, 2006.

1.3. Activation Induced Cytidine Deaminase

Activation Induced Cytidine Deaminase (AID) is an enzyme that was first discovered in 1999 by Honjo and colleagues in an experiment with the murine B lymphoma clone, CH12F3-2, which undergoes robust CSR from IgM to IgA in response to IL-4, TGF-β, and CD40L. By using a cDNA substraction approach they compared the switch-induced and uninduced murine B lymphoma CH12F3-2 cells, and identified *AICDA* as a strongly upregulated gene during B cell activation (Muramatsu et al., 1999).

AICDA, the gene that encodes human AID, is localized in the chromosome 12p13 and extents about 11kb. It is composed by five exons that give rise a protein of 198 amino acids (Muto et al., 2000). Several splicing variants of AID have been reported, but their function remains controversial (van Maldegem et al., 2009; Wu et al., 2008)

AID belongs to the AID / APOBEC cytidine deaminase family of proteins. This family is composed by AID, APOBEC1, APOBEC2, the subgroup of APOBEC3 and APOBEC4 proteins. All of them are zinc-dependent deaminases and share the catalytic cytidine deaminase domain (Conticello, 2008). These enzymes have been related to different important roles by their catalytic activity. APOBEC1, the founder member of this family, has an important role in lipid metabolism by its RNA-editing activity. It is responsible of the apolipoprotein B (ApoB) mRNA editing, which leads to the generation of two different isoforms of ApoB with different roles in lipid transport (Chester et al., 2000). The rest of the members of the AID/APOBEC family are DNA deaminases. Interestingly, the APOBEC3 subgroup has been associated with the innate immunity against retroviruses (Harris et al., 2003). By deaminating and editing the nascent cDNAs, APOBEC3 members interfere with the reverse transcription of RNA genomes from human immunodeficiency virus (HIV) and hepatitis B virus (HBV) (Mangeat et al., 2003; Sheehy et al., 2002; Turelli et al., 2004). Meanwhile, it remains to be determined if APOBEC2 and APOBEC4 have specific and relevant functions (Bransteitter et al., 2009).

1.3.1. AID, the master regulator of antibody maturation

The experiments carried out by Honjo and colleagues mentioned above, suggested for the first time a possible role of AID in activated B cells. However, the critical role of AID during B cell activation was completely elucidated through the study of two different models: knockout mice for *Aid* and a group of patients with a specific subset of Hyper-IgM Syndrome (HIGM).

The *Aid* knockout mice revealed a complete defect in class switching with a phenotype characterized by increased serum levels of IgM (Hyper-IgM) and by the presence of enlarged GCs. Subsequent experiments with splenic B cells from these AID deficient mice showed the specific impairment of CSR and SHM after different stimuli, and suggested the involvement of the catalytic activity of AID in the genetic modifications required in both processes (Muramatsu et al., 2000).

At the same time, Revy and colleagues published the first evidences that linked alterations of human AID with an autosomal recessive form of HIGM. The analysis of a cohort of 18 patients with HIGM diagnosis revealed deleterious mutations scattered across the *AICDA* gene which generate inactive forms of AID. The phenotype observed in these patients was defective CSR, defective SHM and the presence of enlarged GCs with highly proliferating B cells, which resembled the phenotype observed in the AID knockout model (Revy et al., 2000). Thus, both models of AID deficiency, human and mice, demonstrated the absolute requirement of AID during B cell terminal differentiation and antibody affinity maturation.

Interestingly, transgenic mice with constitutive and ubiquitous expression of AID develop T cell lymphomas and lung adenocarcinomas as a result of introduction of mutations in non-Ig genes by AID. Curiously, these mice do not have alterations in B cells or develop B cell malignances, which suggest the presence of additional layers of regulation to control AID activity in B cells (Okazaki et al., 2003).

Cytosine deamination is the specific mechanism by which AID mediates CSR and SHM. This reaction proceeds by a direct nucleophilic attack at position 4 of the pyrimidine ring of cytosine by Zn^{+2} coordinated to AID (Conticello, 2008) (Figure I6).

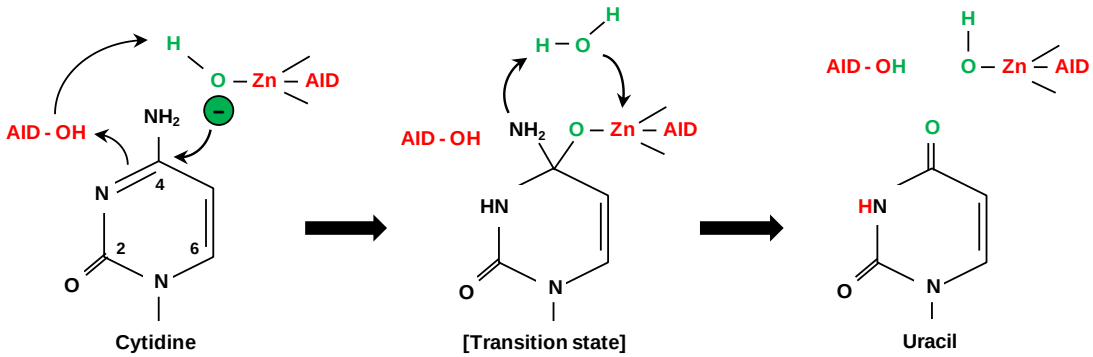


Figure I6. Cytidine deamination reaction. Mechanism based on a bacterial cytidine deaminase that is homologous to APOBEC1 and AID. Zinc ions (Zn^{2+}) coordinated with AID perform a direct nucleophilic attack at position 4 of the pyrimidine ring. Adapted from Chaudhuri & Alt, 2004.

The mistmatches generated by the deamination of cytosines by AID, can be detected by in an error-free manner by BER or MMR machinery. However, uracils that are not detected by DNA repair machinery can be replicated and generate transitions (from C:G to T:A). Another possibility is the processing of the mismatch through a mutagenic pathway that recruits error-prone polymerases, which can produce transitions or transversions at the deaminated site. This mutagenic pathway is the responsible of the DNA changes required during CSR and SHM, as explained above.

1.3.2. Functional Domains of AID

Although AID is a small protein (24kDa), its structure contains several differentiated functional domains. At its N-terminal region it has been described a putative bipartite nuclear localization signal (NLS) (position 8 to 26) with the typical motif KR - $X_{10–12}$ - K(KR)(KR), where X corresponds to any amino acid. It has been postulated that this type of sequences mediate the nuclear import through the importin α / β pathway (Freitas and Cunha, 2009; Ito et al., 2004; Kosugi et al., 2009). Near from the described classical NLS, it has been postulated a conformational nuclear localization signal. This is a non-consecutive motif of basic residues (residues19-RWAK-22,34-KRR-36, 199 50-RNKN-54) that provides a positively charged surface in the folded protein (Patenaude et al., 2009).

Moreover, the catalytic domain is localized between the amino acid residues 55 and 94 and possesses the conserved H(A/V)E-$X_{24–36}$-PCXXC motif present in the cydine deaminase family. Meanwhile, the APOBEC-like domain (position 119 to 183) was defined by alignment with APOBEC-1, as a homologous region in the APOBEC family with uncharacterized functions. And finally, at the C-terminal region of AID, the last 15 amino acid residues correspond to a nuclear export signal (NES) that mediates the AID exclusion from the nucleus by a exportin/CRM1 dependent mechanism (Barreto and Magor, 2011; Ito et al., 2004).

Additionally, other critical regions in the AID structure, based on the crystal structure of APOBEC2, have been proposed. A putative dimerization domain has been associated with the region between 46-53 amino acid position near from the catalytic domain, whereas, amino acids residues 111-113 and 168, have been associated with a putative tetrameric interface (Prochnow et al., 2007).

At the functional level, two regions have been described. In the N-terminal region, a segment required for SHM has been identified (residues 13-23) , whereas the last 26 amino acids have proved to be necessary for CSR. It has been proposed that

these two regions are critical for the association between AID and its regulatory partners (Barreto and Magor, 2011) (Figure I7).

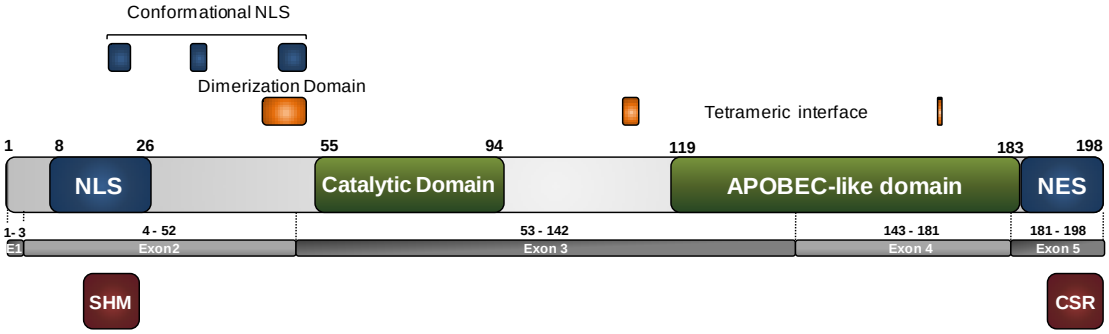


Figure I7. Functional domains of AID. In blue, AID regions linked with AID localization. In green, the domains shared by the AID/APOBEC deaminases members. In orange, putative regions for AID dimerization and tetramerization. In dark red, AID regions linked with the AID functions during antibodies affinity maturation. NLS, indicates nuclear localization signal; NES, nuclear export signal; SHM, somatic hypermutation; CSR, class switch recombination. Adapted from Barreto & Magor, 2011.

1.3.3. AID Expression

Initially, several evidences supported the notion that due to its role, AID expression was restricted to B cells from GC of the spleen and mucosal-associated lymphoid tissues (Peyer's patches, tonsils, lymph nodes). The first analysis of AID mRNA distribution showed undetectable levels in both, T cells and different splenocyte cell populations, while a dramatic increase was observed in B cells after in vitro activation (Muramatsu et al., 1999).

However, recent evidence showed that AID mRNA could be detected outside the B cell compartment under physiological and pathological conditions. As examples, AID expression has been detected in normal tissues from ovary, breast, heart, and lung, although the function of AID in these tissues remains to be elucidated. At the pathological level, AID expression has been reported in lymphomas, leukemias, and malignant epithelial cells from solid tumors. In the latter, the ectopic AID expression has been associated with the chronic inflammation triggered by different infectious agents such as Hepatitis C virus, *Helicobacter pylori*, Epstein Barr virus or HIV, and it has been suggested that ectopic AID expression in these contexts is mediated by aberrant NF-κB activation (Orthwein and Di Noia, 2012).

Moreover, although it remains controversial, it has been proposed that AID is expressed at low levels during lymphopoiesis and it has been suggested that its expression suppresses the development of autoreactivity (Kuraoka et al., 2011; Meyers

et al., 2011). These findings were challenged by the analysis of a mice model in which the expression of a GFP-tagged AID allowed the identification of B cells that actively expressed AID during the immune response. An induction of AID expression in GC B cells after B cell activation was reported as well as undetectable levels of AID in bone marrow–developing B cells, which argue against the potential role of AID in self-tolerance in these stages under physiological conditions (Crouch et al., 2007). More recently, AID expression in the embryonic developmental context has been reported, but these findings will be discussed later.

1.3.4. Regulation of AID expression and activity

As expected, the regulation of the expression of a mutator like AID, is complex, and not only its expression requires to be restricted to B cells, but also has to display a highly specific response to particular stimuli. In addition to the regulation of AID expression, several post translational mechanisms control the activity of AID, thus maintaining the genome integrity.

1.3.5. Transcriptional Regulation of AID

It has been described that there are 4 specific regulatory regions in the AICDA gene that are responsible for the regulation of AID expression, which contain binding sites for at least 19 transcription factors (both, activating and repressive factors). Region 1 is comprised by the *Aicda* promoter and the region immediately upstream. It contains a TATA-less promoter, and binding sites for the transcription factors Sp1, Sp3 and HoxC4. It also contains a binding site for NF-κB transcription factor which has been specifically associated with the response to the TNF-α and Toll-like receptor signaling triggered by viruses. It has also been reported an estrogen response element in region 1, but its function remains to be clarified (Nagaoka et al., 2010).

Region 2 is located in the first intron and is a critical regulatory region. It confers B lineage-cell specificity on *Aicda* expression by containing binding sites for E-proteins and Pax5 transcription factors, both of which are necessary for B cell development, as mentioned before. This region also contains binding sites for the ubiquitous silencers c-Myb and E2f, which are necessary for the transcriptional repression of AID expression (Tran et al., 2010).

Region 3 is located approximately 17 kb downstream of the exon 5. Although its precise role has not been determined, it has been demonstrated that its deletion reduces AID expression. Additionally, a binding site for the transcription factor BATF,

which controls AID and GLT expression has been identified. Finally, Region 4 is located approximately 8 kb upstream of the TSS, and has been associated with the AID induction by environmental stimuli. It contains binding sites for NF-κB, Stat6, Smad and C/EBP proteins (Tran et al., 2010) (Figure I8).

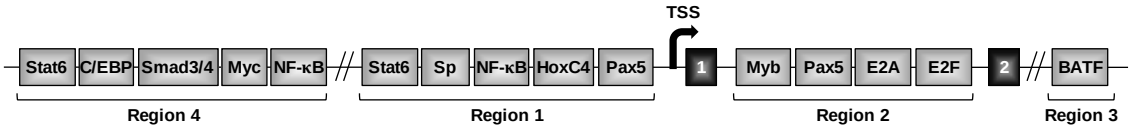


Figure I8. Regulatory regions at the *AICDA* promoter. Black boxes indicate AID exons. Adapted from Matthews et al., 2014.

The regulation of AID expression at the transcriptional level is finally achieved by a highly strict coordination among the activating factors, which are all those elements that respond to the cytokine stimulation during B cell activation and development; and the repressing factors, constituted by ubiquitous silencers that guard the genome integrity by silencing AID expression.

1.3.6. Post transcriptional Regulation of AID

After B cell activation, the increasing amount of AID mRNA is controlled by a second layer of regulation constituted by microRNAs. These small noncoding RNAs target mRNAs and inhibit their translation by altering their stability and translation efficiency. Two microRNAs have been mainly associated with AID regulation: miR-155 and miR-181b. Both of them target the 3'-Untranslated Region (UTR) of the AID encoding mRNA, causing a decrease in the mRNA and protein levels of AID (Stavnezer, 2011).

miR-155 is a lymphocyte-specific microRNA that has a similar expression pattern than AID during CSR. Disruption of the interaction between miR-155 and AID mRNA resulted in deregulation of AID expression that not only affects CSR and affinity maturation, but also lead to aberrant targeting of AID and the increase of chromosomal translocations between *IgH* and *c-myc* locus (Dorsett et al., 2008; Teng et al., 2008).

Moreover, miR-181b was identified in a screening as a microRNA whose expression is able to impair CSR. It exhibits a different expression pattern than miR-155, where the higher levels of miR-181b are observed in resting B cells and are decreased after B cell activation. This leads to the suggestion that miR-181b has a role in preventing premature AID activity (de Yébenes et al., 2008).

The reduction of AID levels by microRNAs has proved to be a critical regulation point to ensure the proper AID expression and function. Altered expression of these microRNAs is associated with lymphomagenesis that mainly result from chromosomal translocations caused by aberrant AID targeting.

1.3.7. AID subcellular localization

AID is a small protein that, because of its size, theoretically should diffuse passively into the nucleus. However, AID localization is mainly restricted to the cytoplasm, since the DNA mutator activity of AID demands a strict control of its nuclear entry. At present, there is evidence that supports the existence of three different mechanisms to regulate the nuclear amount of AID: active nuclear import, nuclear export, and cytoplasmic retention (Patenaude and Noia, 2010).

The nuclear import is active and is mediated by the interaction with importin-α. Although AID possesses a putative bipartite NLS, experiments with deletional mutations suggest that its nuclear entry is mediated by a conformational NLS. Once in the nucleus, the amount of AID is regulated by degradation and nuclear export. The nuclear export is mediated by the recognition of the NES by the CRM1 exportin. Additionally, the cytoplasmic retention mechanism provides another layer that regulates the AID localization. This mechanism is mediated by AID interaction with Hsp-90 chaperone, which protects cytoplasmic AID from being degraded and determines the availability of functional AID (Orthwein et al., 2010; Patenaude et al., 2009)

Together these mechanisms for AID compartmentalization are relevant for the regulation and the accessibility of AID to its DNA substrate, preventing off-target AID activity, and providing a cytoplasmic AID reservoir that leads to a fast nuclear translocation under appropriate stimuli (Xu et al., 2007).

1.3.8. Phosporylation of AID

Several amino acid residues of AID can be phosphorylated: Ser3, Thr27, Ser38, Thr140, Ser41, Ser43 and Tyr184. Phosphorylation at Ser38 and Thr140 has been associated with AID activation, while the phosphorylation at Ser3 and Thr27 has been linked to the inhibition of AID activity. Moreover, the role of phosphorylation at residues Tyr184, Ser41 and Ser43, during CSR and/or SHM remains to be determined (Demorest et al., 2011; Pham et al., 2008; Stavnezer, 2011).

Thr140, phosphorylated by Protein kinase C (PKC), has been linked to SHM. Experiments with mutants forms that are unable to become phosphorylated in this

residue, showed little defects on CSR but an impairment of SHM that was not related to decreased AID activity, which suggests a role of this phosphorylation in mediating key AID interactions during SHM (McBride et al., 2008)

It has been reported that phosphorylation at Thr27 impairs CSR without altering AID activity, whereas phosphorylation at Ser3 impairs the association of AID with the IgH Sμ region. The role of these phosphorylation events *in vivo* remains to be determined (Gazumyan et al., 2011; Pasqualucci et al., 2006)

The best characterized phosphorylation site for AID occurs at Ser38 (p38AID) and is carried out by protein kinase A (PKA). p38AID is detected in a small fraction of AID *in vivo*. Nevertheless AID Ser38 phosphorylation has proved to be critical for CSR and SHM. p38AID is enriched in the chromatin fraction of activated B cells and it is required for the interaction between AID and replication protein A (RPA), which is necessary for the targeting of AID to the IgH locus. In addition, p38AID has been related with the steps downstream DNA deamination by mediating the indirect interaction between AID and APE1, an endonuclease necessary to generate single strand breaks (SSBs) during CSR (Vuong et al., 2013). In fact, a positive feedback loop mediated by p38AID to amplify the generation of DSBs has been suggested.

Thus, phosphorylation at different AID residues, more than leading to a direct alteration of AID activity, constitutes a relevant regulatory mechanism that modulates CSR and SHM through the modification of the interactions established between AID and different proteins.

1.3.9. AID interactions

An additional layer of AID regulation is constituted by the proteins that target AID to the IgH locus and restrict the deamination activity mainly to the Switch and Variable regions for CSR and SHM respectively.

The DNA sequence motifs, RGYW and WA, together with their reverse complementary sequences, WRCY and TW (where R=A or G, Y=C or T, W=A or T), are preferential binding sites for AID. These motifs are considered hotspots for the mutations induced by AID. However, as these motifs are degenerated and are scattered across the entire genome, they cannot account for the specificity observed during CSR and SHM. In line with this observation, it has been found that AID binds transcribed regions across the entire genome of activated B cells, as determined by chromatin immunoprecipitation (ChIP) coupled with ultrasequencing (ChIP-seq)

experiments in mice (Yamane et al., 2011). Thus, the interaction between AID and several proteins is the main AID targeting mechanism that provides insights into how AID activity is targeted to specific Ig loci.

For instance, interaction between AID and the GC associated nuclear protein (GANP) has been linked to the targeting of AID to transcribed variable regions to promote SHM, although the underlying mechanism has not been described (Maeda et al., 2010). Moreover, 14-3-3 adaptors are proteins that are upregulated during B cell activation, bind to the AID hotspots, and also interact with AID, seemingly for the targeting of AID to the S regions for CSR (Xu et al., 2010).

The interaction between AID, PolII and Spt5, as mentioned above, is necessary for the recruitment of AID and explains the well characterized link between transcription and CSR and SHM. Furthermore, the splicing regulator, Polypyrimidine-tract binding protein-2 (PTBP2), has also been implicated in AID binding to S regions through its interaction with S region transcripts (Nowak et al., 2011).

Interestingly, chromatin conformation also plays a role in AID recruitment. Open chromatin states defined by activation histone marks such as trimethyl histone H3 lysine 4 (H3K4me3) and hyperacetylated H3K9 (H3K9Ac), are linked with CSR and its importance is highlighted by experiments in which the activity of the enzymes responsible of these histone marks are altered, leading to an impairment of CSR (Daniel et al., 2010; Wang et al., 2009).

Together, the coordination of different regulatory mechanisms ensure the proper expression and targeting of AID, which is important not only for proper antibody maturation and immune response, but also to guarantee the genome integrity during AID function. The main regulatory mechanisms of AID are summarized in the Figure I9.

1.3.10. AID roles beyond CSR and SHM in B cells

In recent years, new potential AID roles have been proposed since the detection of AID expression in non-B cell contexts and its association with the regulation of DNA methylation, an important epigenetic mechanism.

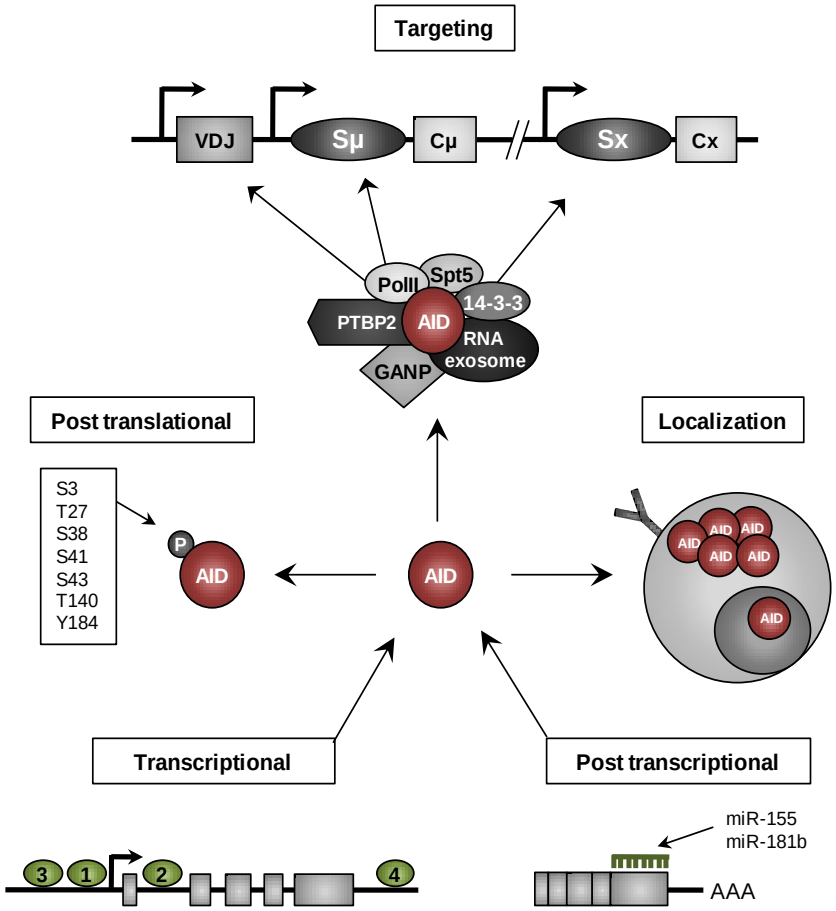


Figure I9. Regulation of AID expression and activity. Different transcription factors regulate AID expression. At the post-transcriptional level, specific micro-RNAs regulate the availability of AID mRNA. The generated AID protein is then controlled by phosphorylation and subcellular localization mediated by the localization signals contained in its structure and by specific protein interactions. Furthermore, protein interactions also control the AID targeting to V and S regions. Adapted from Matthews et al., 2014.

Development of primordial germ cells (PGCs) during embryogenesis, which are the precursors of spermatocytes and oocytes, requires rapid DNA demethylation to the erasure of paternal imprints (Sasaki and Matsui, 2008). AID is expressed in PGCs, and the observation that AID deficient-PGCs exhibit higher levels of global methylation than their WT counterparts, led to the suggestion that AID may participate in active DNA demethylation mechanisms (Popp et al., 2010).

Moreover, in vitro reprogramming of somatic cells to produce induced pluripotent stem cells (iPS), which is another model that requires rapid DNA demethylation, revealed that AID is necessary for promoter demethylation and expression of the critical pluripotency genes, *OCT4* and *NANOG* (Bhutani et al., 2010). Furthermore, the reprogramming of *Aid*-null differentiated cells to induced pluripotent stem cells showed failed stabilization of the pluripotent state due to hypermethylation of pluripotency associated genes (Kumar et al., 2013), which reinforces the role of AID in

the DNA demethylation required for cell reprogramming. In these systems, the proposed demethylation mechanism is the direct deamination of 5-methylcytosines (5mC) by AID to produce thymines which are later replaced by unmethylated cytosines after the action of BER machinery.

Another supporting evidence of the role of AID in DNA demethylation is the observation that cooperative action of AID and ten-eleven translocation 1 (TET1) protein can mediate DNA demethylation. It has been reported that 5-hydroxymethylcytosines (5hmC), the product of the catalytic activity of TET1, can be deaminated by AID producing 5-hydroxymethyluracil (5hmU), which in turn can be detected and processed by thymine DNA glycosylase (TDG) and finally replaced by unmethylated cytosines. The ability of AID to deaminate 5hmC was also demonstrated *in vivo* in mice brain that overexpress AID, nevertheless whether this mechanism take place under physiological conditions remains to be determined. In either case, no AID expression has been detected in mice brain, while the expression of *apobec1* is detected (Guo et al., 2011).

Despite the strong evidence that links AID with epigenetic changes, such as DNA demethylation, it seems that this role is restricted to specific scenarios. In parallel, biochemical analysis has revealed substantially reduced AID activity on 5mC relative to cytosines, and no detectable deamination of 5hmC, which argue against the proposed function of AID in DNA demethylation (Nabel et al., 2012). Thus, it is clear that there are several unresolved issues about AID function and regulation.

1.4. Epigenetic regulation

The concept of Epigenetics was formally introduced in 1940s by Conrad Waddington as the "study of the causal interactions between genes and their products which bring the phenotype into being". Since then, the epigenetic field has experimented important advances and later on epigenetics was redefined as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Russo et al., 1996). More recently, Adrian Bird has proposed that epigenetic events can be defined as "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007). The epigenome is the set of epigenetic modifications that occur along the genomic sequence of an individual. Given that the set of epigenetic marks of a cell type is determined by the set of transcription factors, upstream signaling pathways and extracellular signals, an organism will be characterized for different epigenomes for

each of its cell types. The main groups of epigenetic modifications are represented by DNA methylation and histone post-translational modifications.

1.4.1 DNA methylation

DNA Methylation is the most studied epigenetic modification and it consists in the addition of a methyl group in the 5' position of cytosines localized in CpG dinucleotides. This reaction is catalyzed by DNA methyltransferases (DNMTs) which use S-adenosyl methionine (SAM) as a methyl group donor. CpGs are clustered in specific regions across the genome where they play a regulatory role. CpG islands (CGI) are the most abundant CpG regions, and they are defined as regions of around 1000 base pairs (bp) long with an elevated G+C base composition, little CpG depletion, and frequent absence of DNA methylation. In addition, there are other types of regions defined in relation to CGI. 'Shores' are located between 1-2 kb up- or downstream to CGI, whereas 'shelves', are the adjacent regions to shores (Doi et al. 2009; Vinson and Chatterjee 2012). CGI overlap the promoter regions of 60-70% of all human genes. The majority of promoter associated CGI are unmethylated and are linked with a transcriptionally permissive chromatin state, thus providing a platform to recruit transcription factors.

Promoters and enhancers are common hotspots for DNA methylation changes that are functionally relevant. DNA methylation has an important role in different processes such as the maintenance of chromosome stability, genomic imprinting, and the regulation of gene expression. Specifically, DNA methylation plays a fundamental role in differentiation by driving and stabilizing gene activity states during cell-fate decisions (Eden and Cedar 1994).

DNA methylation patterns are established by three enzymes: DNMT1, DNMT3a and DNMT3b. It has been proposed that the initial CpG methylation pattern is established by DNMT3a and DNMT3b which catalyze *de novo* methylation, whereas DNMT1 maintains these methylation patterns during replication (Bestor, 2000; Okano et al., 1999). Additionally, the DNMT family includes the catalytically inactive member DNMT3L, which is a regulatory factor required for *de novo* methylation by enhancing the activity of DNMT3a and DNMT3b (Bestor, 2000).

The mechanism for the addition of a methyl group to a cytosine by DNMTs is well characterized. DNMTs use a conserved mechanism of "base flipping" in which the target nucleotide is projected out of the double helix allowing the transfer of the methyl group from SAM to the activated cytosine C5, in a process coupled with various

release steps. However, DNA demethylation, i.e. the process by which methyl groups are removed from DNA, are still a matter of intense investigation (Cheng and Roberts, 2001; Piccolo and Fisher, 2014).

1.4.2. DNA demethylation

It has been reported that DNA demethylation can occur through passive and active mechanisms. Passive DNA demethylation is achieved through successive cycles of DNA replication in association with an inefficient or absence of maintenance of the DNA methylation patterns by DNMT1, whereas active mechanisms mediate rapid DNA demethylation in the absence of replication (Kohli and Zhang, 2013).

After several nonproductive efforts to find a protein with DNA demethylation activity, it has been proposed that DNA demethylation in mammals occurs through the sequential activity of specific enzymes. One of the latest models of DNA demethylation, proposes the concerted action of TET hidroxylases, cytidine deaminases, DNA glycosylases and the DNA repair machinery.

The proposed DNA demethylation mechanism is initiated by the successive oxidation of 5mC to produce 5hmC by TET enzymes. The produced 5hmC can be further oxidized by TET enzymes to generate 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) or alternatively 5hmC can be processed by AID/APOBEC deaminases to generate 5hmU. The products of the action of TET enzymes (5fC and 5caC) and AID/APOBEC deaminases (5hmU) are detected and removed by Thymine DNA glycosilase (TDG). The generated abasic sites are repaired and filled with unmethylated cytosines by the BER machinery (Kohli and Zhang, 2013). Although there are several evidences that support this model (as mentioned above in the AID section), there are findings that challenge the deamination of 5hmC by AID/APOBEC enzymes by showing the steric hindrance of the hydroxymethyl group in the deamination activity (Nabel et al., 2012). Therefore, the latest DNA demethylation models still are a subject of intense investigation and debate (Figure I10).

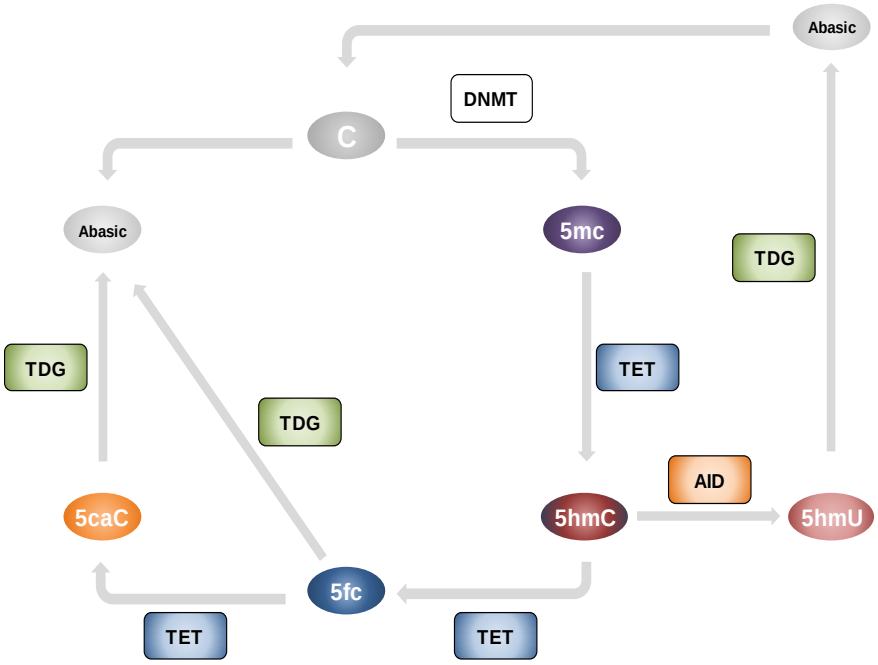


Figure I10. Model of Oxidation-mediated DNA demethylation. Cytosines are methylated by DNA methyltransferases. Methylated cytosines are progressively oxidized by TET enzymes. The oxidized products can be processed directly by TGD which generates abasic sites that are repaired by BER machinery. Alternatively, 5hmC can be deaminated by AID generating 5hmU that can be also excised by TDG. DNMT indicates DNA methyltransferases; TDG, Thymine DNA glycosilase; TET, ten-eleven translocation enzymes; AID, activation-induced cytidine deaminase; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine; 5hmU, 5-hydroxymethyluracil. Adapted from (Kohli and Zhang, 2013)

1.4.3. Histone Modifications

Histone post-translational modifications are the second group of epigenetic marks. DNA and histones are organized in a nucleoprotein complex called chromatin which allows the compaction of the DNA molecule so it can be accommodated in the cell nucleus. The functional subunit of the chromatin, the nucleosome, is composed by a histone octamer around which a DNA segment of 147 bp length is wrapped. The histone octamer is formed by the association of two copies of each of the core histones H2A, H2B, H3, and H4 and nucleosomes are interconnected by DNA segments known as linker DNA which are associated to histone H1 (Tessarz and Kouzarides, 2014).

Chromatin is subjected to a dynamic regulation by covalent modifications of DNA (discussed above) and histone proteins. This chromatin regulation is crucial in DNA-based processes, such as transcription, DNA repair and replication. The covalent modification of histones mainly occurs at the N-terminal tails exposed in the nucleosome, where different chemical groups can be added. To date, the main histone modifications include: lysine acetylation, lysine and arginine methylation, serine and

threonine phosphorylation, lysine ubiquitination, lysine sumoylation, ADP ribosylation, and others. Modification of histone amino acid residues can alter the local charge at the positions that are modified and perturb physical interactions, thus controlling nucleosome interactions and the recruitment of different proteins to DNA. The functional effects of different histone modifications are very specific and it has been proposed that distinct histone modifications, on one or more tails, act sequentially or in combination to form a "histone code" that is read by other proteins to bring about distinct downstream events (Strahl and Allis, 2000).

Methylation of lysine and arginine are among the most studied histone modifications. Arginines can be mono- or di-methylated, whereas lysines can be mono-, di-, or tri-methylated. To date, it has been reported lysine methylation in various amino acid residues at histone H3 and histone H4. Histone H3 can be methylated at residues K4, K9, K27, K36 and K79, whereas histone H4 can be methylated only at the K20 residue. These histone modifications are regulated by the enzymes that add methyl groups, better known as lysine methyltransferases (KMTs), and by enzymes that remove methyl groups, which are the lysine demethylases (KDMs). Histone lysine methylation has proved to play a critical role in different cellular processes such as the regulation of gene expression, cell cycle, genome stability, and nuclear architecture. Furthermore, it has been postulated that histone lysine methylation and DNA methylation coordinate the long-term epigenetic maintenance (Black et al., 2012; Greer and Shi, 2012).

1.4.4. Methylation of histone H4 at lysine 20

The methylation of the lysine 20 constitutes the most common methylated form of histone H4. This lysine residue can be mono-(H4K20me1), di- (H4K20me2) or tri-methylated (H4K20me3), and each of these forms have different physiological roles (Figure I11).

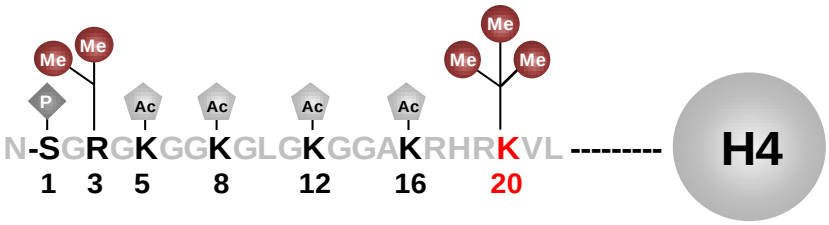


Figure I11. Post translational modifications in histone H4.Histone H4 can be phosphorylated (S1 residue), acetylated (K5, K8,K12 or K16) and methylated (in R3 or K20). Adapted from Sims, Nishioka, & Reinberg, 2003.

In general, H4K20me1 and H4K20me2 have been linked with DNA replication and DNA repair, whereas H4K20me3 has been linked with heterochromatin formation and gene silencing. The H4K20 methylation state varies in a cell cycle-dependent manner. In proliferating cells, the majority of H4K20 is found in the H4K20me2 state, and this mark, together with H4K20me3, does not undergo dramatic changes during cell cycle. Nevertheless, the H4K20me1 undergoes the most important fluctuations. The lowest levels of H4K20me1 are detected during G1 phase and at the beginning of S phase. Then, H4K20me1 levels increase throughout S and G2 phases, reaching a peak during M phase (Sims et al., 2003).

Different KMTs, belonging to the SET domain family, are responsible for the methylation of H4K20 through specific and non redundant roles. PR-SET7 catalyzes the formation of H4K20me1, whereas SUV4-20H1 catalyzes the formation of H4K20me2 (although it can also promote the formation of H4K20me3), and SUV4-20H2 catalyzes the formation of H4K20me3 (Balakrishnan and Milavetz, 2010).

Knockout of the different H4K20 KMTs in mice have provided evidence about the lysine methylation mechanism at histone H4 and the substrate preferences of these enzymes. Although PR-SET7 has been involved exclusively with the formation of H4K20me1, the knockout of this enzyme leads to a global loss of all three methylated forms of H4K20. This finding led to the proposal of a sequential mechanism in which H4K20me2 and H4K20me3 forms are generated from progressive methylation of H4K20me1. Moreover, both SUV4-20H1 and SUV4-20H2 are able to produce H4K20me3 *in vitro,* nevertheless a knockout of SUV4-20H1 in mice cause reduced levels of H4K20me2, whereas the knockout of SUV4-20H2 produce reduced levels of H4K20me3. This suggests that despite the *in vitro* ability of both enzymes, they have particular substrate preferences *in vivo* (Southall et al., 2014).

Further evidence of H4K20 methylation mechanisms were also provided by gain-of-function experiments involving the SUV4-20H enzymes. Tsang and colleagues characterized the localization and activity of SUV4-20H1 and SUV4-20H2 enzymes, and additionally described a splice variant of SUV4-20H1 (SUV4-20H1_i2), which interestingly displayed differential activity and localization. The variant SUV4-20H1_i2 is a truncated version of the canonical form (SUV4-20H1_i1), a variant identical over the first 391 residues followed by two different additional amino acids (Tsang et al., 2010) (Figure I12).

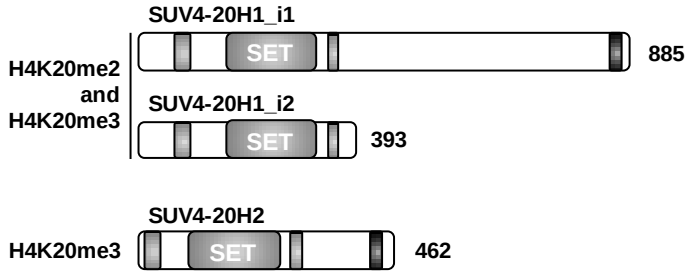


Figure I12. Primary structure of SUV4-20H proteins. All three SUV420H proteins contain the catalytic SET domain, as well as 2 additional regions of conservation (black boxes). The shorter SUV420H1_i2 isoform lacks a 492 amino portion of SUV420H1_i1, which includes a short region of homology with SUV420H2 (gray box). Adapted from Tsang, Hu, & Underhill, 2010.

It has been reported that SUV4-20H1_i1 and SUV4-20H2 are mainly localized at pericentromeric heterochromatin, and this localization is mediated by Heterochromatin protein 1 (HP1) and H3K9me3. Conversely, Tsang and colleagues found that SUV4-20H1_i2 variant is mainly localized throughout the nucleus and its localization is unaffected by HP1. Additionally they reported that whereas SUV4-20H1_i1 and SUV4-20H2 promote the increase of H4K20me3 in PCH, SUV4-20H1_i2 is associated with the increase of this histone mark throughout the nucleus according to its localization. These findings lead to the suggestion that SUV4-20H1_i2 mediates a different H4K20 methylation pathway and raises the complexity of the establishment of H4K20 states and their biological functions (Tsang et al., 2010).

The link between H4K20 methylation, the SUV4-20H enzymes and the immune system was established by the characterization of the mice model with deletion of the *Suv4-20h* genes. Schotta and colleagues generated a *Suv4-20h*-double-null (dn) mice by knocking out *Suv4-20h1*and *Suv4-20h2* genes. These mice showed a genome-wide transition to H4K20me1 and their chromatin had nearly lost all H4K20me3 and H4K20me2. The main consequences of these alterations were the increased sensitivity to damaging stress, and the impaired DNA damage repair. Then, they decided to conditionally delete *Suv4-20h1* and *Suv4-20h2* genes in the hematopoietic system and observed that one of the main effects was the impairment of CSR in B lymphocytes. As explained in previous sections, this is a process that requires the concerted action of different proteins and DNA repair machinery to ensure the proper recombination of gene segments. In the absence of Suv4-20h enzymes there is an accumulation of H4K20me1 that not only impairs CSR, but also increases the translocations and deletions involving the IgH locus (Schotta et al., 2008).

The involvement of H4K20 methylation in the DNA events associated with CSR is only one example of the influence of epigenetic mechanisms during B cell activation. It has been reported that DNA methylation changes and several histone marks have a strong influence and even drive the DNA modifications that occur during CSR and SHM.

1.4.5. Epigenetics mechanisms during antibody diversification

As mentioned before, the transcription is a pre-requisite to initiate both SHM and CSR. The transcription is not only regulated by activators or repressors but also by chromatin conformation. Therefore, SHM and CSR are strongly influenced by epigenetic mechanisms. For instance, DNA hypomethylation and several histone marks have been associated with SHM. It has been described that the V(D)J region is hypomethylated and it has been proposed that this hypomethylation mediates the establishment of activation associated histone marks such as H3K9ac/K14ac, H4K8ac, and H3K4me3. The latter two histone marks have been directly associated with SHM (Li et al., 2013). Increased levels of H4K8ac, induced by the treatment with the histone deacetylase inhibitor trichostatin A, are correlated with an increase of mutations observed at the V(D)J region in the Burkitt's lymphoma derived cell line, BL2 (Woo et al., 2003). Conversely, the knockdown of the histone chaperone suppressor of Ty6 (Spt6) in the same cell line, which decreases the H3K4me3 levels, is correlated with a reduction of V(D)J mutations. These findings suggest an important role of both histone marks during SHM (Begum et al., 2012).

Additionally, it has also been reported that other histone marks such as monoubiquitinated H2A and H2B, and phosphorylated H2B, could help to target the V(D)J region for SHM and to mediate the DNA repair response after AID activity (Borchert et al., 2010; Odegard et al., 2005).

Regarding CSR, important epigenetic changes in the S regions, involved in the establishment of the proper chromatin conformation and the recruitment of CSR machinery, has been described. The Sμ region is constitutively expressed and is poised for switching by displaying an open chromatin state mediated by DNA hypomethylation and the presence of activation histone marks such as H2BK5ac, H3K9ac/K14ac, H3K27ac, H4K8ac, H3K4me3, and H3K36me3 (Li et al., 2013). In the Sμ region it has also been reported low levels of the H3K9me3 repressive mark, which has been linked with the recruitment of the Krüppel-associated protein (KAP)1-

heterochromatin protein (HP)1γ complex, which in turn seems to stabilize AID at the mentioned region (Jeevan-Raj et al., 2011).

Moreover, in resting B cells the remaining S regions are silenced and associated with the repression marks H3K9me3 and H3K27me3. Cytokine stimuli drive the activation of specific S regions and promote the recruitment of histone modifying enzymes that remove the repressive marks and trigger chromatin decondensation, thus allowing the transcription activation necessary to initiate CSR.

Another relevant histone mark during CSR is the combinatorial histone H3K9acS10ph modification, which is recognized by the 14-3-3 adaptors. This mark is associated specifically with S regions and allows the targeting of AID mediated by 14-3-3 adaptors (Xu et al., 2010).

Epigenetic mechanisms play a critical role in regulating the chromatin structure of V and S regions. This regulation not only initiates of SHM and CSR by promoting transcription, but also leads to the recruitment of key proteins in these events. Several epigenetic mechanisms act independently of AID, but there are also epigenetic marks that are associated with AID expression and recruitment. Thus, AID by its proposed role in DNA demethylation and by regulating several histone marks, constitutes a putative epigenetic regulator that could have critical role in the chromatin events during antibody diversification.

1.4.6. Epigenetics mechanisms in the immune system

Throughout time, the physiological role and relevance of different proteins in the immune system has been revealed through their characterization in pathological conditions. Epigenetic mechanisms are not an exception to this scenario and their roles in several aspects of the immune system are evident through the effects caused by their alterations in these processes.

The immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome constitutes a remarkable evidence of the influence of epigenetic mechanisms in the immune system. This is a rare, autosomal recessive disease associated with several recurrent infections that cause high mortality in early childhood. The hallmark of this syndrome is an agammaglobulinemia in the presence of normal B and T lymphocytes counts that occurs as a result of mutations in the *DNMT3B* gene, which is responsible for *de novo* DNA methylation during embryogenesis and germ cell development. The main consequences of mutations in the *DNMT3B* gene are DNA

hypomethylation, mainly restricted to satellite 2 and 3 repeats, and genomic instability in mitogen-stimulated lymphocytes. Interestingly, ICF syndrome patients not only exhibit alterations associated with DNA methylation patterns, but also show alterations in different histone marks and miRNA expression profile, which reflect the interplay between the epigenetic mechanisms (Rodriguez-Cortez et al., 2011)

In the B cell context, epigenetic mechanisms not only are relevant in the antibody diversification during B cell activation, but also regulate important events throughout B cell development. It has been demonstrated that during early B cell development there are important DNA methylation changes associated with gene expression. By comparing the DNA methylation profile among different early B cell stages, Lee and colleagues found a DNA demethylation signature mainly at the CpGs located in gene body regions, and this signature was closely associated with functional transcription factor (TF) sites related to B-cell development. The correlation between DNA demethylation and increased gene expression was observed mainly in the first differentiation stages, therefore they suggested that DNA methylation changes play a smaller role after the B lineage commitment (Lee et al., 2012) .

Nevertheless, in advanced stages of B cell differentiation important epigenetic changes have also been reported. Comparison of the DNA methylomes of naïve, GC, memory and plasma B cells reveals a global shift towards DNA hypomethylation associated with differentiation. It has been described that the major changes occur during GC reaction, thus the main differences in DNA methylation patterns are observed when naïve and GC B cells are compared. Nevertheless, it has also been described differences in DNA methylation profiles when naïve versus memory or naïve versus plasma cells are compared, observing again DNA hypomethylation. Interestingly, despite the differences in the DNA methylome of naive and memory B cells, these two cell types possess a very similar transcriptome and conversely, memory B cells and plasma cells have similar methylomes but very different transcriptomes. It has been suggested that epigenetic mechanisms confer memory B cells the ability to undergo rapid and dramatic gene expression changes by poising its genome through DNA methylation. Thus, memory B cells are "epigenetically prepared" to rapidly differentiate into plasma cells during a secondary immune response (Lai et al., 2013).

Additionally, it has been reported that epigenetic mechanisms not only regulate the expression of the master regulators of B cell differentiation, but also support the gene expression changes driven by them. For instance, the plasma cell transcriptome

is mainly regulated by Blimp-1 transcription factor. The expression of Blimp-1 is epigenetically silenced by histone deacetylases (HDACs) recruited by BCL-6 and by different micro RNAs. The plasma cell differentiation implies the de-repression of Blimp-1. The activation of this transcription factor allows the establishment of plasma cell identity by repressing several key genes through a similar mechanism mediated by HDACs. Furthermore, Blimp-1 represses several genes by increasing the H3K9me3 levels through the recruitment of an H3K9 methyltransferase (Li et al., 2013).

Thus, the concerted action of the different epigenetic mechanisms has a critical role in several immune system events. As mentioned earlier, the vast majority of the knowledge about the normal function of immune system has been acquired by analyzing immune disorders. In this sense, characterizing epigenetic mechanisms in pathological immune scenarios might be helpful to go further in determining epigenetic physiological functions in the immune system. Primary immunodeficiencies constitute an interesting group of disorders that show evidence of epigenetic deregulation, therefore represent a good model to gain insights about the epigenetic influence in the immune system.

1.5. Primary immunodeficiencies

Primary immunodeficiencies (PIDs) are a growing group of heterogeneous disorders which comprises more than 200 alterations in virtually any aspect of the immune response. According to the last report of the International Union of Immunological Societies, there are 9 major groups of PIDs: combined immunodeficiencies, combined immunodeficiencies with associated or syndromic features, predominantly antibody deficiencies, diseases of immune dysregulation, congenital defects of phagocyte number, function, or both; defects in innate immunity, autoinflammatory disorders, complement deficiencies and phenocopies of PID (Al-Herz et al., 2014).

The clinical spectrum of PIDs is broad. Alterations that affect both, immune and non-immune components have been reported, and the clinical manifestations range from the impairment of the immune response against specific pathogens, to several important complications such as autoimmunity and cancer (Fischer, 2004). Different gene mutations have been associated with the development of PIDs, however there are no correlation between the genetic alterations and clinical manifestations, and even different mutations in the same gene are associated with completely different immunodeficiencies. Thus, the huge variability in these disorders highlights the influence of other modifying genetic elements, such as environmental factors and epigenetic mechanisms (Liadaki et al., 2013).

Among the PIDs, the primary antibody deficiencies (PADs) represent the most common inherited immunodeficiencies in humans, in which the hallmark is the reduction or absence of serum immunoglobulins (Durandy et al., 2013). Multiple gene defects can account for the impaired production of antibodies and these alterations are related not only to B cells, but also with another immune cells or molecules. (Conley et al., 2009) (Figure I13).

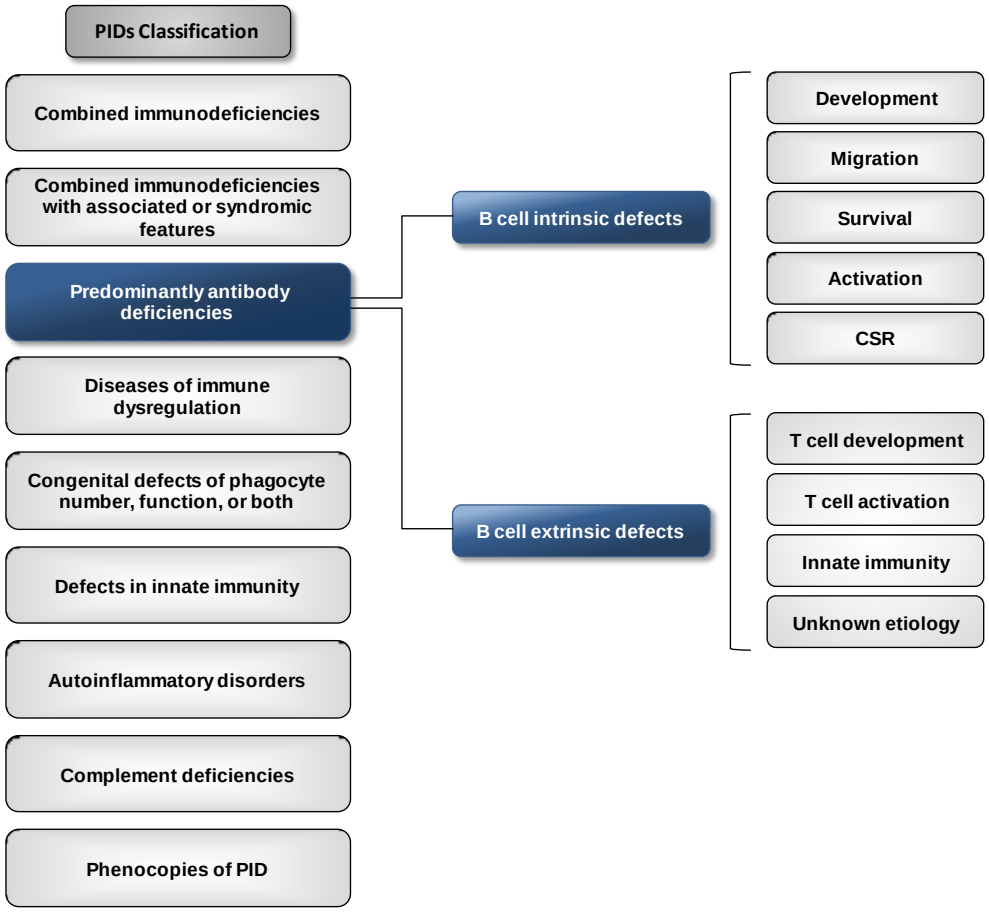


Figure I13. Classification of Primary Immunodeficiencies Adapted from Al-Herz et al., 2014.

Many of the alterations observed in PADs that are related with B cell biology involve B cell intrinsic defects. They could be associated with early B cell development, as is exemplified by mutations in genes that affect the expression of components of the pre-B cell receptor (pre-BCR), such as *IGLL1* or *IGHM*, which are the genes that encode the λ5 chain or the μ-chain of the pre-BCR (Minegishi et al., 1998; Yel et al., 1996). There are also alterations in molecules that mediate the pre-BCR o BCR signaling. For instance, it has been reported mutations in *CD79A* and *CD79B*, the

genes that encode the pre-BCR and BCR co-receptors, Igα and Igβ (Ferrari et al., 2007; Minegishi et al., 1999a). Mutations in *BLNK* gene, which encodes the B-cell linker (BLNK) scaffold protein, that mediates the assembly of the signalosome upon cross-linking of BCR have also been reported (Minegishi et al., 1999b); and the most frequent mutations are found in the gene that encodes the Bruton's tyrosine kinase (BTK), and cause the X-linked Agammaglobulinemia (XLA), which is considered the prototype immunodeficiency (Conley et al., 1998). Alteration of any of these molecules completely abrogates B cell differentiation causing an arrest in the pre-B cell stage. Therefore there is a lack of mature B cells which implies a total absence of antibodies, better known as agammaglobulinemia.

There also are more intrinsic alterations that occur later in the B cell differentiation process that contribute to the development of PADs such as defects in CSR, or impairment of the migration, survival, or activation of B cells. Hyper-IgM syndrome (HIGM) involves specific defects in CSR, while Common variable immunodeficiency (CVID) can be associated with alterations in any of these B cell related processes. Therefore, both disorders constitute two good models of PADs associated with B cell intrinsic defects in later stages of B cell differentiation.

1.5.1. Hyper-IgM Syndromes

Hyper-IgM syndromes, alternatively called Class-Switch Recombination Defects, comprise a group of rare disorders characterized by the presence of normal or elevated levels of serum IgM and the absence or strongly decreased levels of the other Ig isotypes, IgG, IgA, IgE (Durandy et al., 2005).

B cells defects that have been associated with CSR-defects can be associated with the upstream signaling pathways that trigger the antibody production, or with the molecules directly involved in the antibody maturation process (Durandy et al., 2007). To date, six main types of HIGM syndromes have been described (HIGM1 to HIGM6) according to an underlying genetic defect. Additionally, newly described alterations have been incorporated into the HIGM group of disorders, but remain without a specific classification (Davies and Thrasher, 2010) (Table I1).

Table I1. Genetically defined types of HIGM syndromes.

HIGM type	Defect	Inheritance	Infection susceptibility	Lymphoid Hypertrophy	Autoimmunity	Lymphoma	CSR defect	SHM defect	DNA Repair defect
1	XHIM-CD40 L deficiency	XL	Bacterial, opportunistic	-	Yes	No	Yes	Yes	No
2	AID deficiency	AR	Bacterial	++	Yes	No	Yes	Yes	No
*	AID C terminal defect	AD	Bacterial	+	?	No	Yes	No	No
3	CD40 defect	AR	Bacterial, opportunistic	-	Yes	No	Yes	Yes	No
5	UNG deficiency	AR	Bacterial	+	?	Probable	Yes	No	No
6	NF-κB signaling	XL/AD	Bacterial, opportunistic		Yes	No	Yes	Yes	No
	PMS2 deficiency	AR	Bacterial	?	?	?	Yes	No	No
	Complex disorders affecting NHEJ DNA repair	AR	Mainly bacterial some opportunistic	-	Yes	Yes	Yes	No	Yes

Adapted from Davies & Thrasher, 2010. XL, X linked; AR, autosomal recessive; AD, autosomal dominant. AID, activation-induced cytidine deaminase; UNG, uracil N glycosylase; PMS2,vpostmeiotic segregation increased 2; NHEJ, non-homologous end joining.Type 4 HIGM refers to a genetically undefined type.

Specifically, HIGM type 2 (HIGM2) is associated with deleterious mutations in *AICDA* gene leading to defective CSR and SHM, as discussed above. Patients carrying *AICDA* mutations are susceptible to recurrent bacterial infections in the respiratory and intestinal tracts, nonetheless the clinical hallmark of AID deficiency is the enlargement of lymphoid organs and at the histological level, the presence of giant germinal centers (GCs). Additionally, around 20% of the HIGM2 cases are associated with autoimmune complications that are manifested as immune cytopenias, arthritis or hepatitis (Durandy et al., 2005; Revy et al., 2000).

At least 39 different mutations in HIGM2 patients have been described. These autosomal recessive inherited mutations are distributed throughout the AICDA gene, and include amino acid substitutions, premature stop codons, and deletions (Figure I14).

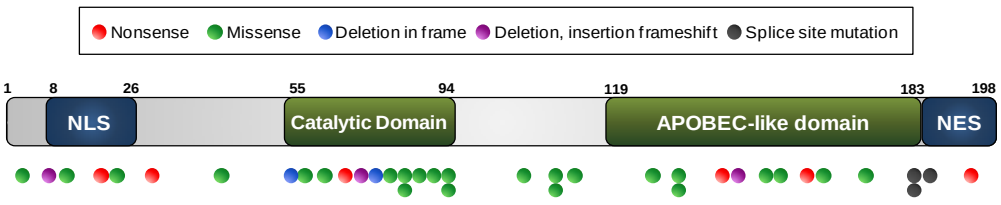


Figure I14. Schematic representation of AID structure and the localization of the different AID mutations reported in HIGM. NLS indicates nuclear localization signal; NES, nuclear export signal. Adapted from (Durandy et al., 2005).

By comparative analysis using the crystal structure of the catalytic domain of APOBE3G as a model, it has been proposed that of the 23 different missense AID mutations found in HIGM2 could be classified in three categories: “catalysis” (class I), “substrate interaction” (class II), “structural integrity” (class III). None of these AID mutants exhibit CSR or SHM activity *in vivo*, however this analysis revealed that 4 of the missense mutations generate catalytic active forms of AID that even show properties similar to WT AID with respect to motif specificity, substrate binding, processivity, mutation spectra and spatial deamination patterns. This interesting findings suggest that the defective CSR and SHM observed *in vivo*, rather than be caused by mutations that affect directly the AID catalytic activity, are related to more subtle effects that involve perturbations to the specificity loop, hydrophobic core, and secondary structures of AID protein (Mu et al., 2012). Thus, these perturbations could impair the interactions needed to achieve CSR and SHM.

The latter assumption is more evident with the analysis of the C-terminal mutants of AID which in fact, constitute a fourth class of AID mutations with particular properties that differentiate them from the rest of AID mutations (Mu et al., 2012). Patients carrying class IV mutations present a milder form of HIGM and AID mutants, despite being catalytic active, are not able to generate CSR but retain the SHM activity, which indicates that the C-terminal region of AID carries a critical binding site for cofactors that regulate AID activity during CSR (Ta et al., 2003).

One specific mutation in the C-terminal region of AID is associated with an autosomal dominant form of HIGM (AD-HIGM). This heterozygous nonsense mutation in the *AICDA* gene replaces the arginine at position 190 by a stop codon (R190X) which results in the truncation of the last nine C-terminal amino acids. It has been proposed that this truncated form of AID exerts a dominant negative effect on CSR by failed recruitment of DNA damage response factors (Imai et al., 2005).

Analysis of the different AID mutants at the clinical level does not provide too much information since all AID mutations (except in AD-HIGM) exhibit the same phenotype with slight variations, but undoubtedly, the molecular analysis of these mutants is extremely useful for understanding the physiological behavior of AID and the critical protein regions that regulate its function.

1.5.2. Common Variable Immunodeficiency

Another immunodeficiency treated in this doctoral thesis is the Common Variable Immunodeficiency (CVID). CVID is the most frequent symptomatic PID in adults and

comprises a heterogeneous group of disorders that are characterized by low immunoglobulin levels and a failure to respond to protein or polysaccharide antigens, a condition that predisposes to recurrent bacterial infections that mainly affect the respiratory and gastrointestinal tracts (Cunningham-Rundles and Ponda, 2005).

CVID patients present a marked decrease in IgG (hypogammaglobulinemia) and in at least one of the isotypes IgM or IgA. The clinical manifestations of CVID are highly variable, but three criteria help to the diagnosis: the onset of immunodeficiency after the age of 2 years, a poor response to vaccines and the exclusion of other previously reported causes of hypogammaglobulinemia (Conley et al., 2009).

CVID patients display several complications such as lymphoproliferative disorders and inflammatory diseases. In addition, more than 25% of CVID cases present autoimmune complications which are poorly understood and sometimes are related with poor prognosis (Baldovino et al., 2013). These cases reflect the complexity of the immune system and the coexistence of two different situations: immunodeficiency and autoimmunity.

Several efforts have been made to find genetic alterations that would help understand the CVID etiology and predict the clinical outcome. However, genetic alterations only can account for about 10-20% of immunodeficiency patients, and the majority lack a definite molecular genetic diagnosis (Salzer et al. 2012). It has been described that genetic defects can affect several immune processes, mainly those related to B cell activation, T-cell signaling and cytokine expression (Kokron et al. 2004). For instance, there are reports of alterations in the genes encoding CD19, CD21, CD81, PLC-γ2, which are proteins involved in signaling through the B cell receptor, and have also been described in the genes encoding ICOS, BAFFR, TACI, MSH5, involved in co-stimulatory pathways necessary for isotype switching and somatic hypermutation during B cell activation (Castigli and Geha 2006).

Different immune cell types have shown to be deregulated in CVID. A deregulation of T-cell subsets with a reduction of T-regulatory and CD4 cells, and an increase in CD8 cells has been reported (Xiao et al. 2014). In addition, there is evidence that a reduction of both myeloid and plasmacytoid dendritic cells are correlated with a greater incidence of autoimmunity, splenomegaly and granulomatous disease, and a higher incidence of clinical complications in CVID patients (Yong et al. 2008). However, the main alterations occur at late B cell differentiation steps which cause a dramatic disturbance of B cell homeostasis. Specifically CVID patients display a severe deficiency of switched memory B cells ($CD27^{+}IgM^{-}IgD^{-}$) (Warnatz et al. 2002)

and, in fact, it has been proposed to classify patients on the basis of the quantitative repartition of naïve/memory B cells according to the dual expression of IgD and CD27 (Piqueras et al. 2003). Moreover, it has been described the association of the increase of certain B cell subsets, such as $CD21^{lo}$ or transitional B cells, with splenomegaly and lymphadenopathy respectively, while the reduction of class-switched memory B cells has been associated with granulomatous disease and autoimmune cytopenias (Wehr et al. 2008).

In summary, HIGM and CVID constitute two examples that illustrate the complexity of the regulation of immune system. A global picture of this process is obtained not only by identifying specific defects that reveal the physiological functions of certain proteins, but also by integrating this knowledge with more general processes, such as epigenetic mechanisms. Thus, this doctoral thesis was conceived to investigate the role of epigenetic mechanisms in the context of the immunodeficiencies HIGM and CVID.

OBJECTIVES

2. OBJECTIVES

The proper function of the immune system requires complex regulatory mechanisms and a highly strict balance in the amount and function of immune and non-immune elements. Part of the primary immunodeficiencies result from mutations in specific genes and their clinical manifestations can be recapitulated through the generation of knockout mice models that support the role of these genes. However, an important number of these disorders cannot be explained by genetic alterations and, to date, there is not an alternative explanation to understand it. This makes difficult the diagnosis process, the establishment of prognosis markers and complicates the finding of specific treatments or the improvement of the existing ones. Epigenetic mechanisms, mainly DNA methylation and histone modifications, are elements of gene control and have emerged to provide explanation to a wide variety of diseases including those related to the immune system. For that reason, this doctoral thesis was focused on investigating the influence of the epigenetic mechanisms in two primary immunodeficiencies: Hyper-IgM Syndrome and Common Variable Immunodeficiency.

This was achieved by the establishment of the following specific objectives:

1. To investigate the influence of activation-induced cytidine deaminase (AID), commonly mutated in Hyper-IgM Syndrome, in the setting of epigenetic modifications in an inducible B cell model.

   It was decided to explore the direct influence of AID in the global levels and sequence-specific profile of histone modifications and DNA methylation by comparing the cells before and after the AID expression. Once interesting variations caused by AID are identified, a further exploration of the possible implicated mechanisms would follow.

2. To analyze the effects of AID mutations in the acquisition of epigenetic alterations.

   The analysis of AID mutant forms are not only useful to gain insight in the AID regions involved in the alteration/regulation of epigenetic mechanisms, but also to understand the underlying mechanisms implicated in HIGM syndrome beyond the impairment of CSR and SHM.

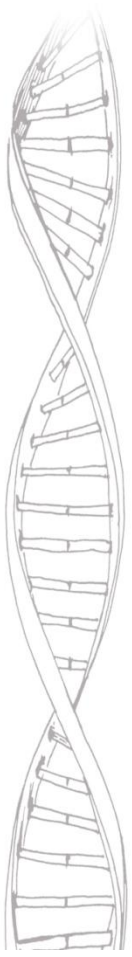
3. To determine the participation of epigenetic alterations in CVID by focusing on the DNA methylation profiling of B cells isolated from monozygotic twins discordant for CVID.

B cells, as the main altered cell type in CVID, represent the best cell population to explore the possible contribution of epigenetic mechanisms in these disorders. In addition, the monozygotic twins discordant for CVID constitute an excellent model in which genetic alterations are not involved in the development of the disease.

4. To expand the results obtained with monozygotic twins discordant for CVID in three different B cell subsets from a cohort of healthy donors and CVID patients.

   This analysis not only will serve to confirm the results obtained in the monozygotic twins discordant for CVID, but also to determine the specific DNA methylation profile of the different B cell subsets and to detect specific alterations in CVID. Class-switched B cells are the main B cell subpopulation altered in CVID, therefore the analysis of DNA methylation in this cell subset contributes to detect variations that could occur during B cell development and that could be important in CVID etiology.

MATERIALS AND METHODS



3. MATERIALS AND METHODS

This doctoral thesis is composed by two studies: Analysis of the role of AID in setting epigenetic marks and its relationship with HIGM, and study of DNA methylation alterations in CVID. In the following section, the materials and methods used in each study are presented separately.

3.1. Methods used for the epigenetic study on AID and HIGM

3.1.1. Cell culture

The cell lines used in this study were grown in a humidified incubator at 37°C and 5% CO_2 and were maintained in log phase growth by changing the culture media every 48 - 72 hours. Burkitt's lymphoma derived cell line Jiyoye was cultured RPMI medium supplemented with 5% FBS (v/v) and antibiotic/antimycotic solution, whereas adherents Hela and 293F cell lines, were grown in DMEM medium with 5% FBS (v/v) and antibiotic/antimycotic solution.

3.1.2. DNA constructs and system for inducible AID expression

A C-terminally Haemagglutinin (HA)-tagged AID was generated by PCR amplification using the primers listed in the 3.1.11. section. The forward primer introduced a BamHI site and the reverse primer introduced an EcoRI site. The amplified fragment corresponding to AID WT was then cloned into BamHI/EcoRI-digested pRetroX-Tight-Pur vector. AID mutants were generated with the Quickchange method (Stratagene) using the primers listed in the 3.1.11. section. To generate an AID inducible system we used the Retro-X ™ Tet-ON ® Advanced Inducible Expression System (Clontech). First, the Jiyoye and Hela cell lines, with no detectable protein levels of endogenous AID, were transduced with the RetroX-Tet-ON advanced vector and Geneticin selected at 1mg/ml (Life Technologies). Second, the cells were transduced with the pRetroX-Tight-Pur vector encoding AID and selected with Puromycin (Sigma-Aldrich) at 0.3μg/ml for Jiyoye cells and 3μg/ml for Hela cells.

The AID expression was induced by the addition of doxycycline (Clontech) at 500ng/ml to the culture media during 24 hours for Jiyoye cells and 48 hrs for Hela cells. The nuclear export was inhibited by the addition of Leptomycin B (LMB) (LC labs) during 2 hours at 10ng/ml for Jiyoye cells and 50ng/ml for Hela cells as previously described (Patenaude et al., 2009) (Figure M1).

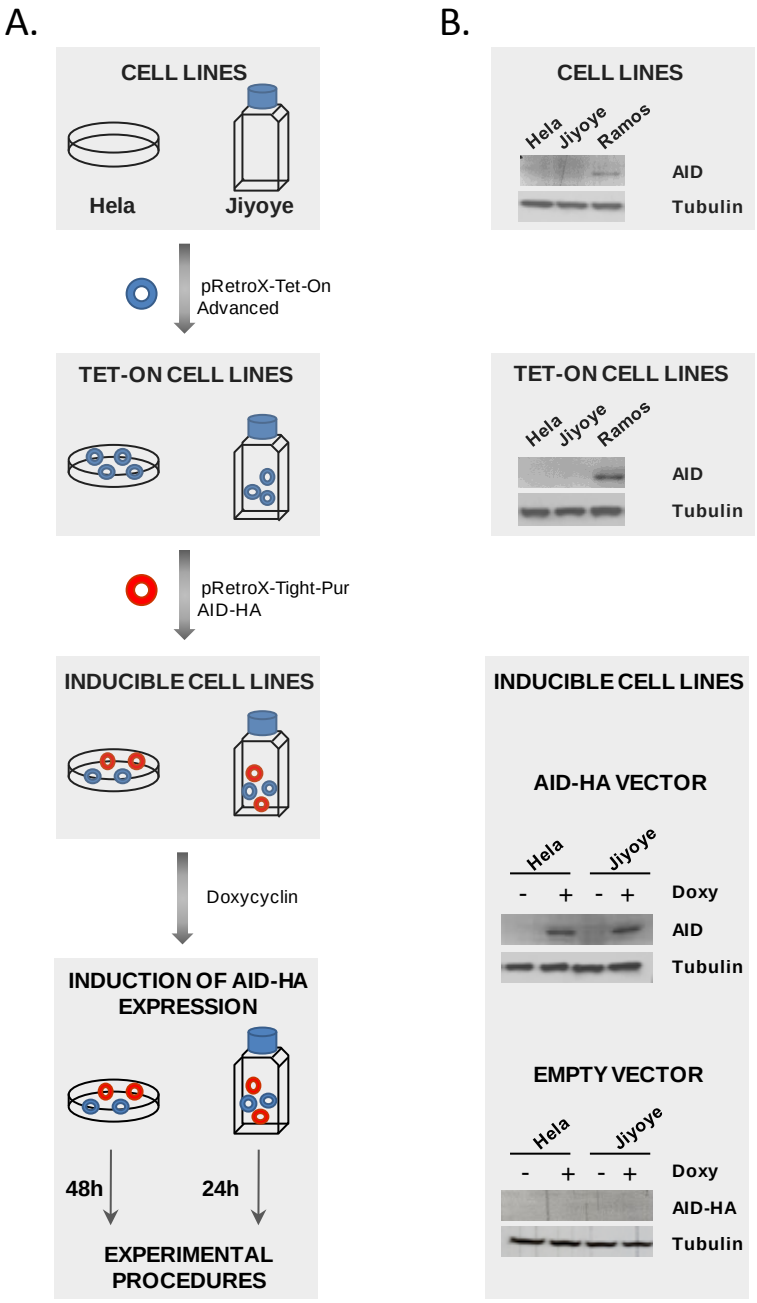


Figure M1. Schematic representation of the procedure to establish the double-stable-inducible cell lines used in the HIGM study. (A) The Hela and Jiyoye cell lines were infected with the RetroX-Tet-ON advanced vector (in blue) and selected with Geneticin .Then the stable cell lines were infected the pRetroX-Tight-Pur vector (in red), which encodes the C-terminal Haemaglutinin tagged AID (AID-HA). After puromycin selection the inducible cell lines for the expression of AID were obtained. The AID-HA expression was induced with Doxycicline treatment.(B) Representative images of western blot assays showing the establishment of AID inducible expression. Hela and Jiyoye cell lines do not express AID. The AID expression is detectable after the infection with the two vectors and doxycycline treatment. Ramos cell line was used as a positive control of AID expression. Negative controls are double infected cells with the RetroX-Tet-ON advanced vector and an empty pRetroX-Tight-Pur vector.

Experimental procedures with the inducible system for AID expression were carried out in four different conditions: Control (C): Cell lines without AID expression. Doxycycline (D): Cell lines after doxycycline treatment to induce AID expression. Doxycycline + LMB (DL): Doxycycline and LMB treatment to induce AID expression and its nuclear accumulation. Control + LMB (CL): Cell lines with LMB.

3.1.3. Western blot assay

Proteins were separated using 8%, 12.5% and 15% acrylamide del for SDS-PAGE and blotted onto a polyvinylidene difluoride membrane of 0.22 μm or 0.45 μm pore size (Immobilon PSQ, Millipore) depending on the size of the protein of interest in each experiment. The membranes were blocked with 5% milk TBS-T (Tris-buffered saline with 0.1% Tween-20) and immunoprobed with the antibodies listed in the 3.1.12. section. Experiments were performed in triplicate. Bands were quantitated by direct scanning of the western blot films with a HP scanjet 4890 and processed with ImageJ software.

3.1.4. Immunofluorescence and Confocal microscopy

Immunofluorescence experiments were carried out as previously described (Patenaude et al., 2009) with some modifications. Briefly, transduced Jiyoye cells were plated in 24 well plates and allowed to attach to poly-Lysine-coated coverslips for 15 min at 37ºC in PBS. Hela cells were plated in 6 well plates and allowed to attach to coverslips 24 hours prior AID expression induction. After this time cells were fixed with 4% paraformaldehyde, permeabilized with 0.5% (v/v) Triton X-100 and blocked with 1 mg/ml BSA + 5% goat normal serum in PBS. Both cells types were stained with the antibodies listed in the 3.1.12 section. Stained preparations were mounted in Mowiol-DAPI mounting medium and confocal optical sections were obtained using a Leica TCS SP5 Spectral confocal microscope (Leica Microsystems). Images were acquired with Leica Application Suite Advanced Fluorescence (LAS AF) software and processed with ImageJ software.

3.1.5. DNAse I digestion-based assay

Through this assay we followed the release of AID and different histone marks after DNase I digestion. DNase I digestion was carried out in nuclei isolated from Jiyoye cell line after the treatment with doxycycline and LMB, to induce AID overexpression and nuclear accumulation respectively. Briefly, cell cytoplasms were disrupted by using RSB buffer (10 mM Tris-HCl pH 7.5, 10mM NaCl, 3mM $MgCl_2$). Then, the isolated

nuclei were washed and resuspended in DNase I buffer prior to the addition of DNase I to initiate the DNA digestion at 37ºC. We monitored the time course of DNase I-induced release of AID and histone proteins at 2, 4, 8, 16 and 32 minutes of digestion. The reaction was stopped by adding EDTA 0.5mM. These procedures were carried out in the presence of Roche Complete Protease Cocktail Inhibitor.

It is well established that sensitivity to nuclease digestion is determined by chromatin structure. After DNase I digestion is possible to isolate a highly nuclease-accessible chromatin fraction and a DNase-resistant fraction. The highly accessible chromatin fraction (soluble fraction) is mainly associated with active genes and therefore with activation associated histone marks, whereas the DNase-resistant fraction (Insoluble fraction) is mainly associated with constitutive heterochromatin and regions of chromatin associated with large multimeric (transcriptional and DNA-repair) protein complexes (Thambirajah et al., 2012).

From each digestion time point sample we obtained both protein fractions (soluble and insoluble) and subjected them to western blot analysis to determine the pattern of appearance or disappearance of AID and the specific histone marks H3K20me3 (heterochromatin mark), H3K27me3 (facultative heterochromatin mark) and H3K4me3 and acetylated H3 (euchromatin marks).

3.1.6. Chromatin immunoprecipitation

Chromatin immunoprecipitation assays were carried out as previously described (Ballestar et al., 2003). Briefly, chromatin was fixed with formaldehyde at final concentration of 1% during 15 min. After fixation, the cells were disrupted with SDS lysis buffer, and the chromatin was sheared to an average length of 0.25 kb with a Bioruptor ultrasonicator (Diagenode). Then, 50μg (for AID immunoprecipitation) or 25μg (for histone marks immunoprecipitation) of chromatin were incubated overnight with the antibodies listed in the 3.1.12 section. IgG was used as a negative control. The formed immunocomplexes were precipitated with the Magna ChIP™ Protein A+G Magnetic Beads (Millipore) and after reverting the crosslinking of protein-DNA interactions, the solution was digested with proteinase K, and the immunoprecipitated DNA was isolated and purified by phenol-chloroform extraction. DNA was amplified and quantified using real-time PCR.

3.1.7. Amplification of UnMethylated Alu's

The Amplification of UnMethylated Alu's (AUMA) protocol was performed as previously described (Hernando et al., 2013; Rodriguez et al., 2008). Briefly, 1μg of DNA of each condition was digested with the methylation sensitive restriction endonuclease SmaI (Roche Diagnostics GmbH, Mannheim, Germany) following the manufacturer instructions. Blue and 5' phosphorilated MCF oligonucleotides adaptors were ligated to the digested product with T4 DNA ligase (New England Biolabs,Beverly, MA, USA). Alu sequences were enriched and amplified from the adaptor-ligated DNA fragments using a chimeric primer comprising the complementary linker sequence (ATTCGCAAAGCTCTGA), the cut SmaI site (GGG) and two additional nucleotides homologous to the Alu consensus sequence AUMA-TT (ATTCGCAAAGCTCTGAGGGTT). A bandrich fingerprint was obtained through high-resolution gel electrophoresis and the pattern was observed by silver staining. The AUMA fingerprints were visually checked and classified following this criteria: hypomethylation (increased intensity), hypermethylation (decreased intensity) and no change (no substantial difference between samples). Only those bands showing clear changes in their fingerprint intensities were considered to represent methylation changes.

3.1.8. Immunoprecipitation

For immunoprecipitation experiments we used 293F cells. These cells were transfected using polyethylemine (PEI) with the following plasmids: pCDNA3.1 constructs encoding the FLAG tagged WT and mutants forms of AID and plasmids encoding the lysine 20 histone 4 methyltransferases, pCDNA4/T0 SUV4-20H1_i1-HA, pCDNA4/T0 SUV4-20H1_i2-HA and pCI SUV4-20H2-HA, that were a generous gift from Dr. D. Alan Underhill (Tsang et al., 2010). Whole cell extracts were obtained with RIPA buffer (50 mM Tris-HCl pH7.8, 150 mM NaCl, 0.5% deoxycholic acid, 0.1% SDS, 1% Nonidet P40) and treated with benzonase 40U (Sigma-Aldrich) during 4 hours prior immunoprecipitation. After this time, samples were centrifuged during 10 minutes at 7000g. The supernatant was then diluted with BC-100 buffer (10 mM Tris-HCl pH 7.8, 0.5 mM EDTA, 0.1 mM PMSF, 1 mM DTT, 50% Glycerol, 100 mM KCl) and overnight incubated at 4ºC with FLAG-agarose (Sigma-Aldrich) for AID immunoprecipitation, or HA-agarose (Sigma-Aldrich) for SUV4-20H immunoprecipitation. The samples were washed twice with BC-100, 0.05% NP-40 and five times with BC-500 (50 mM Tris-HCl pH 7.8, 2.5 mM EDTA, 0.5 mM PMSF, 5 mM DTT, 50% Glycerol, 500 mM KCl), 0.05% NP-40 prior to the protein elution with 0.2 M Glycine pH 2.3. Then, the eluted material

was analyzed by western blot. The results were presented as the ratio among the eluate fraction versus the input fraction, and these ratios were normalized using the AID WT ratio as a reference.

3.1.9. DNA accessibility assay

To analyze the DNA accessibility after AID overexpression, we performed a DNase I digestion followed by amplification by quantitative PCR with the primers listed in the 3.1.11 section, as previously described (Hernando et al., 2014). Briefly, nuclei from Jiyoye cells transduced with AID or an empty vector and treated with doxycyline or doxycycline + LMB, were digested with DNase I. The reaction was stopped at different time points and DNA was isolated and purified by phenol-chloroform extraction to subsequent analysis. The DNA accessibility was represented as the ratio among the final reaction point (30 minutes) and the initial reaction point (1 minute). D4Z4 microsatellite repeats and c-fos gene regions were used as controls of high DNA compaction (low accessibility) and low DNA compaction (high accessibility) respectively.

3.1.10. Expression of μ Germline transcripts

To analyze the expression of μ Germline transcripts (μGLT) of the *IGH* locus, we isolated total RNA with Trizol® Reagent (Life technologies). Then, mRNA was retrotranscribed with the Transcriptor First Strand cDNA Synthesis Kit (Roche). cDNA synthesis reaction was carried out by using anchored oligo$(dT)_{18}$ primer, specially designed to enrich the retrotranscription of mRNAs. We analyzed the expression of mature μGLT by amplifying a 106 bp region within the non-coding Iμ exon (The primers used are listed in the 3.1.11. section). cDNA from Ramos cell line, which is another Burkitt's lymphoma derived cell line and a Jiyoye RNA processed without reverse retrotranscriptase were used as a positive and negative control respectively. The μGLT expression was analyzed by real-time PCR and the results were presented as the mRNA levels relative to the expression of RPL38 (Ribosomal protein L38) housekeeping gene and relative to the mRNA expression in uninduced Jiyoye cells.

In addition to the methods described above, the HIGM study shares with the CVID study the methods related with DNA methylation analysis. Those methods are described in the 3.2 section.

3.1.11. List of primers used in the epigenetic study on AID and HIGM

	Forward Primer	Reverse Primer	Reference
Cloning			
AICDA WT	ATGGATCCAGACACTCTGGACACCACTATG	TAGAATTCCTAAGCGTAATCTGGAACATCGTA	
Site-Directed Mutagenesis			
R24W (70 C>T)	GCTGGGCTAAGGGTTGGCGTGAGACCTACC	GGTAGGTCTCACGCCAACCCTTAGCCCAGC	
W68X (203 G>A)	AAATGTCCGCTGGGCTAAGGGTCGGCGTGA	TAGCACAGGTAGGTCTCACGCCGACCCTTA	
C87R (259 T>C)	CCTCCTGGAGCCCCCGCTACGACTGTGCCC	GGGCACAGTCGTAGCGGGGGCTCCAGGAGG	
M139V (415 A>G)	TGCAAATAGCCATCGTGACCTTCAAAGATT	AATCTTTGAAGGTCACGATGGCTATTTGCA	
R174S (522 A>C)	GTTCGTCTCTCCAGCCAGCTTCGGCGCATC	GATGCGCCGAAGCTGGCTGGAGAGACGAAC	
R190X (568 C>T)	AGGTTGATGACTTATGAGACGCATTTCGTA	TACGAAATGCGTCTCATAAGTCATCAACCT	
ChIP			
Sμ	TGAGATGGCTTTAGCTGAGACAAG	CAGCTCACCTGGTGCAACTTAG	
Cμ	CACGTGGTGTGCAAAGTCCAGCACC	ACGCCAGACCCCACCTGCTT	(Xu et al., 2010)
Bisulfite pyrosequencing			
Sμ - amplicon	GTTGAATTGGGTTGAGTAGGTT	[Btn]CCAACTCACCCTAATTCAACTT	
Sμ - sequencing primer 1	GGTTGAGTAGGTTGT		
Sμ - sequencing primer 2	GTTGAGTTGAATTGGG		
Cμ - amplicon	AATAAAGAAAAGAAYGTGTTTTTT	[Btn]CAACCAAAACACCTAAATCTACC	
Cμ - sequencing primer 1	GGGATAGAGAGGGAG		
Cμ - sequencing primer 2	AGTGATTGTCRAGTTGTTTTTTAAA		
Cμ - sequencing primer 3	GGTTTTTTCRGTAATTTT		
PAX5 - amplicon	[Btn]GGGATTCRGTTTGGAGGGG	CCTACCTATCCAACATCAAAAAATAATCA	
PAX5 - sequencing primer	AAATAATCAAAAACTCCCAT		
AUMA			
Blue adaptor	CCGAATTCGCAAAGCTCTGA		(Rodriguez et al., 2008)
P-MCA adaptor	^{P}-TCAGAGCTTTGCGAAT		(Rodriguez et al., 2008)
ALU up 5 primer	ATTCGCAAAGCTCTGAGGGTT		(Rodriguez et al., 2008)

3.1.11. List of primers used in the epigenetic study on AID and HIGM (continued)

	Forward Primer	Reverse Primer	Reference
µGLT expression			
Iµ	ATGGATCCAGACACTCTGGACACCACTATG		(Fujieda et al., 1996)
Iµ		TGCTCTGAGGTATCGAAAAAG	
DNA accesibility assay			
Sµ	TGAGATGGCTTTAGCTGAGACAAG	CAGCTCACCTGGTGCAACTTAG	
Cµ	CACGTGGTGTGCAAAGTCCAGCACC	ACGCCAGACCCCACCTGCTT	(Xu et al., 2010)
D4Z4	CTCAGCGAGGAAGAATACCG	ACCGGGCCTAGACCTAGAAG	(Hernando et al., 2014)
c-fos	CGAGCATCTGAGAAGCCAAG	GAAGCCCGAGAACATCATCG	

[Btn] indicates Biotinylated primers

3.1.12. List of antibodies used in the epigenetic study on AID and HIGM

Antibody	Company	Reference	Application
α- hemagglutinin (HA)	Sigma	H6908	WB,IF, ChIP
α- FLAG	Sigma	F7425	WB
α- AID	Invitrogen	39-2500	WB
α- histone H3	Abcam	ab1791	WB
α- H3K4me3	Millipore	upstate 07-473	WB, ChIP
α- H3K27me3	Millipore	upstate 07-449	WB, ChIP
α- histone H4	Abcam	ab10158	WB
α- H4K20me3	Millipore	upstate 07-463	WB
α- H4K20me3	Millipore	upstate 07-749	ChIP
Goat anti-rabbit conjugated to horseradish peroxidase (HRP)	Amersham	RPN4301	WB
Sheep anti-mouse-HRP	Amersham	RPN4201	WB
Alexa Fluor® 488 Donkey Anti-Rabbit IgG (H+L) Antibody	Life technologies	A-21206	IF

3.2. Methods used for the study on DNA methylation alterations in CVID

3.2.1. Patients and ethics statement

Human blood samples used in this study were obtained from CVID patients, diagnosed according to established criteria, including one of the siblings of the monozygotic twin pair, and from blood donors. They were collected at the University Hospital Dr. Negrín of Gran Canaria (monozygotic twins discordant for CVID) and the University Hospital La Paz in Madrid (CVID cohort and healthy donors). The blood donors received oral and written information about the possibility that their blood would be used for research purposes, and any questions that arose were then answered. Before giving their first blood sample the donors signed a consent form at their corresponding hospital, which adhered to the principles set out in the WMA Declaration of Helsinki. The protocol used to isolate B cells from these anonymous donors was approved by IDIBELL's Committee of Biosecurity (CBS) on May 5 2011 and the Ethics Committee of the University Hospital of Bellvitge (CEIC) on May 28 2011.

3.2.2. Sample preparation and isolation of B cell subsets

Peripheral blood mononuclear cells (PBMCs) from monozygotic twins were isolated by Lymphoprep™ (Stem Cell Tech Inc., Vancouver, BC, Canada) density gradient centrifugation. The collected cells were washed twice with ice-cold PBS, followed by

centrifugation at 2,000 rpm for 5 min. The CD19+ cells were isolated by positive selection using CD19 MicroBeads (Miltenyi Biotec, Cologne, Bergisch Gladbach, Germany).

For the isolation of CVID and healthy control group B cells, PBMCs were obtained from peripheral blood by Ficoll-Paque Premium (VWR Internacional, Eurolab) density centrifugation. PBMCs were enriched in B cells by negative depletion using CD3 and CD14 Microbeads (Miltenyi Biotec). For the isolation of naïve (CD19+CD27-IgD+), unswitched memory (CD19+CD27+IgD+) and switched memory (CD19+CD27+IgD-) B lymphocytes, B cell enriched PBMCs were stained with CD19 FITC, CD27 APC (BD, Biosciences, Becton Dickinson) and IgD PE (Southern Biotech). Cells were sorted on a FACS® Aria (Becton Dickinson). Purity check was >95% for all selected fractions. Purified samples were pelleted and stored at -80ºC.

3.2.3. DNA isolation and Bisulfite modification

Pelleted cells were resuspended in 750µl of lysis buffer (50mM Tris pH 8.8, 10 mM EDTA pH 8.3, 100 mM NaCl, 1% SDS). 50µl of proteinase K (10mg/ml) and 1µl of glycogen (20mg/ml) were added followed by overnight incubation at 37ºC. After this time the degraded proteins were precipitated by adding 340µl of 5M NaCl and centrifuging at maximum speed during 15 minutes. The supernatant was collected and 450µl of 100% Isopropanol was added. The isolated DNA was then transferred to a new eppendorf tube and washed with 75% Ethanol. The DNA was resuspended en 10µl of DNase free water. For the subsequent DNA methylation analysis, DNA samples were bisulfite converted using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA) following the manufacturer's instructions.

3.2.4. DNA methylation profiling using universal bead arrays

The methylation status of bisulfite modified DNA of $CD19^{+}$ cells from monozygotic twins was analyzed using the Infinium Human Methylation450 Bead Chips (Illumina, Inc.). This platform allows the interrogation of >485,000 methylation sites per sample at single-nucleotide resolution, comprising an average of 17 CpG sites per gene in the 99% of RefSeq genes. The 96% of CpG islands are covered, with additional coverage in CpG island shores and the regions flanking them. Samples were hybridized in the array following the manufacturer instructions. Each methylation data point is obtained from a combination of the Cy3 and Cy5 fluorescent intensities from the M (methylated) and U (unmethylated) alleles. Background intensity computed from a set of negative controls was subtracted from each data point. For further analysis we used the B-value

of each CpG, which is the ratio of the methylated probe intensity and the overall intensity (sum of methylated and unmethylated probe intensities). The Beta value ranges from 0 (non-methylation) to 1 (total methylation).

3.2.5. Bioinformatic analysis and detection of differentially methylated CpGs

We estimated the β values following what was defined as the optimal 3-step pipeline by Marabita and colleagues (Marabita et al., 2013). For the differential methylation we transformed β values into M-values (as suggested by Du and colleagues) and used limma Bioconductor package for the differential methylation analysis (Du et al., 2010). Given the following observations (1) technical variability was similar to biological variability, (2) many DNA methylation changes we expected to be mild, and (3) the reduced number of samples (1 twin pair) we made a conservative selection of probes as candidate differentially methylated sites: (i) 10% or greater difference in the β-value DNA methylation between monozygotic twins, (ii) p-value smaller than 0.01 and (iii) in the three replicates the β-value estimated of one twin were always larger or always smaller than the other twin. This selection criterion returned a selection of 311 differentially methylated candidate sites. Those sites are not statistically significant if FDR or FWER is considered however they provide a robust initial candidate selection.

3.2.6. Gene Ontology Analysis

Genes Symbol of probes differentially methylated were uploaded to AmiGO database (Carbon et al., 2009). We performed Biological Process GO enrichment based on genes that shows Differential Methylation with Fold Change at least 2 points up or below for Fold Change in each probe. To filter non-enrichment GO terms we applied a threshold p value and adjusted p value of less than 0.05. The hypermethylated and hypomethylated genes were also analyzed with Ingenuity program and GO database. The hypermethylated genes were selected by their constantly presence in relevant B cell signaling pathways, by their implication in common regulatory networks and functions related with BCR signaling.

3.2.7. Bisulfite pyrosequencing

In order to validate the results obtained from the DNA methylation array, we selected seven relevant genes in CVID that showed differential methylation to perform the pyrosequencing analysis. We analyzed the samples of CD19+ cells from monozygotic twins and additionally we analyze a panel of 80 samples obtained from CVID patients and healthy donors with three different cell populations, Naïve, Unswitched and

Switched B cells. Biotinylated amplicons for each gene were generated by PCR using the HotStart Taq DNA polymerase PCR kit (Qiagen). Specific primers were designed using the PyroMark Assay Design Software (QIAGEN- version 2.0.01.15). Pyrosequencing reactions and quantification of DNA methylation were performed with Pyromark™ Q24 system (Qiagen). Results from bisulfite pyrosequencing are presented as a percentage of methylation.

3.2.8. Gene expression data analysis and overlapping between expression and DNA methylation data

We obtained expression array values from GSE17269 Affymetrix Human Genome U133 Plus 2.0 Array microarray (Rakhmanov et al. 2009) to compare expression data of B cells from CVID donors and healthy donors. To obtain differential expression values between diseases and control samples, we used Bioconductor package Affy functions to read Affymetrix samples with .CEL format. After that, Limma package was used to estimate the differential expression of each probe ID, using the same parameters estimated in original data study. Raw differential expression data was filtered by Fold Change lower/higher than 0.5/1.5 and p-value lower than 0.05. At the end, overlap of differentially methylated and differentially expressed gene values was performed using Gene Symbol as ID and Fold Change as numeric values.

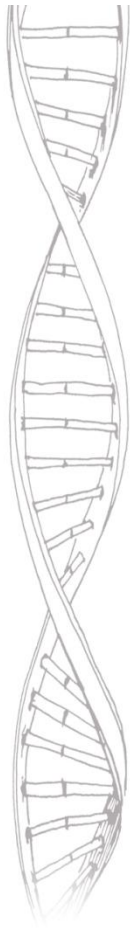
3.2.9. Statistical Analysis

The B cell subsets from CVID patients and healthy donors were compared using a student *t* test. *p* values less than 0.05 were considered statistically significance.

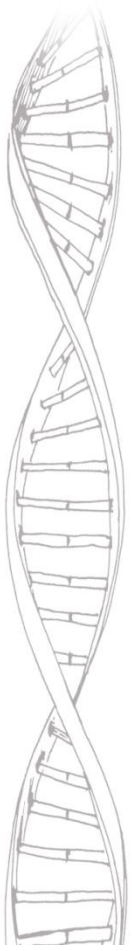
3.2.10. List of primers used in the CVID epigenetic study

Gene	Forward Primer	Reverse Primer	Pyrosequencing primer	Analyzed CpG
BCL2L1	[Btn]AGTAGTAAAGTAAGYGTTGAGGGAGGTAGG	CACAATACRACCCCAATTTACCCCATCC	AATATCAAATCACTAAATACCC	cg13989999
TCF3	[Btn]TATAGGTTTTYGAGGGATTATAGTTGGT	ATCAAAAACCRACCTCTCAAATCACTT	CCACCCCCCATACCC	cg26615224
PIK3CD	[Btn]ATAAGGATTGTTTYGGTGTTGTTATTGTA	AACTCAAATCCAACCTACTAACTATTC	ACTAATACATCTATCCATTACA	cg03265564
RPS6KB2	GGGGTTTGAGGTTTGTGGGATTA	[Btn]AACCCTCCTTAATACTATCCAAAACA	AGGTTTGTGGGATTAG	cg03559915
KCNN4	ATGTTGTTTTGTGTGGTTAGAATAT	[Btn]AAAAACATACCTATAATAACCCCCAATC	ATTATTATTGCRAGTATTTGTG	cg26890181
KCNC4	GTTYGTAGTATTGTGGGGTGGTG	[Btn]ATCAACTCAATTCACTTTCCATTTATA	GTAGTGYGTTATGATAGGAGTAAAA	cg26189021
CORO1B / PTPRCAP	[Btn]AGTTTTGGAGGTTGAGGAGTG	CCCAACACCACCCCTACCTA	AAATCCCRCTACTTACTAAACAC	cg12044599
	[Btn]AGTTTTGGAGGTTGAGGAGTG	ACCAAAACCACACCTAAAACTC	CATAACRAACCRACTATCAAACAAC	cg23468927

[Btn] indicates Biotinylated primers



RESULTS



4. RESULTS

The results of this thesis are presented in two separate sections that correspond to the two studies addressing the analysis of epigenetic mechanisms in relation with two primary immunodeficiencies: HIGM and CVID. The first study focused on the role of AID in the ability to influence epigenetic changes at the histone modification and DNA methylation level. Given that AID is the main altered molecule in HIGM2, after the characterization of the wild type (WT) AID effect on epigenetic mechanisms, it was analyzed the role of six HIGM-associated AID mutants in order to detect mutation-associated epigenetic effects that might contribute to the pathogenesis of HIGM. The second study, focusing on CVID, was centered in the identification of DNA methylation changes in CVID by using a pair of monozygotic discordant twins for CVID, followed by the determination of the DNA methylation levels of three B cell subsets in a cohort of CVID patients and healthy individuals.

4.1. AID ability to target epigenetic changes: impact on Hyper-IgM Syndrome

The starting point of this part of the thesis focused on establishing the two cell systems used to analyze the AID effects. After the verification of the functioning of the cell systems, we started the characterization of different epigenetic marks at a candidate DNA sequence level, by analyzing the *IGH* locus (the main target of AID), and at the global level, by using high throughput approaches. Following the identification of changes at some histone marks, we characterized interactions of putative relevant enzymes with the AID WT form followed by analysis with AID HIGM mutants to determine possible aberrant effects.

4.1.1. AID preferentially associates with heterochromatin

To investigate the potential impact of AID in the epigenetic regulation of its cognate sites in B cells as well as how these changes may be impaired in HIGM individuals, we generated an inducible retroviral system to express AID WT (Figure R1A) and a set of AID mutant forms occurring in HIGM. Two cell types were chosen, Jiyoye B cells and Hela cells, for the experimental design. Jiyoye cells, which are a Burkitt's lymphoma derived B cell line, provide a B cell context in which AID is not expressed. On the other hand, Hela cells are ideal to test the cell distribution, given the convenience of adherent cells for immunofluorescence. In addition, this cell type has been used for many biochemical studies of human AID. For the inducible system, Jiyoye and Hela cells were infected with the RetroX-Tet-ON advanced vector with geneticin resistance followed by transduction with the pRetroX-Tight-Pur vector encoding AID tagged with

HA at its C-terminal region. Expression of AID was achieved with the addition of doxycycline (Figure R1B). To obtain accumulation of AID in the cell nucleus, nuclear export was inhibited by the addition of LMB and accordingly, it was observed the nuclear accumulation of AID-HA after LMB treatment in both Hela and Jiyoye cells. (Figure R1C).

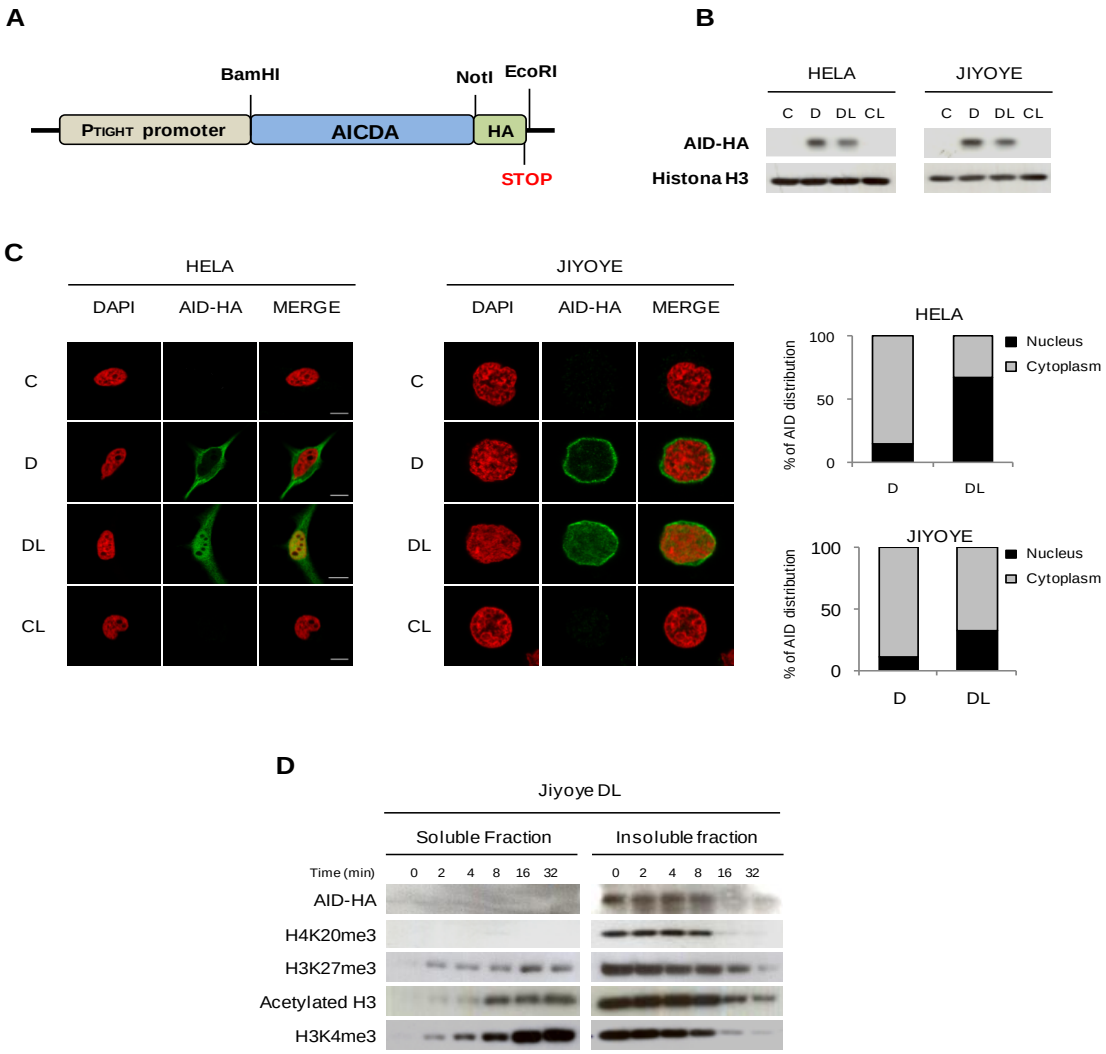


Figure R1. An inducible system for AID expression and accumulation in the nucleus. (A) HA-tagged AID retroviral construct (B) AID expression is induced following doxycycline treatment and retained in the nucleus under additional treatment with leptomycin B, which specifically inhibits nuclear export. Anti-HA antibody was used to detect AID expression and histone H3 was used as a loading control (C) Immunolocalization of AID in both Hela and Jiyoye cells infected with the retroviral inducible system after doxycycline treatment and leptomycin B treatment. Nuclear DNA was counterstained with DAPI (red). (D) Association of AID to heterochromatin. Time course digestion of HA-tagged AID Jiyoye cells nuclei with DNase I, where the soluble fraction mostly contains the euchromatic marks (as demonstrated by the appearance of H3Ac and H3K4me3), whereas insoluble fraction contains the heterochromatic marks(as shown by the progressive decrease of H4K20me3 and H3K27me3).

To discriminate whether AID is associated with heterochromatin or euchromatin, we performed a DNase I digestion-based assay to compare the ability of AID to be released from chromatin in comparison to euchromatic (H3Ac and H3K4me3) and heterochromatic (H3K27me3 and H4K20me3) histone modifications. Using this strategy, we observed that euchromatic histone modifications were released following digestion and are present in the soluble fraction at around 2 min of DNase I digestion, whereas the heterochromatic mark, H4K20me3, was absent in the soluble fraction and appeared in the insoluble fraction. Moreover, H3K27me3 is a histone modification characteristic of facultative heterochromatin, and appeared in both, soluble and insoluble fractions, displaying a similar pattern to that of euchromatic marks. In this assay, AID displayed a similar behavior to that observed for H4K20me3, suggesting that AID is mainly localized in the same chromatin compartment that H4K20me3 (Figure R1D).

4.1.2. AID expression does not alter the DNA methylation status of IgH locus but results in sequence specific and global increase in H4K20me3

As explained above, both CSR and SHM depend on AID activity. It has been described that AID binds to the S regions at the *IGH* locus to initiate productive CSR. These regions constitute the main target of AID and are very well defined to investigate a potential effect of AID on the epigenetic status of its associated sequences.

To explore potential changes in AID-bound regions we first analyzed the binding of AID to the Sμ region of the *IGH* locus in the two inducible cell models. ChIP assays revealed specific binding of AID to Sμ following induction of its expression and a further increase after inhibition of nuclear export of AID with LMB. Enrichment at the Sμ region was particularly high in Jiyoye B cells, although it was also observed in HeLa cells expressing ectopic AID. This binding did not occur at the Cμ sequence, as previously reported. Unlike Jiyoye cells, AID-expressing Hela cells presented detectable levels of AID in Cμ region, perhaps due to a mistargeting of AID caused by the overexpression. It is also likely that Hela cells, as a non-B cell, probably lack some of the AID interacting proteins or the chromatin structure that mediate proper AID targeting (Figure R2C).

To test the potential effect of AID on DNA methylation, we performed bisulfite pyrosequencing of specific CpG sites located within the Sμ and Cμ regions (Figure R2A). No changes were observed following induction, and therefore binding of AID (Figure R2B). The Sμ region presented very low levels of DNA methylation even before

AID overexpression in both, Jiyoye and Hela cell lines. This observation prevented the possibility of concluding the potential demethylating activity of AID to this region, given that it is already demethylated prior to AID binding. However, the analysis of PAX5, another reported AID target (Yamane et al., 2011) that has shown to be highly methylated in Jiyoye and Hela cells, did not display DNA methylation changes after AID overexpression either (Figure R2B). In fact, the high-throughput comparison of the DNA methylation profiles of control versus AID-expressing Jiyoye B cells using 450K Illumina arrays, which contain over 450.000 CpG sites, did not show significant changes associated with AID expression (Figure R3A). In addition, the DNA methylation analysis of repetitive elements, such as Alu repeats, using AUMA assays also showed absence of apparent changes (Figure R3B) discarding the existence of DNA demethylation events in association with AID binding, at least in our two inducible models.

Following the DNA methylation analysis, we investigated the possibility that AID associates with changes in histone modifications. ChIP assays were performed with three different histone modifications focusing on their association with the Sμ and Cμ regions. Specifically, we analyzed the association of H4K20me3, H3K27me3 and H3K4me3 with these two regions before and after AID induction. Previous studies have reported changes in H3K4me3 and H3K27me3 in association with the complex regulation of the *IGH* locus. It has been reported that H3K4me3 is enriched in the μ enhancer, S regions and Iγ promoters, and this enrichment increases in specific regions depending on the B cell activating stimuli. Meanwhile, it has been reported lower levels of H3K27me3 across the *IGH* locus, with enrichments at the 3' region of the *IGH* locus and at intergenic regions (Chowdhury et al., 2008). Moreover, as aforementioned, H4K20 methylation has been linked with CSR. *Suv4-20h-dn* B cells showed an accumulation of H4K20me1 (by losing the H4K20me2 and H4K20me3 marks) that impairs CSR and promotes chromosomal translocations. Thus, the three selected histone modifications (H4K20me3, H3K27me3 and H3K4me3) have been shown to be relevant for AID mediated events, although, there are no reports indicating a direct association among AID and modulations of these histone marks.

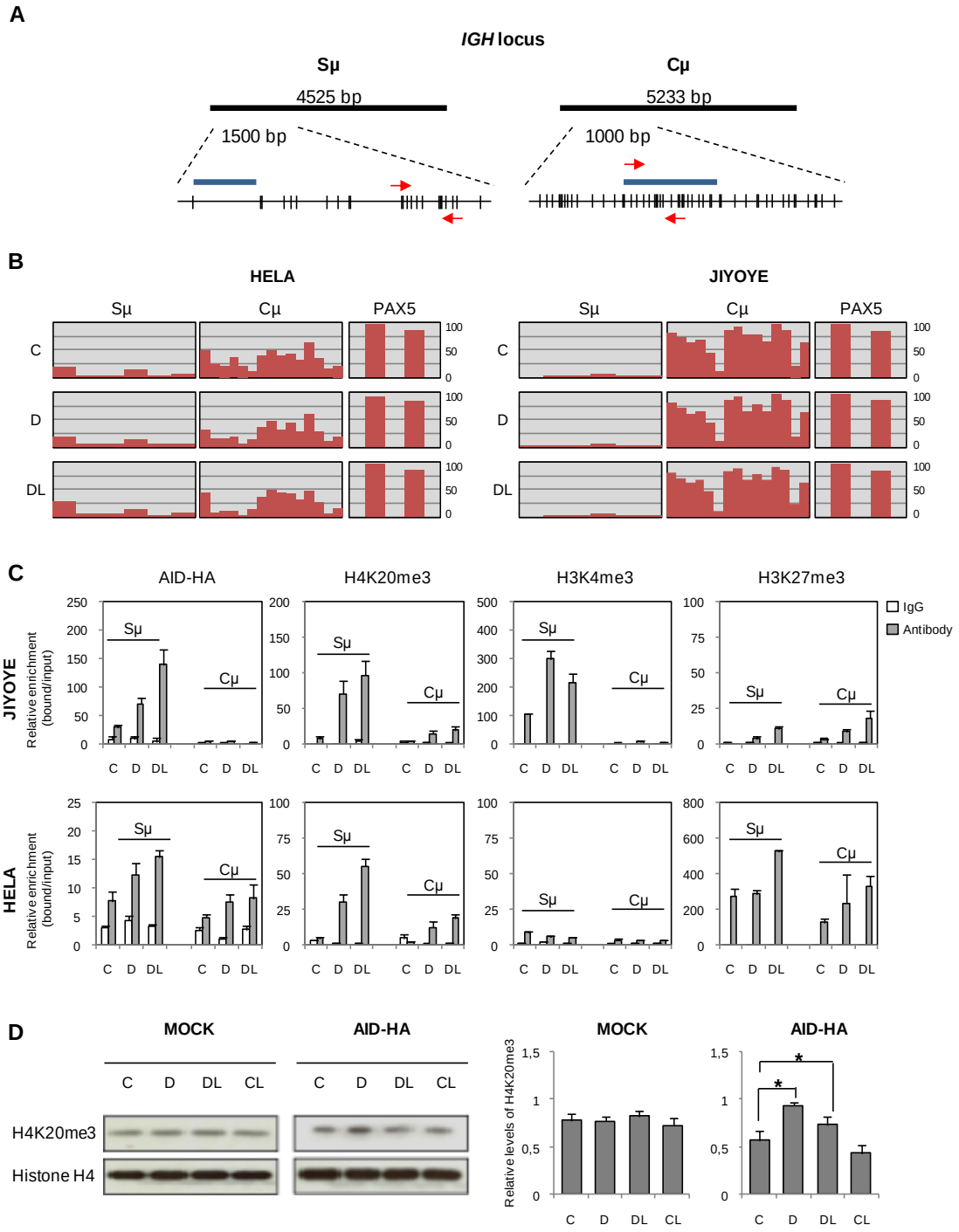


Figure R2. Effects of AID binding to the epigenetic status of the *IGH* locus. (A) Schematic representation showing the Sμ and Cμ regions of the *IGH* locus, the *bona fide* binding and non-binding sites for AID respectively. The scheme show the specific areas analyzed by pyrosequencing (red arrows) and ChIP assays (blue line). (B) Bisulfite pyrosequencing of Sμ, Cμ and PAX5 regions in both Hela and Jiyoye cells before and after induction with doxycycline, and following inhibition of nuclear export with leptomycin B. Percentage of DNA methylation is represented by red bars and each bar corresponds to a CpG site. (C) Changes in AID, H4K20me3, H3K4me3, H3K27me3 at the Sμ and Cμ regions using ChIP assays. For both Hela and Jiyoye cells we used uninduced, doxycycline, and doxycycline + leptomycin B conditions and IgG was used as negative control. (D) Effects of AID on the global content of H4K20me3 of Jiyoye cells as analyzed by western blot and quantitation of three independent experiments. Mock infected cells were used as an additional negative control.

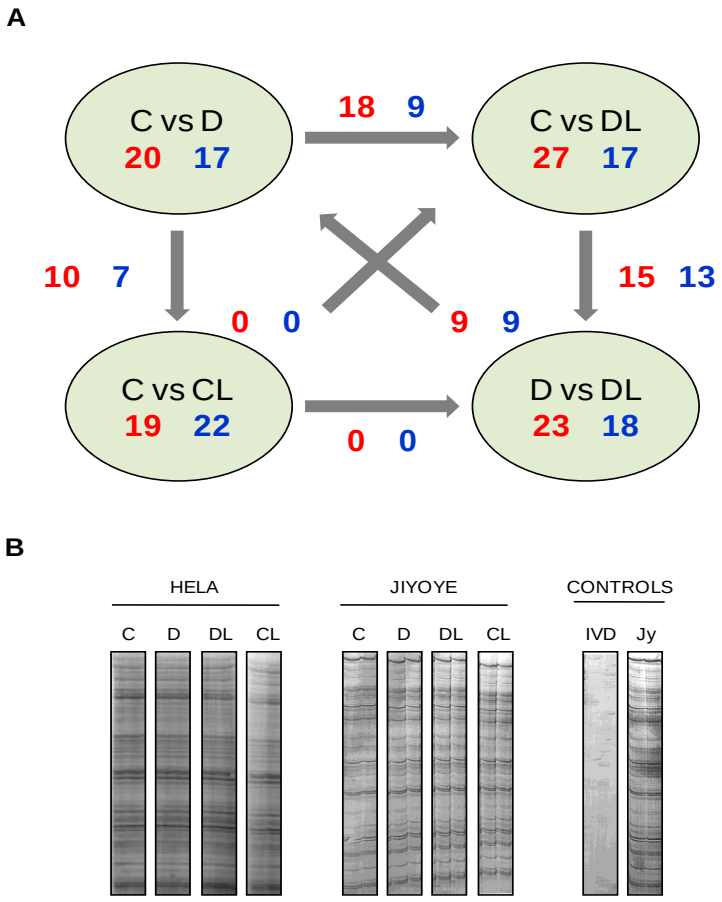


Figure R3. DNA methylation status of the IgH locus after AID overexpression. (A) Schematic representation of the results obtained from the DNA methylation profiling of the Jiyoye cell line with the Illumina 450K DNA methylation array. Green ovals represent the different comparisons of DNA methylation profiles. In red, it is indicated the number of significantly hypermethylated CpGs. In blue, the number of significantly hypomethylated CpGs. Red and blue numbers next to the arrows represent the hypomethylated or hypermethylated CpGs that coincide among the different comparisons. (B) Band pattering corresponding to DNA methylation analysis of Alu repeats, using AUMA assays. No significant differences were observed in the analyzed conditions neither in Hela nor in Jiyoye cells. The controls of the assay are constituted by DNA from Jiyoye cell line (without retroviral vectors) and the *in vitro* methylated DNA from the same cell line (IVD), which is the control of DNA methylation.

ChIP assays showed distinctive patterns for H3K4me3, H3K27me3 and H4K20me3 in B cells at the *IGH* locus. It has been described that the Sμ regions present an open chromatin conformation, whereas Cμ regions present a closed chromatin conformation in B cells. Accordingly, in the control condition (C), where AID is absent, H3K4me3 histone mark was enriched at Sμ region in Jiyoye B cells, whereas displayed low levels in Hela cells (a non B cell context) (Figure R2C). Moreover, the Cμ region presented low levels of H3K4me3 in both Hela and Jiyoye cells. After AID overexpression (D) it was observed a two fold increase in the H3K4me3 levels in Jiyoye cells but this increase was not observed in Hela cells. Furthermore, it was not observed in Jiyoye cells a further increase in H3K4me3 after LMB treatment, despite the AID nuclear levels was increased by two fold by LMB treatment. Thus, although it

seemed that AID caused an effect on H3K4me3 levels in Jiyoye cells when is overexpressed, this effect is not potentiated by AID nuclear accumulation after LMB treatment.

Moreover, the H3K27me3 histone mark showed lower levels compared with H3K4me3 as previously reported (Chowdhury et al., 2008) and did not show important variations after AID overexpression.

The most interesting observation from the ChIP assays was the specific enrichment of H4K20me3 at Sμ but not at Cμ regions, consistent with the specific binding of AID. This enrichment occurred in both Hela and Jiyoye cells, and appeared to be dependent on the AID binding. Jiyoye cells presented higher enrichment of H4K20me3 histone mark, according with the higher enrichment of AID observed in Jiyoye cells in comparison with Hela cells.

To determine whether changes in H4K20me3 also occurred at global levels, we performed western blot assays with Jiyoye cells before and after induction with doxycycline. We observed a discrete but significant increase in H4K20me3 following AID induction reinforcing the notion of a direct role of AID in inducing an increase of this modification at its binding sites (Figure R2D). This increase was not observed in control cells.

4.1.3. AID interacts with SUV4-20H family proteins, and modifies its recruitment to the chromatin

Changes in H4K20me3 following binding of AID suggest a relationship between AID and histone methyltransferases of the SUV420H family, which specifically target H4K20. As mentioned in the introduction, there are two genes in this family, SUV420H1 and SUV420H2, the first one producing two isoforms, SUV420H1_i1 and SUV420H1_i2 through alternative splicing. SUV420H1_i1 and SUV420H1_i2 are mainly involved in the addition of a methyl group to H4K20me1 to yield H4K20me2, whereas SUV420H2 is mainly involved in the transition from H4K20me2 to H4K20me3. To explore the potential physical interaction between these enzymes and AID, we performed immunoprecipitation experiments.

In these experiments, we used an alternative cell system, the 293F cell line, which allows the co-transfection of tagged SUV4-20H proteins and AID, leading to the co-expression of the C-t FLAG-tagged AID together with the HA-tagged version of each of the SUV4-20H enzymes (Figure R4A). These experiments revealed that AID

interacts with the three of them, as demonstrated both by immunoprecipitation of Flag-tagged AID and the reciprocal experiment immuprecipitating with HA-tagged SUV420H proteins (Figure R4B).

Then, we investigated the association of the SUV4-20H enzymes to the *IGH* Sμ and Cμ sites in relation with AID expression. ChIP experiments revealed that SUV4-20H1_i1 is associated with the IgH Sμ and Cμ regions only in the absence of AID (Figure R4C), and that co-transfection with AID impairs its association with the Sμ and Cμ sites. In the case of SUV4-20H1_i2, we observed a similar behavior, although this enzyme only associates with the Sμ site. Most importantly, we observed that SUV4-20H2 only associates with the Sμ site in the presence of AID, suggesting a primary role of AID in the replacement of these enzymes to chromatin in these regions (Figure R4C).

Since SUV4-20H2 is the main enzyme involved in the generation of H4K20me3, the recruitment of SUV4-20H2 to the Sμ region by AID might be responsible for the increase of H4K20me3 levels after AID expression and subsequent binding to this sequence.

To investigate the potential impact of AID in the accessibility at the *IGH* locus, a DNase I digestion under limiting conditions followed by amplification by quantitative PCR with primers for the Cμ and Sμ sequences was performed. The DNase-resistant fraction (size range ≥5000 bp) was cut at different time points and the above sequences were amplified. This strategy allowed monitoring the digestion dynamics, thus providing an estimation of the accessibility to DNase I. The analysis of the accessibility of D4Z4 and c-fos genes (Figure R5A) confirmed the validity of the accessibility assay, since it was possible to discriminate among highly compacted regions (therefore less accessible DNA regions) represented by D4Z4, and low compacted (therefore more accessible DNA regions) represented by c-fos. However, no differences in accessibility were observed at the Cμ sequences following AID expression (and subsequent binding) (Figure R5A).

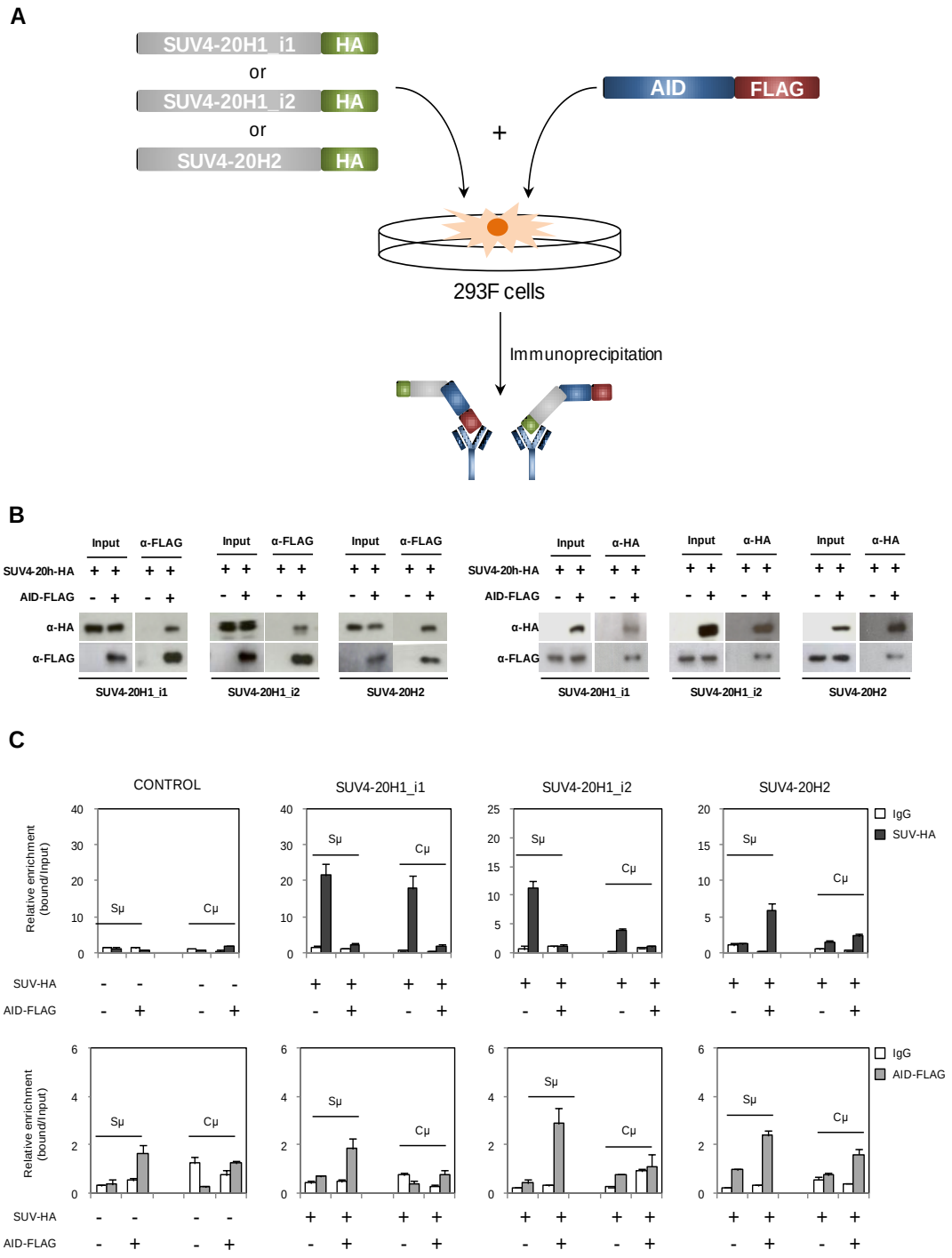


Figure R4. AID interacts with SUV4-20H enzymes and recruits them to IgH Sμ regions (A) Graphic representation of the imunoprecipitation experiment in 293F cells. Co-transfection of Flag-tagged AID and each of the HA-tagged SUV4-20H enzymes: SUV420H1_i1, SUV420H1_i2 and SUV4-20H2 (B) Immunoprecipitation. Immunopreciptated material with anti-HA was blotted and visualized with anti-FLAG to determine its potential interaction with AID. The reverse experiment was also performed, using anti-FLAG to immunoprecipitate AID and anti-HA was used to test he interaction with each of the SUV4-20H enzymes. AID and SUV4-20H enzymes were also transfected alone as controls. (C) ChIP assays demonstrating the association of SUV4-20H enzymes to the Sμ and Cμ regions in relation with AID binding.

We also checked whether AID association and binding to the Sμ region has an effect on the levels of the transcript produced by the transcription start site (TSS) related to this site, the μGLT. In B cells the μGLT is constitutively expressed and increases its levels after B cell activation. (Figure R5B). Nevertheless, the solely AID overexpression or its nuclear accumulation was not able to induce any changes in the expression of μGLT. Therefore the association of AID with the SUV4-20H enzymes and the increase of H4K20me3 after AID overexpression and nuclear accumulation did not appear to be associated with changes in the accesibility of Sμ region, nor with changes in the expresion of μGLT. However, this does not discard the potential effect of H4K20me3 in CSR.

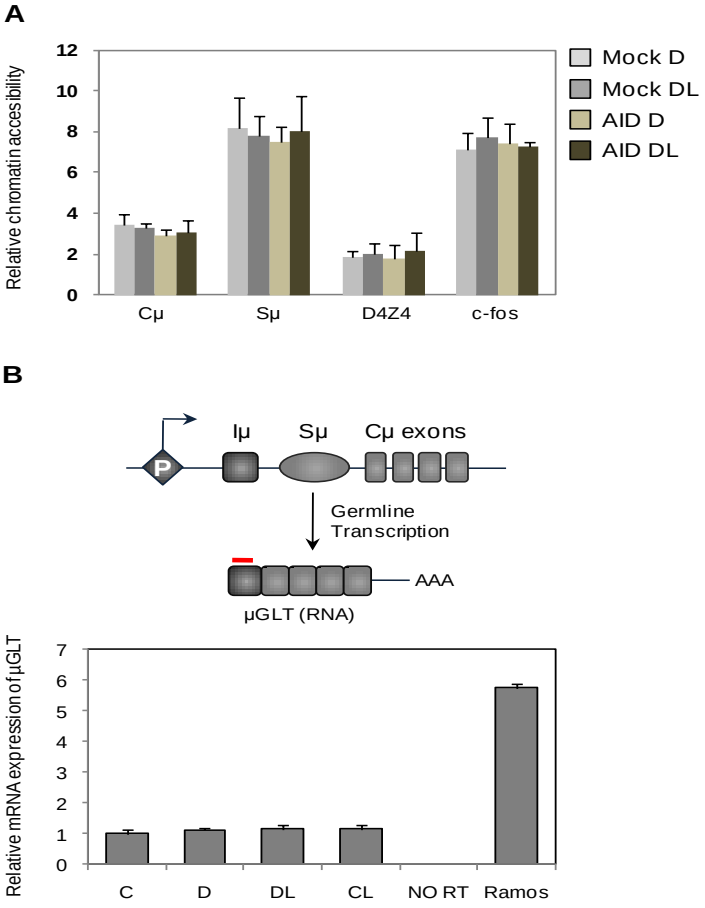


Figure R5. Potential role of AID in the accesibility and expression at the IGH locus (A) DNA accessibility assay. Amplification by quantitative PCR of Sμ, Cμ, D4Z4 and c-fos regions after DNAse I digestion. Accesibility is represented as the ratio among the final reaction point (30 minutes) and the initial reaction point (1 minute). It was possible to detect variations in DNA accessibility levels as demonstrated by the differences detected among D4Z4 and c-fos regions, but it was not detected variations in DNA accessibility after AID overexpression and its nuclear accumulation. This experiment was carried out with Jiyoye cell line transduced with empty vector (Mock) or with AID encoding vector. The DNA accessibility was assessed in this cell lines after treatment with doxycycline (D) or doxycycline + LMB (DL). (B) μGLT expression after AID overexpression and nuclear accumulation. The μGLT expression was normalized to the RPL38 expression. It was not observed any changes in μGLT expression after AID overexpression and its nuclear accumulation. Ramos cell line was used as a positive control of μGLT expression, whereas Jiyoye sample processed without retrotranscriptase was used as a control of genomic DNA contamination.

4.1.4. Hyper-IgM Syndrome AID mutant forms display different effects on nuclear localization and export

The immunoprecipitation experiments demonstrated a novel interaction between AID and SUV4-20H enzymes. We wondered whether these interactions are impaired in the context of HIGM syndrome mutations. These experiments not only help to understand the functional implications of these mutations but also can contribute to dissect the role of different domains in establishing these interactions. To this end we generated a version of the AID inducible system in Hela and Jiyoye cells for several HIGM forms, including AID R24W, W68X, C87R, M139V, R174S and R190X (Figure R6A). R24W AID is mutated in the nuclear localization signal (NLS), W68X and C87R are respectively a nonsense and a missense mutation at the catalytic domain of AID. W68X produces severely truncated protein (Figure R6B). Both M139V and R174S are missense mutations at the APOBEC-like domain. Finally, we also tested R190X AID, a nonsense mutant form that generates a shorter protein without part of the nuclear export signal (Figure R6B). As aforementioned, the C-terminal region containing the NES is critical to CSR probably by the recruitment of proteins involved in the AID targeting during this process. Furthermore, recent findings suggest that the C-terminal region of AID mediates the recruitment of DDR factors and mutants lacking this region display a defective end joining during CSR (Zahn et al., 2014).

To test the AID mutants, their cellular localization was investigated by immunofluorescence and confocal microscopy using a rabbit monoclonal anti-HA antibody. As aforementioned, WT AID is expressed following treatment with doxycycline and is accumulated in the nucleus after LMB treatment both in Hela and Jiyoye cells. The R24W, C87R, M139V AID mutants had a similar behavior than WT AID, however the two truncated forms of AID, W68X and R190X, where the nuclear export signal is totally or partially missing, displayed accumulation in the nucleus without the addition of LMB and therefore a constitutively nuclear localization (Figure R6C).

Like for Hela cells, missense AID mutations R24W, C87R, M139V did not alter the AID subcellular localization in Jiyoye cells, and the nonsense R190X was constitutively nuclear, therefore its localization was unaffected by LMB treatment. The truncated W68X mutant was detectable by western blot, however, it was not possible to detect its constitutively nuclear localization by immunofluorescence. This truncated form of AID showed to be less stable than the other AID mutants, which was also suggested by the western blots experiments.

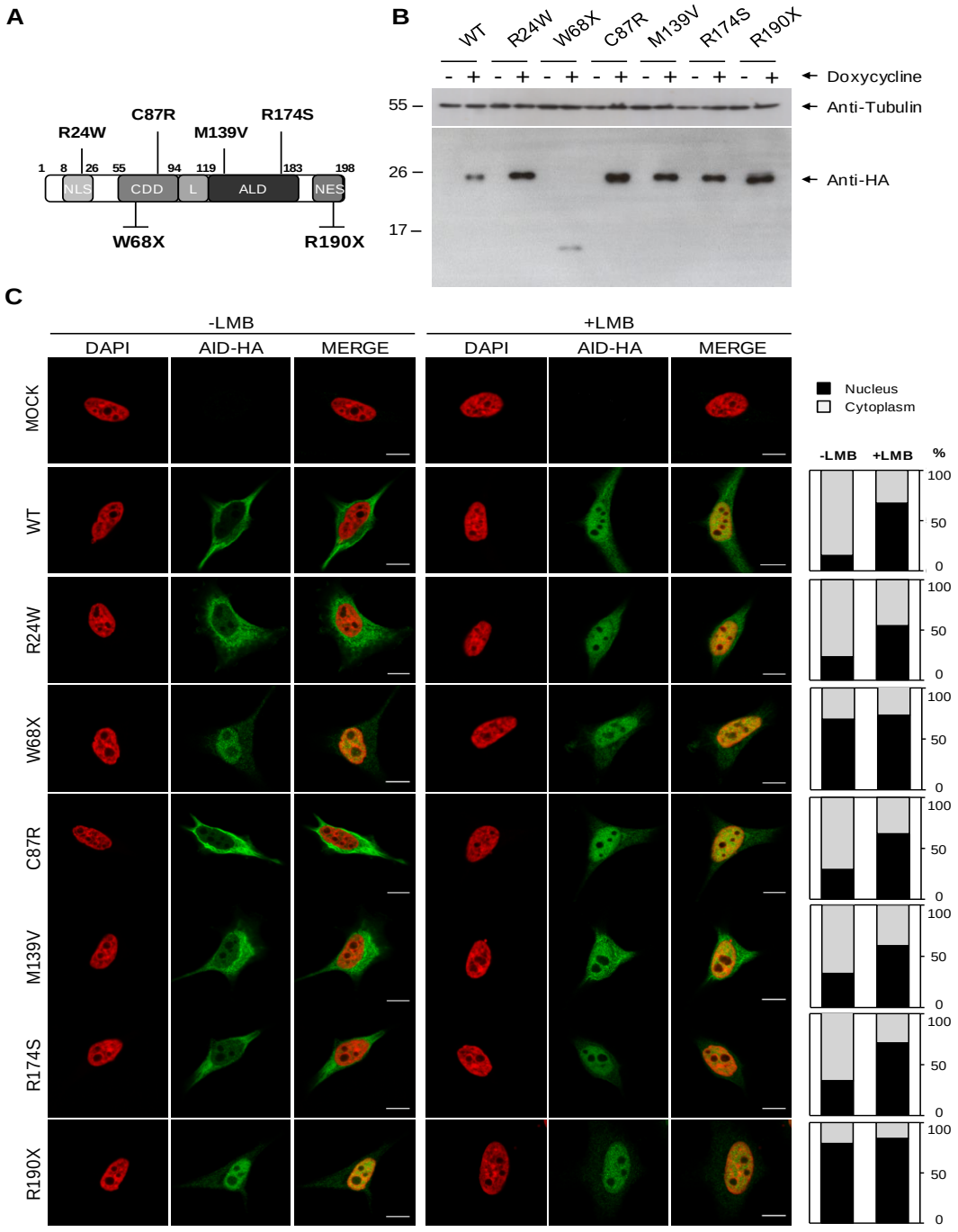


Figure R6. Subcellular distribution of AID and HIGM mutants forms in Hela cell line. (A) Primary structure of AID. In the upper part of the scheme, the four missense mutations related with HIGM selected to our study are indicated.In the lower part of the scheme, the two selected nonsense HIGM mutations are indicated. NLS indicates nuclear-localization signal; CDD, cytidine deaminase domain; L, linker region; ALD, APOBEC-like domain; NES, nuclear-export sequence. (B) Western blot image showing the inducible expression of AID WT and the different HIGM mutants, before and after the treatment with Doxycycline (Doxy) 500ng/ml during 48 hours. (C) Representative confocal images showing the subcellular localization of C-terminally Hemagglutin (HA) tagged human AID in inducible Hela cells. When the nuclear export was inhibited with Leptomycin B (LMB) 50ng/ml during 2 hours, most of AID translocates from the cytoplasm to the nucleus. Protein products of missense HIGM mutations showed similar response to AID WT after LMB, while truncated forms of AID lacking NES, were constitutively nuclear. Scale bar: 10μm.

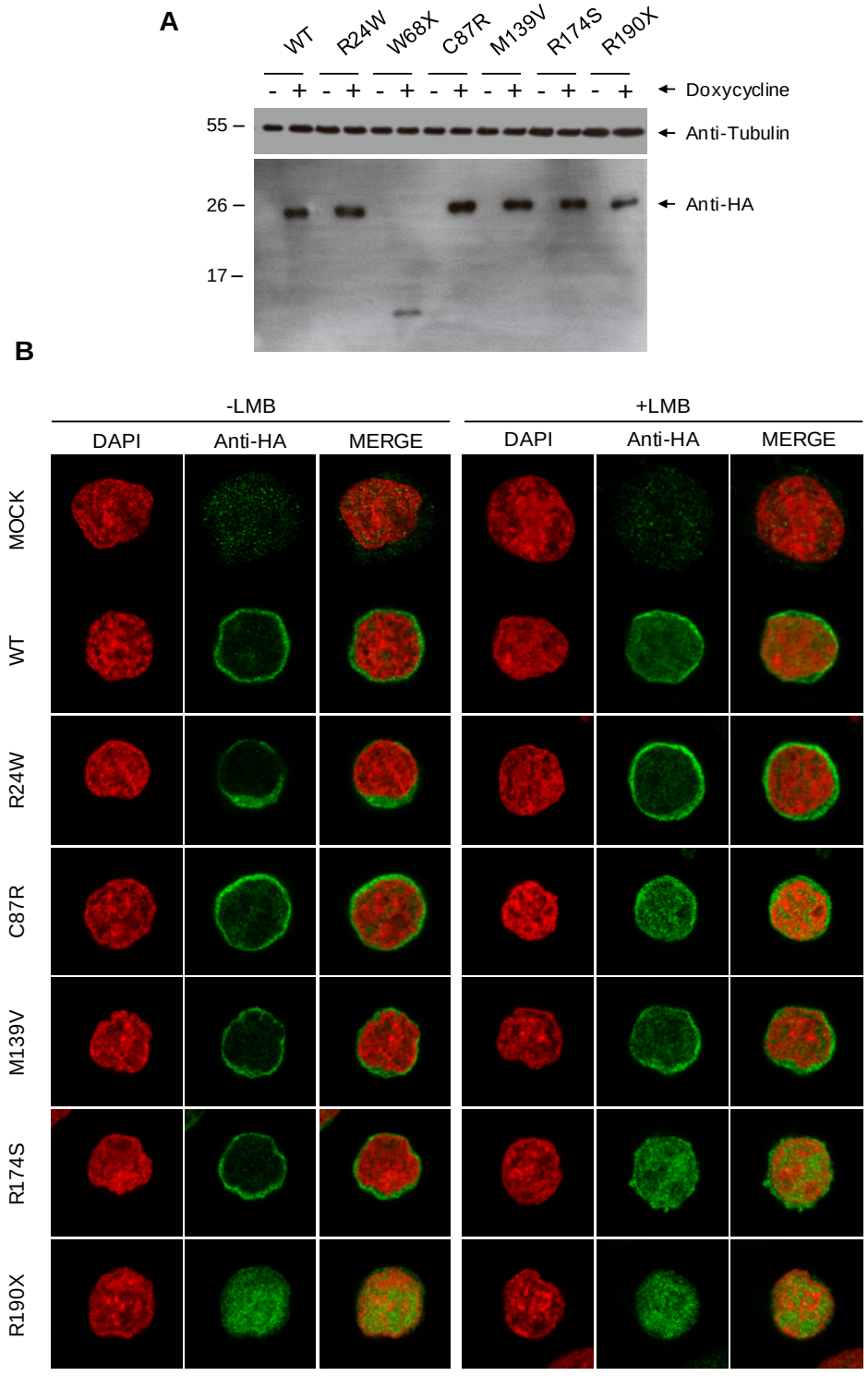


Figure R7. Subcellular distribution of AID and HIGM mutants forms in Jiyoye cell line (A) Western blot showing the inducible expression of AID WT and the different HIGM mutants, before and after the treatment with Doxycycline (Doxy) 500ng/ml during 24 hours. (B) Representative confocal images showing the subcellular localization of C-terminally Hemagglutin (HA) tagged human AID in inducible Hela cells. When the nuclear export was inhibited with Leptomycin B (LMB) 50ng/ml during 2 hours, most of AID translocates from the cytoplasm to the nucleus. Protein products of missense HIGM mutations showed similar response to AID WT after LMB, while truncated forms of AID lacking NES, were constitutively nuclear.

4.1.5. Hyper-IgM Syndrome AID mutants differentially interact with SUV4-20H enzymes

The ability of HIGM AID mutants to interact with SUV420H enzymes was then tested. Co-immunoprecipitation experiments with the anti-FLAG antibody in cells co-transfected with HA-tagged AID and each of the three SUV420H enzymes showed that WT AID and all the mutant AID forms interact at similar extent with SUV4-20H1_i1, with the exception of R24W mutant, that showed an impaired interaction when compared with WT AID. SUV4-20H1_i2 displayed a different interaction pattern with AID mutants, where the two AID truncated forms, W68X and R190X, displayed partial or total impairment of interaction. Moreover, AID mutants R24W, W68X, R174S and R190X showed an impaired interaction with SUV4-20H2. These results suggest that N-terminal region of AID could be important for the interaction of AID and SUV4-20H1_i1, the C-terminal region of AID could mediate the interaction between AID and SUV4-20H1_i2, whereas both regions of AID, N-terminal and C-terminal, could be involved in the interaction between AID and SUV4-20H2 (Figure R8).

Then, we investigated the association of SUV4-20H2 to the Sμ sequences depending on the form of AID used in co-transfection experiments (AID WT or W68X). We also included H4K20me3 in these experiments to test the effects on this mark in association with AID mutant form. In agreement with previous ChIP experiments, we observed that SUV4-20H2 was not enriched in the Sμ region, and this enzyme was recruited to this sequence when was co-expressed with AID WT, showing high levels of enrichment. When SUV4-20H2 was co-expressed with the truncated mutant W68X showed a similar enrichment to the one observed when co-expression wild type AID, despite its decrease affinity. In agreement with this, increases in the H4K20me3 levels were unaffected and experienced a similar enrichment to the one observed when SUV4-20H2 was co-expressed with AID WT (Figure R8). Although we cannot provide a satisfactory explanation for the unaffected recruitment of SUV4-20H2 and H4K20me3 levels when SUV4-20H2 and AID W68X were co-expressed, we consider that differences observed in the immunoprecipitation experiments could be related to functional alterations of the interaction between AID and SUV4-20H enzymes. To confirm this hypothesis, it would be necessary to analyze additional mutant AID forms and the recruitment of SUV4-20H1_i1 and SUV4-20H1_i2 to Sμ region and other AID target sequences.

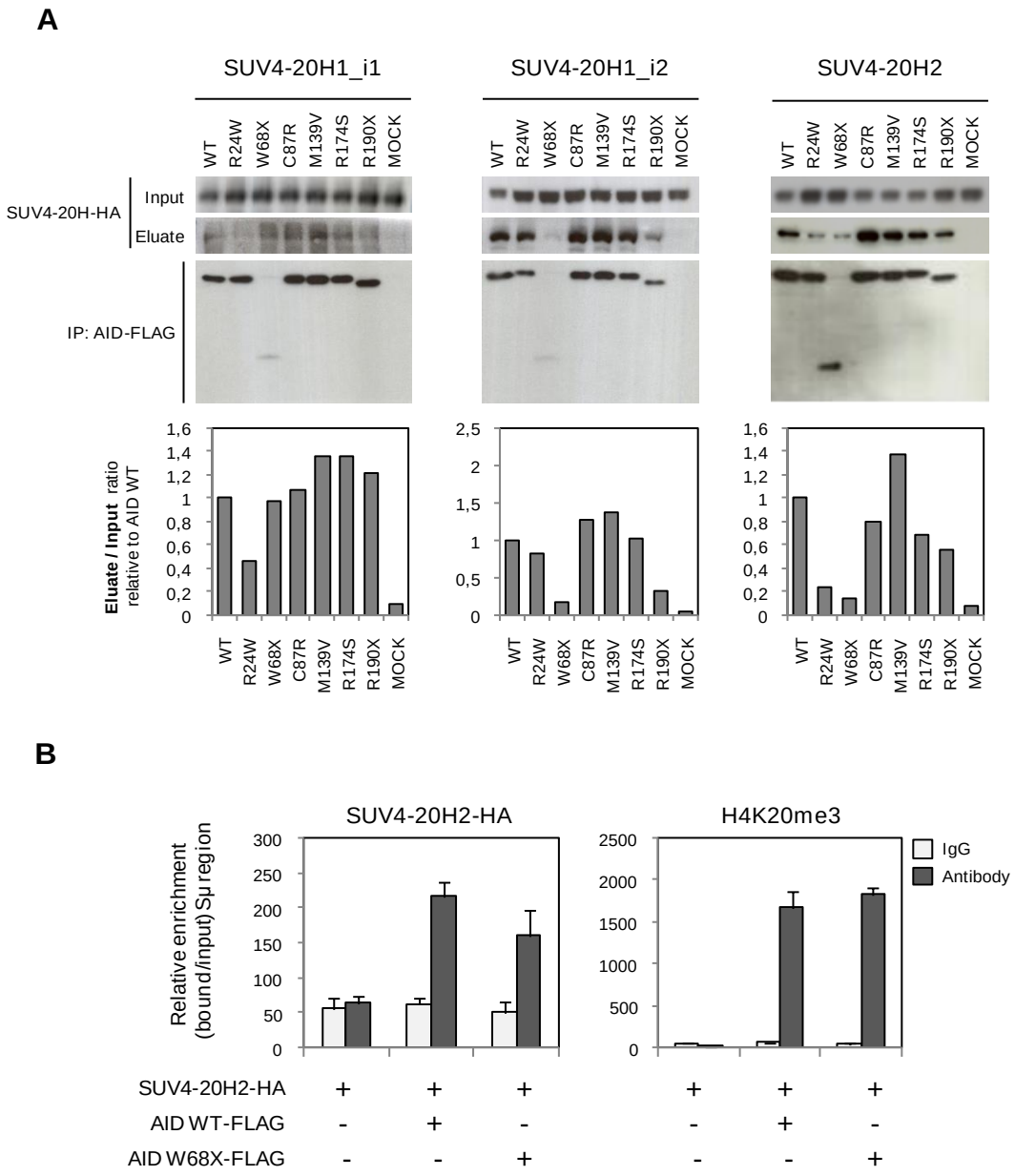


Figure R8. Subcellular distribution of AID and HIGM mutants forms in Jiyoye cell line (A) Co-immunoprecipitation of AID WT and HIGM mutants with the three SUV4-20H enzymes. All the HIGM mutants were also able to interact with the three SUV4-20H enzymes, but the interaction was impaired in the co-immunoprecipitation experiments with the truncated forms of AID, which suggests that the C-terminal region of AID could be important to the interaction with SUV4-20H enzymes. (B) ChIP analysis to analyze the binding of SUV4-20H2 enzyme alone or when is co-expressed with AID WT or the truncated mutant W68X. Additionally we analyzed the levels of H4K20me3.

4.2. DNA methylation alterations in Common Variable Immunodeficiency

The study of the acquisition of epigenetic alterations in CVID had as a starting point the comparison of DNA methylation profiles of B cells from a pair of monozygotic twins discordant for CVID. After this analysis, once obtained a list of genes displaying DNA methylation differences between the CVID and the healthy sibling, we compared the

DNA methylation levels of three B cell subsets in a cohort of CVID patients and healthy donors.

4.2.1. Comparison of DNA methylation profiles of monozygotic twins discordant for CVID reveals hypermethylation of a large set of genes

To investigate the potential occurrence of DNA methylation alterations in B cells of CVID individuals, we performed a DNA methylation screening on total $CD19^+$ cells isolated from a pair of monozygotic twins discordant for CVID. One of the individuals had been diagnosed with archetypical CVID, with a virtual absence of switched memory B cells, whereas his healthy sibling only displayed a discrete deficit of IgA, with IgG that had been increasing with age to normal levels. For the analysis, we used bead arrays interrogating the DNA methylation status of > 450,000 CpG sites across the entire genome covering 99% of RefSeq genes. Samples were processed and analyzed in triplicate. We observed that array wide technical variability (three replicates per twin: variability within a twin) and biological variability (differences between twins) were similar. For this reason we designed a robust CpG criteria selection (see Bioinformatics analysis in Materials and Methods) in order to identify differentially methylated candidate regions that were profiled later replication in an independent replication cohort. By the selection procedure we identified 311 CpG sites that were differentially DNA methylated between the CVID individual and his healthy matching twin. Given the small range of changes, we considered a CpG to be differentially methylated when the difference in β was greater than 10%. Approximately 75% of these differentially methylated CpGs (230 out of 311) displayed higher methylation levels in the CVID individual than in his sibling, and the remaining 25% (81 CpGs) were less methylated in the CVID compared with the healthy sibling. The majority of these changes occurred in annotated genes (79.1% and 80.2% in the hypermethylated and hypomethylated sets respectively) (Figure R9A). Among the genes associated hypermethylated CpGs, 131 were located at gene bodies, 28 at the 3' UTR and 54 at gene promoters. In the hypomethylated set of CpGs, 26 CpGs located at gene bodies, 17 CpGs at 3' UTR, and 37 in promoter regions (Figure R9B).

We then performed GO analysis to determine whether the differentially methylated genes were associated with potentially relevant biological processes in CVID. We independently analyzed the lists of hypermethylated and hypomethylated genes. In the hypermethylated set of genes, there was enrichment of important GO categories such as immune system process (GO:0002376), intracellular signal transduction (GO:0035556), defense response (GO:0006952), positive regulation of

macromolecule biosynthetic process (GO:0010557), and positive regulation of cellular metabolic process (GO:0031325). We found 12 to 19% of the hypermethylated genes to be associated with these categories. A smaller percentage of hypermethylated genes was associated with other interesting categories in CVID such as regulation of lymphocyte chemotaxis (GO:1901623), or positive regulation of T cell chemotaxis (GO:0010820). In the hypomethylated group, the functional categories were less obviously relevant in the context of B cell biology. We found that 24% of genes were associated with cell adhesion (GO:0007155), 20% were in the cell-cell adhesion (GO:0098609) and homophilic cell adhesion (GO:0007156) categories. Around 3 to 5% of hypomethylated genes were associated with other categories, such as response to cAMP (GO:0051591), cell fate specification involved in pattern specification (GO:0060573), and cell fate specification (GO:0001708) (Figure R9C).

The hypermethylated genes (Table R1) were particularly interesting for their relevance in BCR signaling pathways such as: PI3K signaling in B lymphocytes, FcRIIB signaling in B lymphocytes, CD27 signaling, P38MAPK, CD40 signaling, NF-κB signaling, APRIL mediated signaling, B cell activating factor signaling, pathway of inositol phosphate compounds. All these genes have associated network functions that the programme identified as: cell death and survival, cell mediated immune response and cellular movement.

The GO database also reveals the hypermethylated genes implicated in lymphocyte chemotaxis (CCL5) and phospholipids translocation (KCNN4), and some other relevant common pathways such as class I PI3K signaling events and mTOR signaling pathway. In addition to the overall analysis of functional categories, inspection of individual genes made it possible to identify those essential for B cell biology and function from amongst the list of hypermethylated genes. The list of hypomethylated genes had less obvious candidates in terms of B cell biology. From the list of hypermethylated genes, we identified *TCF3*, *PIK3CD*, *KCNN4, BCL2L1*, and *RPS6KB2*, among others (Table R1). *PIK3CD* encodes the catalytic subunit delta of the phosphatidylinositol-4,5-bisphosphate 3-kinase, a molecule that regulates natural antibody production, marginal zone and B-1 B cell function, and autoantibody responses (Durand et al., 2009). Another interesting example is KCNN4, a molecule that is activated during B cell activation and helps to maintain elevated Ca2+ levels during signal transduction (Wulff et al. 2004). *BCL2L1* encodes an apoptotic inhibitor that plays a critical role in the germinal center (Tuscano et al., 1996).

We then compared the DNA methylation data with existing gene expression data obtained from B cells of CVID patients versus healthy donors (accession no. GSE17269) (Figure R9D) (Rakhmanov et al., 2009) . Despite the differences between this study focusing on a small cohort and also using the naïve CD19+ subset, we found a significant number of genes with differences in expression among those displaying differences in DNA methylation. This suggests that the methylation changes occurring in CVID B cells may affect the overall expression of the genes. (Figure R9D).

4.2.2. DNA methylation changes for B cell regulators in CVID are B cell specific

To confirm the results obtained using DNA methylation arrays, we performed bisulfite pyrosequencing of a selection of the aforementioned genes hypermethylated in B cells isolated from the same pair of twins discordant for CVID (Table R2). We focused on looking at the same CpG sites for which significant changes had been identified (Figure R10A). Despite the modest differences in DNA methylation quantitated using the initial high-throughput strategy, we confirmed by bisulfite pyrosequencing a robust increase of DNA methylation in the CVID sibling with respect to the healthy sibling (Figure R10B), with very similar values to the ones obtained using bead arrays (Figure R10C). We also tested The DNA methylation changes in $CD4^+$ cells (T lymphocytes) and $CD14^+$ cells (monocytes) from the same individuals (Figure R10B). In contrast with the robust increase in B cells from the CVID sibling with respect to the healthy sibling, we observed no differences or even changes in the opposite direction in T cells and monocytes, highlighting the specificity of the observed CVID-associated hypermethylation of these genes to B cells.

As mentioned earlier, CVID individuals have been reported to contain fewer switched memory cells. Our results showing DNA methylation differences between this CVID individual and his matching healthy sibling using total $CD19^+$ may just reflect changes in the proportion of different B cell subsets between these two individuals if the DNA methylation levels for these genes were different for naïve, unswitched and switched memory B cells. For this reason, it is essential to analyze DNA methylation using separate B cell subsets, to distinguish between the change in B cell subset proportions and the existence of *bona fide* changes in the DNA methylation status of these genes in specific B cell subpopulations.

Table R1. Relevant genes in B cell biology in the hypermethylated set of genes

Gene	CpG number	Localization	ΔBeta value	Relevance of gene products in B cell context
BCL2L1	cg13989999	Body	0.13	The longer isoform encoded by this gene, BCL2-X_L, is an apoptotic inhibitor and plays a critical role in regulating cell survival in the germinal center (GC). It has been reported that T lymphocytes from CVID patients are more sensitive to apoptosis due to a deregulation of this molecule
TCF3	cg26615224	Body	0.13	Important transcription factor that mainly acts as gene activator. Is required for normal T and B-cell development
PIK3CD	cg03265564	5´UTR	0.11	The p110δ catalytic subunit is mainly expressed in hematopoietic cells. This molecule mediate the chemokine-induced migration, BCR signaling, BCR-induced proliferation, and differentiation into Antibody-producing cells
RPS6KB2	cg03559915	Body	0.15	This kinase is a downstream signaling protein of mTOR pathway. S6K2 regulates cell growth and protein synthesis
KCNN4	cg26890181	TSS1500	0.11	Is upregulated during B cell activation and helps to maintain the enhanced cytosolic Ca^{+2} levels during signal transduction
KCNC4	cg26189021	Body	0.14	This K+ channel modulates the membrane potential directly by regulating the generation of cytosolic Ca2+ signals. KCNC4 channels are key regulators in the stress response of irradiated leukemia cells
CORO1B / PTPRCAP	cg12044599	Body / TSS1500	0.17	Coronin 2, the protein product of CORO1B gene, has been implicated in actin-based processes such as cell migration. Coronin1A, another member of coronin protein familiy, has been related to T cell survival and has been associated with T cell immunodeficiencies
	cg23468927	Body / TSS1500	0.15	The protein encoded by *PTPRCAP* is required for normal antigen-receptor signaling and function in lymphocytes. Its association with the protein phosphatase CD45, seems to modulate signal transduction by regulating the CD45-LCK interaction
XCL1	cg21872093	TSS1500	0.13	This chemokine is important in the regulation of T and B lymphocytes and neutrophil trafficking
WNT5A	cg24049183	Body	0.17	WNT5A protects isolated GC B cells from apoptosis by initiating the noncanonical b-catenin–independent signaling pathway
AKT3	cg24455383	Body	0.14	AKT, as the major effector downstream of phosphatidylinositol 3-kinase (PI3K) signaling pathway, plays a key role in peripheral B-cell maturation and survival
RPTOR	cg00701918	Body	0.15	This component of the mTORC1 complex is one of the main targets of AKT. Is required for proliferation of splenic B cells and promotes B cell responses to LPS in the absence of PI3K activation
IKBKE	cg26859016	TSS1500	0.11	IKBKE is a non-canonical IKK family member that plays an important role in the regulation of inflammatory signaling pathway. This kinase activates NF-κB and is able to activate AKT in an PI3K independent way.
STK11	cg08317252	Body	0.11	It has been reported that this master kinase is important for cessation of the GC reaction, plasma cell differentiation, and suppression of tumorigenesis
DUSP2	cg02431562	Body	0.13	This phosphatase is involved in the regulation of a number of MAP kinase-dependent physiological processes that occur during the proliferation and differentiation of hemopoietic cells
TNFRSF10A	cg23303108	TSS1500	0.11	Receptor of the proapoptotic protein TRAIL, that has been involved in the apoptosis necessary to eliminate primary plasma cells after the synthesis and secretion of large amounts of Antibodies
HDAC4	cg11231069	Body	0.12	Chromatin modifier that is recruited by BCL-6 transcriptional repressor to regulate lymphocyte function, survival, and differentiation
MTA3	cg06342490	TSS1500	0.15	MTA3 is a subunit of the transcriptional corepressor Mi-2/NuRD, with a prominent role in B cell fate determination trough its interaction with BCL-6

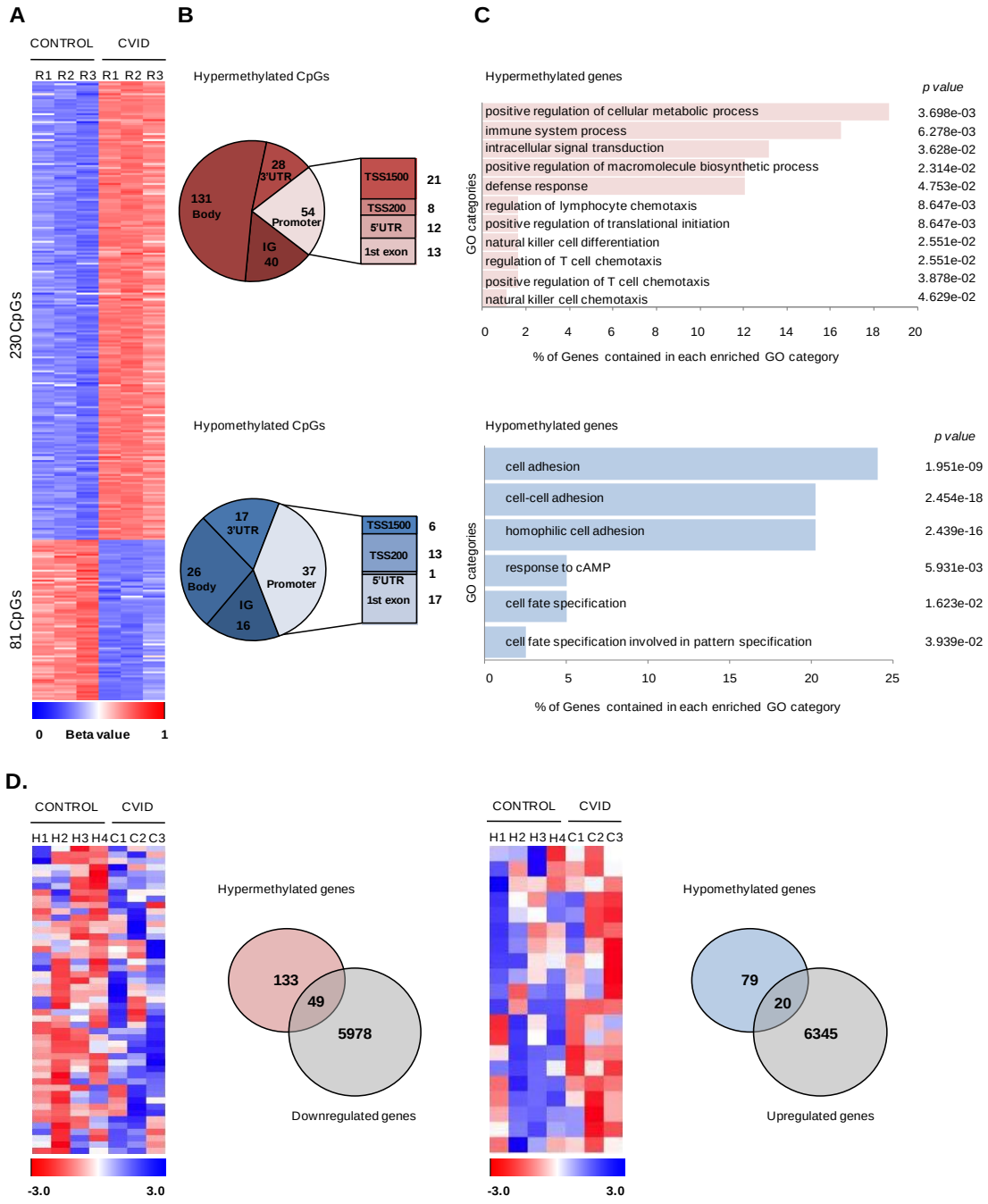


Figure R9. Comparison of DNA methylation profiles of B cells from monocygotic twins discordant for CVID. (A) Heatmap showing the differentially methylated CpGs between the B cells from twins discordant for CVID obtained through the DNA methylation array. A technical triplicate was analyzed for each sample (R1, R2, R3). The heatmap scale shows the B value range, where 1 indicates total methylation (red), and 0 the absence of methylation (blue). After comparison, B cells from CVID twin showed 230 hypermethylated CpGs, while 81 CpGs were hypomethylated (ΔB value ≥ 0.1) (B) Genomic distribution of differentially methylated CpGs. The analyzed CpGs could be associated with genes and localized at the 3' UTR region, at the gene body, or at the promoter region. At the promoter level, the CpGs could be localized within 1500bp of the transcription start site (TSS1500), within 200bp of the transcription start site (TSS200), at the 5' UTR or in the first exon (1st exon). The CpGs that are not associated with genes are considered intergenic. (C) Gene Ontology enrichment analysis of genes associated with differentially methylated CpGs. The bar charts show the most relevant and significantly enriched GO categories, the *p* values, and the percentage of hyper or hypomethylated genes that are contained in each category. (D) Heatmaps showing the expression differences between a small cohort of CVID and healthy individuals for hypermethylated genes (left panel) and hypomethylated genes (right panel). Beside of each heatmap, a Venn diagram shows the overlap between genes that are hypermethylated and genes that are downregulated in CVID with respect to healthy individuals, or the overlap between hypomethylated and genes that are upregulated in CVID with respect to healthy individuals.

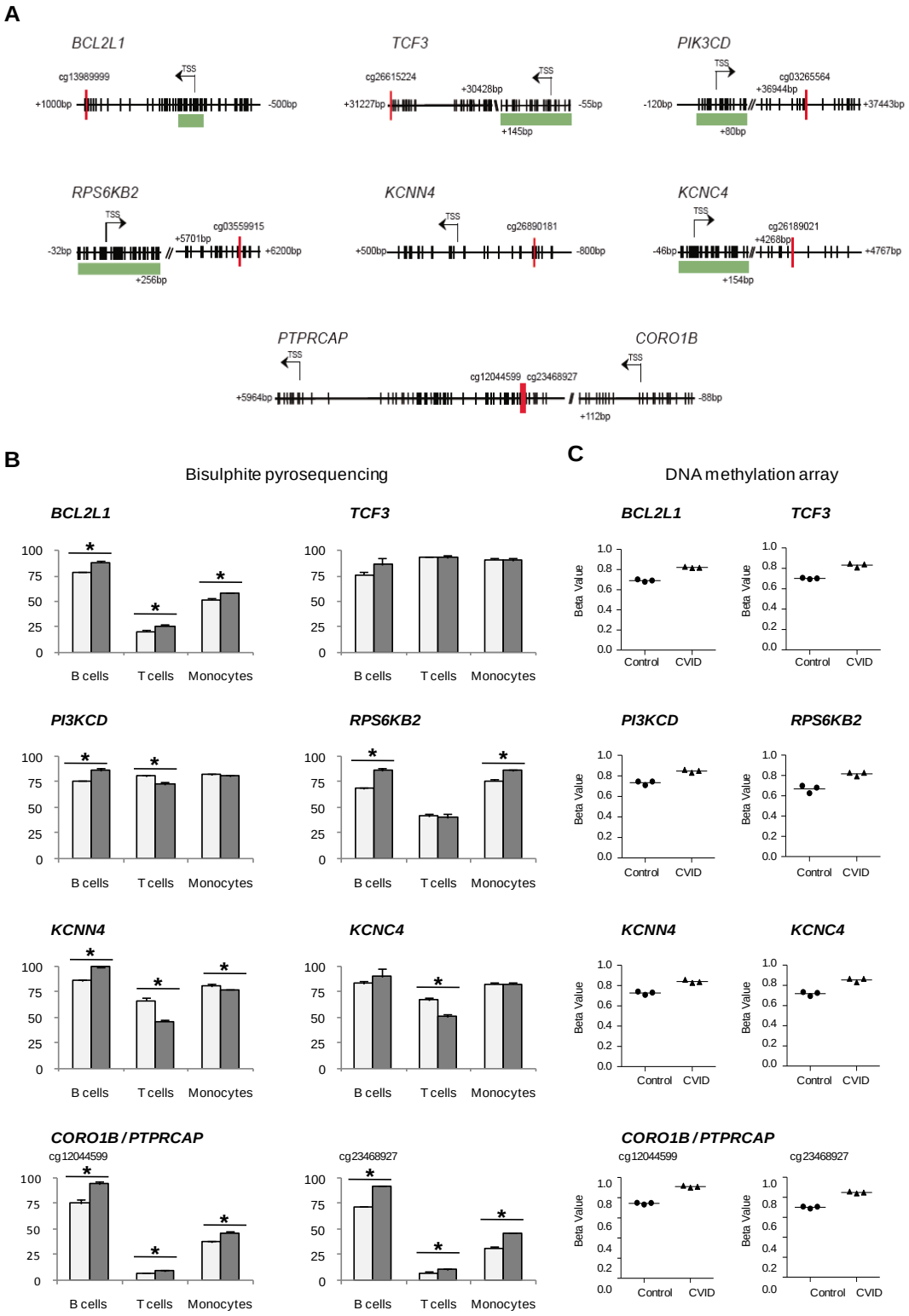


Figure R10. Validation of DNA methylation results by bisulfite pyrosequencing. (A) Schemes depicting a selection of genes indicating the CpG that it is differentially methylated (red line) and its relative location with respect to the TSS (arrow). Additional CpG sites are also represented with vertical black lines. CpG islands are represented with a green bar. (B) Bisulfite pyrosequencing of selected genes. In addition to $CD19^+$ B lymphocytes, $CD4^+$ T lymphocytes and $CD14^+$ cells (monocytes) were analyzed. Methylation levels are represented as a percentage (C) Beta values obtained in the DNA methylation array of the B cells from CVID twin show higher levels of DNA methylation (ΔB value ≥ 0.1) in the analyzed genes when compared with B cells from its healthy brother. Black filled circles indicate the B values triplicate in the control twin. Black filled triangles indicate the B values triplicate in the CVID twin. Black line indicates the average of B values in each condition.

Table R2. Selection of hypermethylated genes to the validation process

Gene	Full Name
BCL2L1	BCL2-like 1
TCF3	Transcription factor 3
PIK3CD	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit delta
RPS6KB2	Ribosomal protein S6 kinase, 70kDa, polypeptide 2
KCNN4	Potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4
KCNC4	Potassium voltage-gated channel, Shaw-related subfamily, member 4
CORO1B	Coronin, actin binding protein, 1B
PTPRCAP	Protein tyrosine phosphatase, receptor type, C-associated protein

4.2.3. The efficiency of demethylation of B cell regulators is partially impaired in the transition from naïve to memory cells in CVID patients

To address the aforementioned question, we therefore proceeded to analyze the DNA methylation levels in selected hypermethylated genes for three B cell subsets in a cohort of 12 healthy donors and 16 CVID patients. Specifically, we measured the methylation levels of all these genes in naïve (CD19+CD27-IgD+), unswitched (CD19+CD27+IgD+) and switched (CD19+CD27+IgD-) B lymphocytes. The relative numbers of these B cell subsets differed between CVID and healthy individuals. For an equivalent number of naïve cells, the unswitched B cell subset was present at similar levels in CVID and healthy individuals, but switched memory cells were much less frequent in CVID individuals (Figure R11A).

We then compared the methylation levels of the B cell subsets in the control group, observing a progressive loss of methylation from naïve cells to switched memory B cells (Figure R11B). For instance, *BCL2L1* displayed an average 79% methylation in naïve cells, 49% in unswitched memory cells and only 13% in switched memory cells (Table R3). Similar changes towards lower DNA methylation levels from naïve to switched memory cells were found for all the other genes in our list (Table R3), highlighting the close association between the differentiation of naïve cells to memory cells and the loss of methylation in these relevant genes. These findings are compatible with the increased expression levels of these genes and the function of their products in memory cells. These results are also in agreement with the progressive loss of methylation and gain of expression during B cell differentiation (Lai et al., 2013).

Notably, when we analyzed the DNA methylation levels of these genes in the three B cell subsets in CVID patients, we observed impairment in the loss of methylation as cells transition from naïve to switched memory B cells (Figure R11B). For instance, focusing again on *BCL2L1*, the average methylation levels were significantly higher for memory cells (both unswitched and switched) in CVID individuals than in healthy controls (Table R3). This behavior was generally observed for all genes (see Figure R11 and Table R3).

We observed that the levels of methylation for all genes in naïve B cells from CVID individuals were almost identical to those from healthy controls. However, the average levels for all these genes in unswitched and switched B cells were higher for CVID than for control individuals. In general, unswitched and switched memory B cells tended to be hypermethylated relative to controls. This finding suggests impaired loss of methylation for these genes during B cell differentiation in CVID patients.

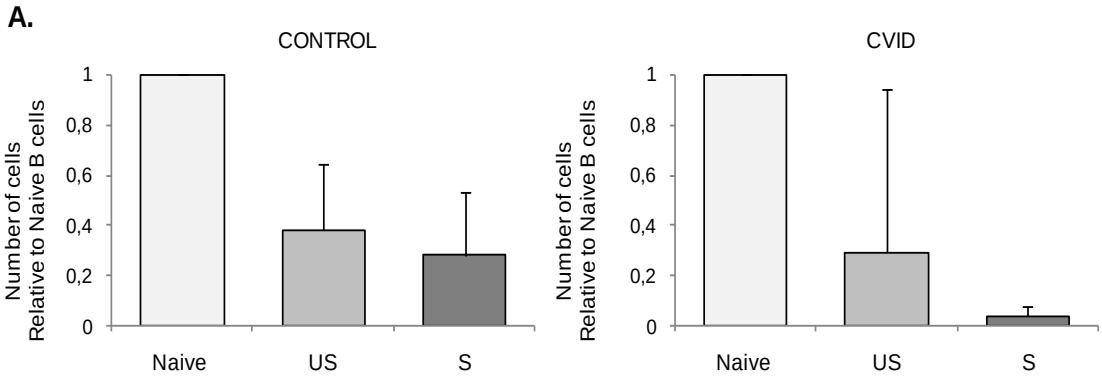


Figure R11. Comparison of DNA methylation levels of selected genes in different B cell subsets in a cohort of CVID patients and healthy donors. (A) Graph showing the relative amount of cells in each B cell subset (naïve, unswitched (US), switched (S) memory cells) in CVID patients and healthy donors. The cell number is relative to the naïve B cell subset. (B) Box and whisker plots showing the percentage of DNA methylation obtained by pyrosequencing in the 8 validated genes. White boxes represent the DNA methylation levels in healthy donors. Gray boxes represent DNA methylation levels in CVID patients. The analyzed B cell subsets were Naïve, Unswitched, and Switched B cells. The *p* value is showed in the cases with a statistically significant difference

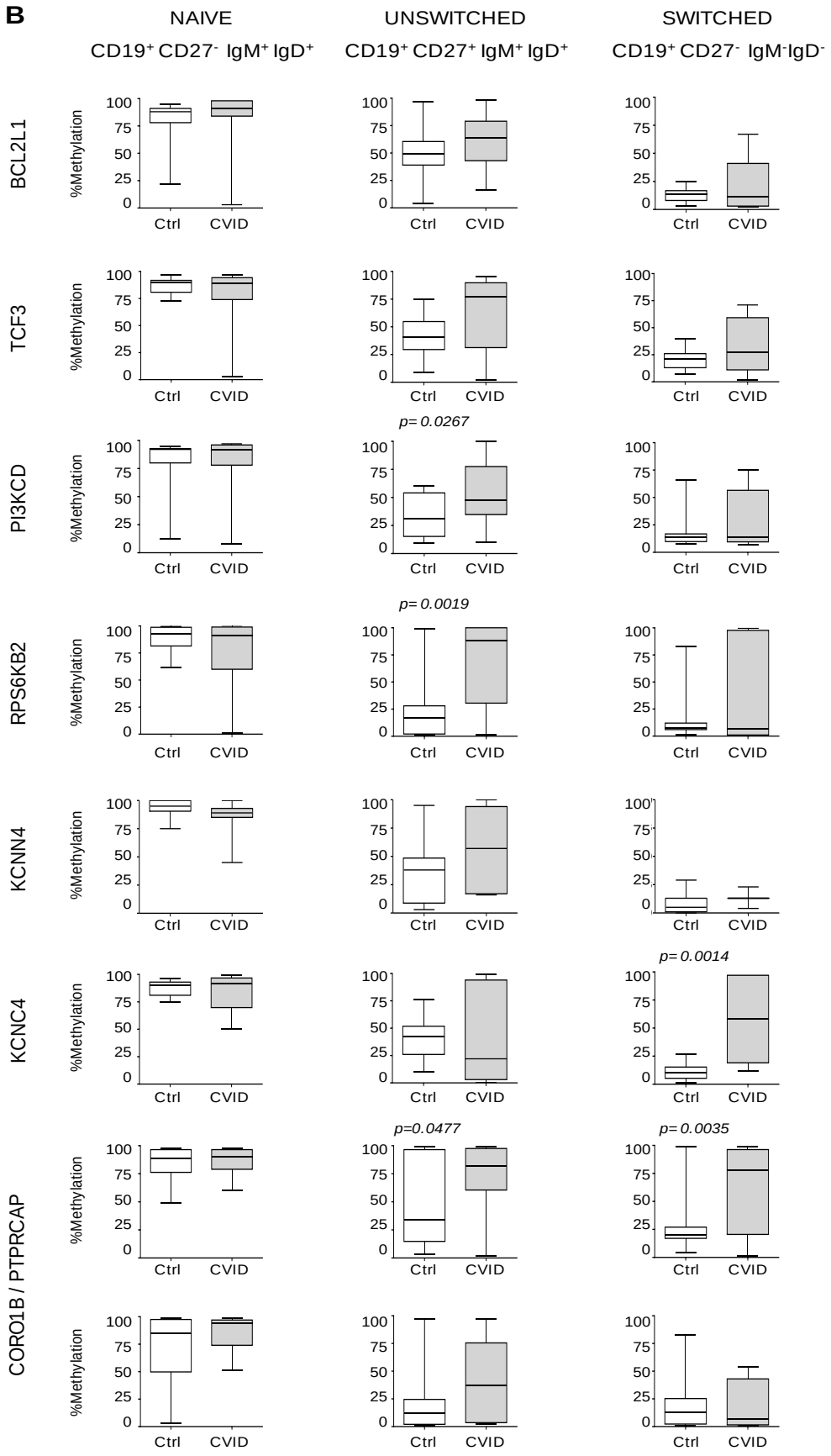
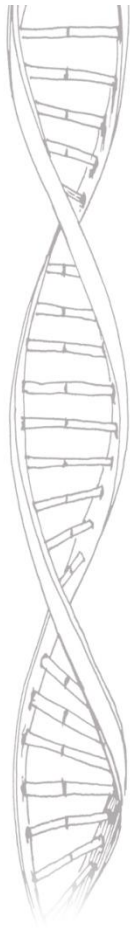


Figure R11. Comparison of DNA methylation levels of selected genes in different B cell subsets in a cohort of CVID patients and healthy donors. Continued from the previous page.

Table R3. Average of the percentage of DNA methylation obtained in the three B cell subsets from CVID patients and healthy donors

Gene		Naive		Unswitched		Switched	
		AVG %Meth	SD	AVG %Meth	SD	AVG%Meth	SD
BCL2L1	Control	**79**	22	**49**	25	**13**	6
	CVID	**80**	31	**62**	25	**21**	24
TCF3	Control	**87**	7	**42**	20	**21**	10
	CVID	**77**	29	**62**	32	**32**	25
PIK3CD	Control	**80**	23	**33**	18	**18**	16
	CVID	**82**	23	**54**	29	**31**	26
RPS6KB2	Control	**87**	13	**22**	28	**14**	22
	CVID	**76**	33	**66**	40	**35**	45
KCNN4	Control	**93**	8	**35**	28	**8**	9
	CVID	**86**	15	**53**	37	**11**	9
KCNC4	Control	**87**	7	**40**	20	**11**	8
	CVID	**84**	17	**41**	41	**59**	37
CORO1B / PTPRCAP	Control	**81**	18	**46**	38	**26**	25
	CVID	**87**	11	**72**	31	**64**	39
	Control	**81**	31	**21**	29	**18**	23
	CVID	**88**	15	**44**	40	**31**	38



DISCUSSION

5. DISCUSSION

The two studies contained in this doctoral thesis represent two different strategies to explore the participation of epigenetic alterations in the context of two different immunodeficiencies, HIGM syndrome and CVID. Firstly, we have investigated the ability of AID, mutated in HIGM syndrome, in targeting epigenetic modifications and have identified that it is able to interact with H4K20 methylating enzymes and increase H4K20me3 levels at AID target sites, perhaps related to efficient CSR by modulating H4K20me3 levels. This could be impaired in HIGM AID mutants, which might contribute to the pathogenesis of HIGM. However we have been unable to identify those changes in our cellular model. Secondly, we have identified impairment in the demethylation of certain key genes during differentiation of B cells from CVID patients, which might be related with the alterations of proportions in B cell subsets observed in this disease. In the following section, the results obtained during this doctoral thesis are further discussed.

5.1. AID ability to target epigenetic changes: impact on Hyper-IgM Syndrome

Several lines of evidence point to the epigenetic mechanisms as key regulators of the events associated with antibody affinity maturation. SHM and CSR involve important alterations in DNA structure and epigenetic mechanisms have shown not only to be crucial for the establishment of the proper DNA conformation during these processes, but also for the recruitment of the machinery required.

The targeting of AID activity to specific DNA regions has been one of the most studied aspects related to B cell activation. Nevertheless, it still remains to be elucidated. Most of the research efforts related to epigenetics mechanisms and antibody affinity maturation have been focused on describing epigenetic landscapes that contribute to the AID targeting and binding for CSR and SHM and, accordingly, AID recruitment has been associated with specific histone marks and specific DNA methylation patterns.

Nevertheless, there are also evidences of a direct role of AID in establishing epigenetic mechanisms. AID expression and activity is one of the main events triggered during B cell activation, therefore, the proposal of a regulatory role of AID in addition to its deamination catalytic activity, is a very attractive and plausible idea.

B cell activation models are invaluable tools to determine the main events during this process and to gain insights closer to the physiological context related with non-epigenetics and epigenetics aspects. Nevertheless, regarding epigenetic

mechanisms, specific effects attributed to AID can be masked by the upstream and downstream events around AID expression and activity that occur during B cell activation. This is evident by analyzing the differences in histone acetylation levels across the *IGH* locus after two different stimuli to undergo CSR. Splenic murine B cells were activated with LPS+ CD-40L or with LPS+CD-40L+IL-4. Both stimuli trigger the expression and activity of AID, however, these stimuli differentially modulate the chromatin status at the *IGH* locus. Stimulation with LPS+CD-40L is associated with an increase in histone acetylation in Sμ and Iγ3 regions, with almost no variation in S regions and promoters of γ1 and ε loci. Conversely, LPS+CD-40L+IL-4 stimulation increase the histone acetylation in the γ1 and ε loci, whereas the Sμ and γ3 regions are unaffected (Wang et al., 2006). These findings are in agreement with the role of the cytokines in driving the decisions about the immunoglobulin isotype selected for switching, but since both stimulations require AID expression and activity, these models of B cell activation difficult the establishment of a direct role for AID in regulating the chromatin status at the *IGH* locus, at least regarding to histone acetylation levels, but also extensive to other epigenetic marks.

For that reason, we decided to analyze the two main epigenetic mechanisms, DNA methylation and histone modifications, in two different cells systems with inducible expression of AID to evaluate its possible direct regulatory role in epigenetic mechanisms. On the one hand, Hela cells, which have been widely used in different molecular biology approaches, provides a good study system due to its rapid growth rate and its easy handling during experimental procedures. On the other, the cell line Jiyoye provides a B cell context that lacks AID expression and therefore constitutes a good model to evaluate the effects of AID overexpression.

Our first observation was that following DNase I chromatin digestion, AID presented a similar release pattern than the exhibited by H4K20me3 heterochromatic mark. It is well established that AID is targeted to actively transcribed open chromatin regions and it has been reported that its localization is associated with active histone marks (Yamane et al., 2011), therefore it was expected to observe a coincidence in the release pattern of AID and the active histone marks represented by H3K4me3 and H3Ac in our experiment. Unexpectedly, AID was absent in the soluble fraction of DNase I digestion and mirrored the pattern of release of H4K20me3 in the insoluble fraction of the digestion. This fraction is not only associated with constitutive heterochromatin, but also contains regions of chromatin associated with large multimeric (transcriptional and DNA-repair) protein complexes (Thambirajah et al., 2012), which we presume are the most probably AID associated regions.

Then, we decided to analyze the epigenetic status of *IGH* locus, specifically at the Sμ and Cμ regions, the *bona fide* respective binding and non-binding sites for AID. Uninduced Hela and Jiyoye cells exhibit a similar pattern of DNA methylation in Sμ and Cμ regions; however, the Sμ regions appear to be more methylated in Hela than Jiyoye cells, whereas Cμ regions are more methylated in Jiyoye than Hela cells, which might reflect cell-specific epigenetic patterns. Accordingly, Jiyoye pronouncedly showed the epigenetic pattern previously described in B cells for Sμ and Cμ regions, with low and high DNA methylation levels respectively. Following the induction of AID expression and nuclear accumulation, it was not observed any alteration of this DNA methylation status neither in Jiyoye nor in Hela cells. Given that the Sμ region displayed very low levels of DNA methylation before the AID induction, we decided to analyze PAX5, an alternative target of AID, also reported previously (Yamane et al., 2011). The region analyzed in PAX5 presented high levels of DNA methylation before AID expression in both, Hela and Jiyoye cells, and similarly to Sμ regions, PAX5 did not change its DNA methylation levels after AID expression.

Moreover, we analyzed the global levels DNA methylation in Jiyoye cells before and after AID induction by using the 450K Illumina DNA methylation array and there were very few DNA methylation changes associated with AID overexpression. and at a similar extent of the ones observed when comparing different control samples.

This constituted an unexpected finding, given that at the moment in which this project started, AID had emerged as one of the strongest candidates to participate in the DNA demethylation mechanisms. As aforementioned, although there are biochemical evidences against the role of AID in DNA demethylation (Nabel et al., 2012), evidence from studies in embryonic contexts still supports the participation of AID in DNA demethylation mechanisms (Bhutani et al., 2010; Morgan et al., 2004; Popp et al., 2010). Additionally, different facts contributed to the proposal of AID role in DNA demethylation in B cells. It has been reported a marked shift in DNA methylation patterning in GC B cells versus resting/naive B cells, in which GC B cells were predominantly hypomethylated compared with naive B cells and AID binding sites were highly overrepresented among hypomethylated loci (Shaknovich et al., 2011). Supporting these findings, the analysis of the complete DNA methylome of the three mature B-cell subpopulations, naïve B cells, class switched and unswitched memory B cells; it has been reported that the transition from naïve to class switched memory B cells is associated with a global shift toward hypomethylation, and it has been suggested that these changes probably take place in GC (Kulis et al., 2012). Thus, DNA methylation analysis regarding B cell activation, together with the involvement of

AID in the DNA demethylation pathway, points AID as a strong candidate to drive the DNA methylation changes described in this context.

To dissipate any doubts about the involvement of AID in DNA demethylation events in B cells, and supporting the results obtained in this doctoral thesis, it was recently published a work in which it was confirmed the inability of AID to drive DNA methylation changes in mice models of B cell activation. By analyzing the transcriptome and methylome of Aid-deficient, wild-type and Aid-overexpressing activated B cells, Fritz and colleagues were not able to detect any change in DNA methylation profile to support the proposed role for AID in regulating DNA methylation (Fritz et al., 2013). The analysis of our model system also revealed the lack of effects on DNA methylation, suggesting that contribution of the cytidine deaminase activity of AID in DNA methylation, if any, is restricted to specific situations and is not simply associated with a high expression of AID, perhaps more related to differentiation processes. Thus, the results about the role of AID on DNA methylation reported in this doctoral thesis are in agreement with actual investigations.

Regarding the direct role of AID in modulating epigenetic mechanisms, specifically histone marks, there is evidence that suggest that AID is involved in the modulation of histone acetylation during B cell activation. It has been reported that AID-deficient B cells showed reduced levels of acetylation in both, H3 and H4 histones in S regions. Thus, the increase of H3Ac and H4Ac levels observed during B cell activation depends on AID expression (Wang et al., 2006), therefore this finding constitute an evidence of a regulatory role of AID in epigenetic mechanisms.

During this doctoral thesis, we proposed to determine the direct effects of AID in the epigenetic modulation at the global level and at the *IGH* locus. One of the most relevant findings of this doctoral thesis was the identification of the mechanistic link between AID and H4K20 methylation changes. Our results demonstrated a novel interaction between AID and SUV420H enzymes that resulted in the increase of H4K20me3 and changes in the accessibility to the SUV420H enzymes to the Sμ sites at the *IGH* locus. These interactions and changes were impaired in the AID W68X mutant and probably in the R190X mutant. H4K20 methylation is a histone modification that has been linked with diverse epigenetic functions. In B cell context, it has been demonstrated that alterations in the methylation states of H4K20 have profound effects on CSR. The characterization of *Suv4-20h*-dn mice revealed that loss of H4K20me2 and H4K20me3 histone marks, with the consequent transition to an H4K20me1 state, results in compromised chromatin that is insufficient to protect genome integrity and to

the proper processing of DNA recombination events, such as CSR (Schotta et al., 2008).

The increase in the H4K20me3 levels induced by AID overexpression in our cell system was not related with changes in the accessibility of S regions or with alterations in the transcription of the μGLT, therefore there is not a clear evidence of the role of this change during CSR. Recently, it has been proposed that hMOF-mediated H4K16Ac and SUV420H2-mediated H4K20me3 play opposing roles in the regulation of Pol II pausing. H4K16Ac promoted the release of Pol II from pausing, whereas Pol II remained paused in the presence of H4K20me3 (Kapoor-Vazirani et al., 2011). It is well established that the stalling of Pol II is a crucial step for AID targeting. Stalled Pol II not only acts a docking protein for the AID recruitment through Spt5 interaction (Pavri et al., 2010), but also facilitates the recruitment of histone modifiers to generate accessibility required for AID activity (Wang et al., 2009). On the basis of these findings, we suggest that AID might reinforce the stalling of Pol II by recruiting SUV4-20H2 and the increase of the H4K20me3 levels at the Sμ region, thus generating a positive feedback loop that reinforce its recruitment to mediate CSR. A further investigation is required to support this hypothesis.

Previous findings and the results that have been shown in this doctoral thesis provide evidence of an AID role beyond its catalytic activity during antibodies affinity maturation. The discovery of AID partners and the function of the different interactions established by AID is a crucial step to fully understand the additional roles of this enzyme during CSR and SHM. Thus this doctoral thesis is contributing to the understanding of AID roles by revealing a new protein interaction with SUV4-20H enzymes with potential physiological role. Moreover, there is evidence that this interaction is impaired in HIGM AID mutants, which can provide an additional mechanism of deficient B cell function in HIGM.

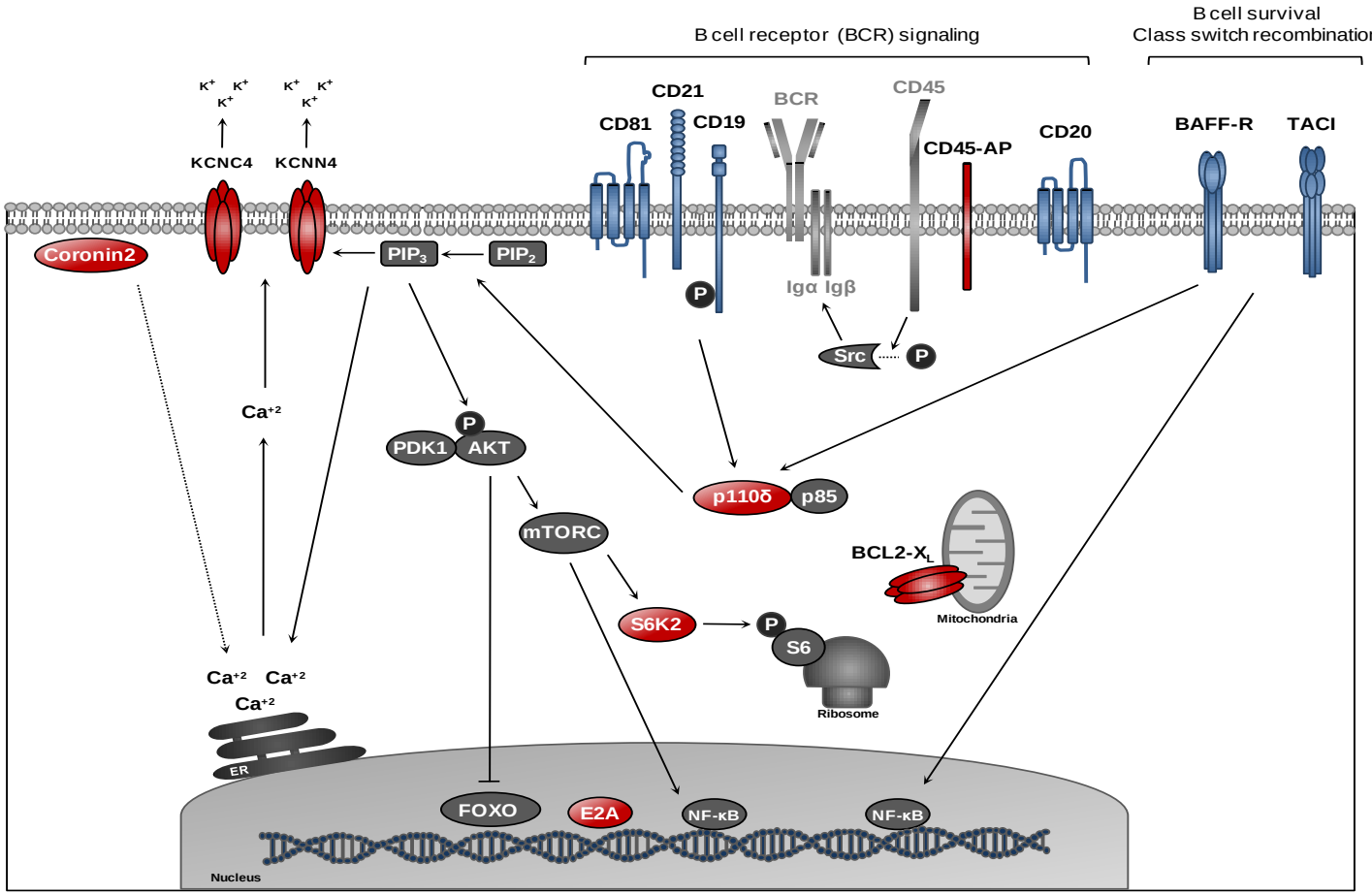
5.2. DNA methylation alterations in CVID

Our study demonstrates for the first time the existence of DNA methylation alterations in CVID, most notably, in genes relevant to B cell function. The observed changes in CVID individuals indicate that the unswitched and switched memory cell subsets are impaired to achieve the DNA demethylation found in healthy individuals. In contrast, the naïve B cell subset of CVID patients displays normal levels of methylation. The observed changes might be one of the mechanisms participating in the defective generation of memory cells in CVID individuals. Also, our results prove that studies of monozygotic twins are an extremely valuable for identifying DNA methylation

differences, since interference with the genetic contribution is minimized and it is possible to tease out the epigenetic component.

Various lines of evidence indicate that both genetic and epigenetic factors play a role in the development of primary immunodeficiencies (Knight, 2013). Despite enormous efforts to elucidate the genetic basis of CVID, its molecular basis remains elusive for most patients. Several factors and pathways relevant to B cell biology have been associated with CVID in genetic studies. Our analysis here has revealed additional elements of these pathways that are altered at the DNA methylation level (Figure D1). An important target for alterations in CVID is the B cell receptor (BCR) signaling pathway. BCRs initiate and control several key processes, such as proliferation, differentiation, migration, activation and survival of B cells, and signaling through this receptor is important not only during the first steps of B cell differentiation in the bone marrow, but also for the proper function of fully differentiated B cells (McHeyzer-Williams et al., 2011). Although the BCR has not been reported to be genetically altered in CVID patients, and is by itself the source of a separate group of PIDs, there is evidence of alterations in several molecules that are necessary to its function in this disorder. For instance, genes that encode molecules such as CD19, CD21, CD81, CD20 (Kuijpers et al., 2010; Thiel et al., 2012; van Zelm et al., 2006; Zelm and Smet, 2010) are mutated in some CVID patients. CD19, CD21 and CD81 make up the B cell co-receptor complex, whose function is to decrease the threshold of B cell activation, conferring an enhanced response against pathogens at low concentrations. Specifically, CD81 is necessary for the expression of CD19, while CD21 is a receptor for the complement component C3d. Moreover, CD19 is responsible for BCR signal enhancement mainly through the activation of PI3K signaling.

In our study, one of the most interesting genes displaying increased DNA methylation in memory B cells from CVID individuals was *PI3KCD*. This gene encodes p110δ, the catalytic subunit of one of the PI3K isoforms, a master regulator of B cell signaling due to its role in the transduction of signals from different molecules and receptors. Through the generation of inositol-1,4,5-trisphosphate 3 (PIP3), p110δ recruits and activates important proteins that drive the downstream effects of BCR signaling (Durand et al., 2009). Mutations in this gene have been described for several PIDs (Angulo et al., 2013; Lucas et al., 2014; Salzer et al., 2013; Thiel et al., 2012).



Gene	Protein
BCL2-L1	BCL2-X_L
KCNN4	KCNN4
KCNC4	KCNC4
PIK3CD	p110δ
RPS6KB2	S6K2
TCF3	E2A
CORO1B	Coronin2
PTPRCAP	CD45-AP

Figure D1. Scheme depicting B cell factors that have been implicated in CVID from genetic studies or from the present study, looking at DNA methylation changes. In blue, molecules whose encoding genes have been reported to be altered in CVID. In red, the products of genes that were found to be hypermethylated in memory B cells in our study. In gray, molecules important in B cell biology and related to the genetically or epigenetically altered molecules. The table on the right panel contains the names for the genes and proteins. In addition, the following proteins/genes are represented: BCR, B cell receptor; Igα and Igβ, BCR co-receptors; CD45, receptor-like tyrosine phosphatase that regulates the action of the Src family kinases. The table on the right bottom side contains the names for the genes and proteins. In addition, the following molecules, proteins, and organelles are depicted in dark gray: P, phosphorylation; PIP_2, Phosphatidylinositol-4,5-biphosphate; PIP_3, inositol-1,4,5-trisphosphate; PDK1, 3-phosphoinositide-dependent protein kinase-1; p85, regulatory subunit of PI3K; S6, S6 ribosomal protein; Src, family of kinases that triggers BCR signaling by phosphorylation of Igα and Igβ; ER, endoplasmic reticulum.

One of the downstream effects of PI3K signaling is the increase in the intracellular Ca^{2+} levels to generate a "Ca^{2+} signal". PIP3 can induce the release of the calcium stores from the endoplasmic reticulum to the cytosol. Another downstream effect of PI3K signaling is the activation of the AKT/mTORC/SP6K axis. PIP3 promotes the activation of AKT, which in turn phosphorylates and activates the mTORC complexes. The activated mTORC complexes then phosphorylate and activate different substrates, such as the S6K2 kinase (encoded by *RSP6KB2*) (Figure D1). We have also found this gene to be hypermethylated in memory B cells in CVID patients. The S6K2 kinase has several substrates, but its main function is to phosphorylate the ribosomal protein S6, promoting cell growth and protein synthesis.

Once the Ca^{2+} signal has been initiated, other proteins become involved in the enhancement and maintenance of the raised Ca^{2+} levels. Although Coronin2, encoded by *CORO1B*, also hypermethylated in CVID memory B cells, has not been related to B cells directly, it has been reported that one member of its family, CORO1A, has an important role in the immune system through Ca^{2+} mobilization mediated by actin remodeling changes. It is possible that CORO1B has similar roles in the B cell context, but this remains to be determined. Other molecules with an important role in calcium signaling are the potassium channels, like those encoded by other genes hypermethylated in CVID B cells, *KCNN4* and *KCNC4* (Figure D1). They are responsible for maintaining the elevated calcium levels through the mobilization of K+ ions to the extracellular medium.

B cell differentiation and activation require high rates of protein synthesis and proliferation. In normal conditions both may provide an apoptotic stimulus, requiring the action of anti-apoptotic molecules such as BCL2-XL (Figure D1), encoded by the *BCL2-L1* gene, and which are also hypermethylated in CVID B cells. This protein maintains the mitochondrial and blocks the release of cytochrome c from mitochondria.

Finally, at the end of BCR signaling there are the transcription factors, which can transform the signals received at the cell surface into gene expression changes that enable the cell to adapt and respond. To ensure the proper gene expression pattern is achieved for the different stimuli, it is important not only for several transcription factors to be activated, but also for those that are not required to be blocked. E2A (encoded by another hypermethylated gene in CVID memory B cells, namely *TCF3*) (Figure D1) and NF-κB are two examples of transcription factors activated during BCR signaling, while FOXO transcription factors are blocked after BCR activation.

Our findings regarding the impaired ability to demethylate genes during the transition from naïve or memory cells reinforces the notion emerging from various studies that CVID individuals are deficient in differentiating towards memory B cells. For instance, we have known for over 40 years that CD73 activity is reduced in B lymphocytes of patients suffering from immunodeficiency syndromes (Edwards et al., 1978; Johnson et al., 1977) and we now know that CD73-dependent adenosine generation favors class-switch recombination, endowing the B cell with an intrinsic control of differentiation towards immunoglobulin class-switched plasma cells (Schena et al., 2013). As mentioned above, CVID patients have severely deficient levels of switched memory B cells (Warnatz et al., 2002). These altered proportions of B cell subsets are also reflected in the cell counts in the cohort of CVID individuals analyzed in this study. Interestingly, at the DNA methylation level, genes like *PI3KCD*, *RSP6KB2*, *KCNN4*, *KCNC4*, *CORO1B*, and *BCL2L1* are hypermethylated in memory B cells of CVID individuals relative to healthy individuals, or, in other words, they do not undergo the demethylation in the transition from naïve to memory cells observed in healthy individuals. This could also be interpreted as meaning that if the DNA methylation patterns for these genes in memory cells from CVID individuals were reminiscent of those from naïve cells, their ability to respond/maintain appropriate responses could therefore be diminished.

Finally, our investigation demonstrates the utility of twin studies for identifying epigenetic changes in complex diseases. Given that monozygotic twins share their entire genotype (and potentially specific susceptibility gene variants), our results reinforce the notion that, for a particular genetic background, epigenetic changes are related to the onset of CVID. However, currently, we cannot distinguish whether these epigenetic changes are a cause or a consequence; in other words, whether the DNA methylation changes are caused by these individual's genetic predisposition, perhaps influenced by the environment, or whether environmental factors cause the CVID by some unknown mechanism, which is then accompanied by immune responses that are secondary to the CVID.

A potential constraint on the use of this approach arises from the fact that many monozygotic monochorionic twins exchange blood through shared vascular communication. If any DNA methylation changes occur in the blood of one of the twins before birth, it is common for both of them to show it. Only the sibling of the pair in whom the change has originally occurred may present the phenotype, because additional cells of the body could be affected. Therefore, monozygotic twins are a good model system for diseases, like PIDs, that start to develop after birth, when there is no

blood sharing. Our data do not allow us to draw any conclusions about the timing of the DNA methylation changes observed, although it is likely that they are associated with the onset of the disease.

A previous study by our team revealed the existence of epigenetic differences between monozygotic twins that are discordant for SLE (Javierre et al., 2010), reinforcing the notion of an epigenetic component in complex, immune-related diseases. Epigenetic differences in monozygotic twins become more pronounced with age, supporting the idea that “epigenetic drift” plays a role in the divergence of monozygotic phenotypes (Fraga et al., 2005). Data obtained from monozygotic and dizygotic twins have provided further evidence that differences in epigenomes can explain phenotypic differences (Kaminsky et al., 2009). Thus, the present report, in which the discordance for CVID between a pair of monozygotic twins is used to identify epigenetic targets, highlights the potential of this strategy for learning about the contribution of epigenetics to immune-related disorders as well as to other complex diseases.

CONCLUSIONS

6. CONCLUSIONS

The conclusions obtained during this doctoral thesis can be summarized as follows:

1. Analysis of a B cell lines-based inducible system for AID shows that during DNase I digestion, AID is associated with the insoluble fraction of chromatin, together with the heterochromatin histone modification H4K20me3.

2. AID induction and its nuclear accumulation have no effects on global and gene-specific DNA methylation, in agreement with recent results by other laboratories. These results discard a role for AID in active DNA demethylation, at least in B cell models.

3. AID induction and its nuclear accumulation causes a discrete but robust increase in the global levels of H4K20me3.

4. Association of AID with the Sμ region of the IGH locus is concomitant with a specific local increase of H4K20me3 in both, Jiyoye and Hela cells, whereas H3K4me3 increases after AID overexpression and nuclear accumulation only in Jiyoye cells, and H3K27me3 do not experiment significant changes. This suggests a specific modulation of H4K20me3 levels by AID.

5. AID interacts with the three members of the SUV4-20H family and modifies their binding to the Sμ region of the *IGH* locus. Co-expression of AID and SUV4-20H1_i1 and SUV4-20H1_i2 results in a diminished binding of these enzymes to the Sμ regions, whereas the co-expression of AID and SUV4-20H2 is translated to an increased binding of this enzyme to the Sμ regions. This provides a possible mechanism to explain the increased levels of H4K20me3 following AID overexpression and nuclear accumulation.

6. Hyper-IgM AID mutations have different effects on AID nuclear localization. Missense mutations present similar subcellular localization when compared with WT AID, whereas the nonsense mutations, W68X and R190X, which lack the nuclear export signal show a constitutively nuclear localization that is unaffected by the LMB treatment.

7. The W68X mutant for AID displays a lower ability to interact with SUV4-20H2 enzyme when compared with WT AID. Accordingly, the co-expression of W68X mutant and SUV4-20H2 shows an impaired recruitment of the latter to the Sμ region.

8. The results obtained using the AID system suggest that this enzyme has a regulatory role of epigenetic mechanisms, specifically by the direct modulation of the

levels of H4K20me3. Additionally, this regulatory role could be impaired in some HIGM mutations. However, we have been unable to detect any alteration on the levels of H4K20me3 for the mutant that we have analyzed, specifically the W68X mutant.

9. The comparison of the DNA methylation patterns of total B cells between a pair of monozygotic twins discordant for CVID reveals the predominant gain of DNA methylation in the CVID sibling. Specifically, it has been identified hypermethylation in 230 CpG sites and hypomethylation in 81 CpG sites.

10. The identified changes in DNA methylation in CVID are specific to B cells, as demonstrated by the analysis of T cells and monocytes.

11. Analysis of the functional categories from our high-throughput screening of DNA methylation through GO analysis reveals an enrichment of important biological processes in the differentially methylated genes. The significant enriched categories are not only related to general biological processes, but are also involved in central pathways of B cell biology and differentiation.

12. Analysis of DNA methylation in a selection of relevant genes among the hypermethylated set of genes in naïve B cells, and the unswitched and switched memory B cells, indicates a progressive loss of DNA methylation in healthy individuals that is impaired in CVID patients. The DNA methylation profile of memory B cells from CVID patients resembles the DNA methylation profile observed in naïve B cells, suggesting a defect in the establishment of the proper epigenetic pattern during B cell differentiation in this disorder.

BIBLIOGRAPHY

7. BIBLIOGRAPHY

Abbas, A.K., and Janeway, C.A. (2000). Immunology: improving on nature in the twenty-first century. Cell *100*, 129–138.

Al-Herz, W., Bousfiha, A., Casanova, J.-L., Chatila, T., Conley, M.E., Cunningham-Rundles, C., Etzioni, A., Franco, J.L., Gaspar, H.B., Holland, S.M., et al. (2014). Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency. Front. Immunol. *5*, 162.

Allen, C.D.C., Okada, T., Tang, H.L., and Cyster, J.G. (2007). Imaging of germinal center selection events during affinity maturation. Science *315*, 528–531.

Angulo, I., Vadas, O., Garçon, F., Banham-Hall, E., Plagnol, V., Leahy, T.R., Baxendale, H., Coulter, T., Curtis, J., Wu, C., et al. (2013). Phosphoinositide 3-kinase δ gene mutation predisposes to respiratory infection and airway damage. Science *342*, 866–871.

Balakrishnan, L., and Milavetz, B. (2010). Decoding the histone H4 lysine 20 methylation mark. Crit. Rev. Biochem. Mol. Biol. *45*, 440–452.

Baldovino, S., Montin, D., Martino, S., Sciascia, S., Menegatti, E., and Roccatello, D. (2013). Common variable immunodeficiency: crossroads between infections, inflammation and autoimmunity. Autoimmun. Rev. *12*, 796–801.

Ballestar, E., Paz, M., Valle, L., and Wei, S. (2003). Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. EMBO *22*.

Barreto, V.M., and Magor, B.G. (2011). Activation-induced cytidine deaminase structure and functions: a species comparative view. Dev. Comp. Immunol. *35*, 991–1007.

Basu, U., Meng, F., Keim, C., Grinstein, V., Eccleston, J., Zhang, T., Myers, D., Wesemann, D.R., Januszyk, K., Gregory, R.I., et al. (2011). NIH Public Access. *144*, 353–363.

Begum, N. a, Stanlie, A., Nakata, M., Akiyama, H., and Honjo, T. (2012). The histone chaperone Spt6 is required for activation-induced cytidine deaminase target determination through H3K4me3 regulation. J. Biol. Chem. *287*, 32415–32429.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum. Mol. Genet. *9*, 2395–2402.

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.

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol. Cell *48*, 491–507.

Blom, B., and Spits, H. (2006). Development of human lymphoid cells. Annu. Rev. Immunol. *24*, 287–320.

Boboila, C., Yan, C., Wesemann, D.R., Jankovic, M., Wang, J.H., Manis, J., Nussenzweig, A., Nussenzweig, M., and Alt, F.W. (2010). Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. J. Exp. Med. *207*, 417–427.

Borchert, G.M., Holton, N.W., Edwards, K. a., Vogel, L. a., and Larson, E.D. (2010). Histone H2A and H2B are monoubiquitinated at AID-targeted loci. PLoS One *5*.

Bransteitter, R., Prochnow, C., and Chen, X.S. (2009). The current structural and functional understanding of APOBEC deaminases. Cell. Mol. Life Sci. *66*, 3137–3147.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288–289.

Casali, P., Pal, Z., Xu, Z., and Zan, H. (2006). DNA repair in antibody somatic hypermutation. Trends Immunol. *27*, 313–321.

Casola, S., Otipoby, K.L., Alimzhanov, M., Humme, S., Uyttersprot, N., Kutok, J.L., Carroll, M.C., and Rajewsky, K. (2004). B cell receptor signal strength determines B cell fate. Nat. Immunol. *5*, 317–327.

Cerutti, A., Cols, M., and Puga, I. (2013). Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. Nat. Rev. Immunol. *13*, 118–132.

Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat. Rev. Immunol. *4*, 541–552.

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature *422*, 726–730.

Cheng, X., and Roberts, R.J. (2001). AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res. *29*, 3784–3795.

Chester, A., Scott, J., Anant, S., and Navaratnam, N. (2000). RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. Biochim. Biophys. Acta *1494*, 1–13.

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. Proc. Natl. Acad. Sci. U. S. A. *105*, 15872–15877.

Cobaleda, C., Schebesta, A., Delogu, A., and Busslinger, M. (2007). Pax5: the guardian of B cell identity and function. Nat. Immunol. *8*, 463–470.

Conley, M.E., Mathias, D., Treadaway, J., Minegishi, Y., and Rohrer, J. (1998). Mutations in btk in patients with presumed X-linked agammaglobulinemia. Am. J. Hum. Genet. *62*, 1034–1043.

Conley, M.E., Dobbs, a K., Farmer, D.M., Kilic, S., Paris, K., Grigoriadou, S., Coustan-Smith, E., Howard, V., and Campana, D. (2009). Primary B cell immunodeficiencies: comparisons and contrasts. Annu. Rev. Immunol. *27*, 199–227.

Conticello, S.G. (2008). The AID/APOBEC family of nucleic acid mutators. Genome Biol. *9*, 229.

Crouch, E.E., Li, Z., Takizawa, M., Fichtner-Feigl, S., Gourzi, P., Montaño, C., Feigenbaum, L., Wilson, P., Janz, S., Papavasiliou, F.N., et al. (2007). Regulation of AID expression in the immune response. J. Exp. Med. *204*, 1145–1156.

Cunningham-Rundles, C., and Ponda, P.P. (2005). Molecular defects in T- and B-cell primary immunodeficiency diseases. Nat. Rev. Immunol. *5*, 880–892.

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.

Davies, E.G., and Thrasher, A.J. (2010). Update on the hyper immunoglobulin M syndromes. Br. J. Haematol. *149*, 167–180.

Demorest, Z.L., Li, M., and Harris, R.S. (2011). Phosphorylation directly regulates the intrinsic DNA cytidine deaminase activity of activation-induced deaminase and APOBEC3G protein. J. Biol. Chem. *286*, 26568–26575.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.-H., Robbiani, D.F., Di Virgilio, M., Reina San-Martin, B., Heidkamp, G., Schwickert, T. a, et al. (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity *28*, 630–638.

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics *11*, 587.

Durand, C. a, Hartvigsen, K., Fogelstrand, L., Kim, S., Iritani, S., Vanhaesebroeck, B., Witztum, J.L., Puri, K.D., and Gold, M.R. (2009). Phosphoinositide 3-kinase p110 delta regulates natural antibody production, marginal zone and B-1 B cell function, and autoantibody responses. J. Immunol. *183*, 5673–5684.

Durandy, A., Revy, P., Imai, K., and Fischer, A. (2005). Hyper-immunoglobulin M syndromes caused by intrinsic B-lymphocyte defects. Immunol. Rev. *203*, 67–79.

Durandy, A., Taubenheim, N., Peron, S., and Fischer, A. (2007). Pathophysiology of B-cell intrinsic immunoglobulin class switch recombination deficiencies. Adv. Immunol. *94*, 275–306.

Durandy, A., Kracker, S., and Fischer, A. (2013). Primary antibody deficiencies. Nat. Rev. Immunol. *13*, 519–533.

Edwards, N.L., Magilavy, D.B., Cassidy, J.T., and Fox, I.H. (1978). Lymphocyte ecto-5'-nucleotidase deficiency in agammaglobulinemia. Science *201*, 628–630.

Ferrari, S., Lougaris, V., Caraffi, S., Zuntini, R., Yang, J., Soresina, A., Meini, A., Cazzola, G., Rossi, C., Reth, M., et al. (2007). Mutations of the Igbeta gene cause agammaglobulinemia in man. J. Exp. Med. *204*, 2047–2051.

Fischer, A. (2004). Human primary immunodeficiency diseases: a perspective. Nat. Immunol. *5*, 23–30.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U. S. A. *102*, 10604–10609.

Freitas, N., and Cunha, C. (2009). Mechanisms and signals for the nuclear import of proteins. Curr. Genomics *10*, 550–557.

Fritz, E.L., Rosenberg, B.R., Lay, K., Mihailović, A., Tuschl, T., and Papavasiliou, F.N. (2013). A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. Nat. Immunol. *14*, 749–755.

Fujieda, S., Lin, Y.Q., Saxon, A., and Zhang, K. (1996). Multiple types of chimeric germ-line Ig heavy chain transcripts in human B cells: evidence for trans-splicing of human Ig RNA. J. Immunol. *157*, 3450–3459.

Gazumyan, A., Timachova, K., Yuen, G., Siden, E., Di Virgilio, M., Woo, E.M., Chait, B.T., Reina San-Martin, B., Nussenzweig, M.C., and McBride, K.M. (2011). Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. Mol. Cell. Biol. *31*, 442–449.

Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat. Rev. Genet. *13*, 343–357.

Guo, J.U., Su, Y., Zhong, C., Ming, G., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423–434.

György, I., Boller, S., Nechanitzky, R., Mandel, E., Pott, S., Liu, E., and Grosschedl, R. (2012). Transcription factor Ebf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. Genes Dev. *26*, 668–682.

Han, L., Masani, S., and Yu, K. (2011). Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA deamination mediates innate immunity to retroviral infection. Cell *113*, 803–809.

Hasham, M.G., Donghia, N.M., Coffey, E., Maynard, J., Snow, K.J., Ames, J., Wilpan, R.Y., He, Y., King, B.L., and Mills, K.D. (2010). Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. Nat. Immunol. *11*, 820–826.

Hernando, H., Shannon-Lowe, C., Islam, A.B., Al-Shahrour, F., Rodríguez-Ubreva, J., Rodríguez-Cortez, V.C., Javierre, B.M., Mangas, C., Fernández, A.F., Parra, M., et al.

(2013). The B cell transcription program mediates hypomethylation and overexpression of key genes in Epstein-Barr virus-associated proliferative conversion. Genome Biol. *14*, R3.

Hernando, H., Islam, A.B.M.M.K., Rodríguez-Ubreva, J., Forné, I., Ciudad, L., Imhof, A., Shannon-Lowe, C., and Ballestar, E. (2014). Epstein-Barr virus-mediated transformation of B cells induces global chromatin changes independent to the acquisition of proliferation. Nucleic Acids Res. *42*, 249–263.

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.

Ito, S., Nagaoka, H., Shinkura, R., Begum, N., Muramatsu, M., Nakata, M., and Honjo, T. (2004). Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. Proc. Natl. Acad. Sci. U. S. A. *101*, 1975–1980.

Javierre, B.M., Fernandez, A.F., Richter, J., Al-Shahrour, F., Martin-Subero, J.I., Rodriguez-Ubreva, J., Berdasco, M., Fraga, M.F., O'Hanlon, T.P., Rider, L.G., et al. (2010). Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus. Genome Res. *20*, 170–179.

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.

Johnson, S.M., North, M.E., Asherson, G.L., Allsop, J., Watts, R.W., and Webster, A.D. (1977). Lymphocyte purine 5'-nucleotidase edficiency in primary hypogammaglobulinaemia. Lancet *1*, 168–170.

Kaminsky, Z.A., Tang, T., Wang, S.-C., Ptak, C., Oh, G.H.T., Wong, A.H.C., Feldcamp, L.A., Virtanen, C., Halfvarson, J., Tysk, C., et al. (2009). DNA methylation profiles in monozygotic and dizygotic twins. Nat. Genet. *41*, 240–245.

Kapoor-Vazirani, P., Kagey, J.D., and Vertino, P.M. (2011). SUV420H2-mediated H4K20 trimethylation enforces RNA polymerase II promoter-proximal pausing by blocking hMOF-dependent H4K16 acetylation. Mol. Cell. Biol. *31*, 1594–1609.

Klein, U., and Dalla-Favera, R. (2008). Germinal centres: role in B-cell physiology and malignancy. Nat. Rev. Immunol. *8*, 22–33.

Knight, J.C. (2013). Genomic modulators of the immune response. Trends Genet. *29*, 74–83.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472–479.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009). Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J. Biol. Chem. *284*, 478–485.

Kracker, S., and Durandy, A. (2011). Insights into the B cell specific process of immunoglobulin class switch recombination. Immunol. Lett. *138*, 97–103.

Kuijpers, T.W., Bende, R.J., Baars, P. a, Grummels, A., Derks, I. a M., Dolman, K.M., Beaumont, T., Tedder, T.F., van Noesel, C.J.M., Eldering, E., et al. (2010). CD20 deficiency in humans results in impaired T cell-independent antibody responses. J. Clin. Invest. *120*, 214–222.

Kulis, M., Heath, S., Bibikova, M., Queirós, A.C., Navarro, A., Clot, G., Martínez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. *44*, 1236–1242.

Kumar, R., DiMenna, L., Schrode, N., Liu, T.-C., Franck, P., Muñoz-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 *500*, 89–92.

Kuraoka, M., Holl, T.M., Liao, D., Womble, M., Cain, D.W., and Reynolds, A.E. (2011). Activation-induced cytidine deaminase mediates central tolerance in B cells.

Lai, A.Y., Mav, D., Shah, R., Grimm, S.A., Phadke, D., Hatzi, K., Melnick, A., Geigerman, C., Sobol, S.E., Jaye, D.L., et al. (2013). DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at Alu elements during B-cell activation. Genome Res. *23*, 2030–2041.

Langerak, P., Krijger, P.H.L., Heideman, M.R., van den Berk, P.C.M., and Jacobs, H. (2009). Somatic hypermutation of immunoglobulin genes: lessons from proliferating cell nuclear antigenK164R mutant mice. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *364*, 621–629.

Lebecque, S.G., and Gearhart, P.J. (1990). Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. J. Exp. Med. *172*, 1717–1727.

LeBien, T.W., and Tedder, T.F. (2008). B lymphocytes: how they develop and function. Blood *112*, 1570–1580.

Lee, S.-T., Xiao, Y., Muench, M.O., Xiao, J., Fomin, M.E., Wiencke, J.K., Zheng, S., Dou, X., de Smith, A., Chokkalingam, A., et al. (2012). A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. Nucleic Acids Res. *40*, 11339–11351.

Li, G., Zan, H., Xu, Z., and Casali, P. (2013). Epigenetics of the antibody response. Trends Immunol. *34*, 460–470.

Liadaki, K., Sun, J., Hammarström, L., and Pan-Hammarström, Q. (2013). New facets of antibody deficiencies. Curr. Opin. Immunol. *25*, 629–638.

Longerich, S., Basu, U., Alt, F., and Storb, U. (2006). AID in somatic hypermutation and class switch recombination. Curr. Opin. Immunol. *18*, 164–174.

Lucas, C.L., Kuehn, H.S., Zhao, F., Niemela, J.E., Deenick, E.K., Palendira, U., Avery, D.T., Moens, L., Cannons, J.L., Biancalana, M., et al. (2014). Dominant-activating germline mutations in the gene encoding the PI(3)K catalytic subunit p110δ result in T cell senescence and human immunodeficiency. Nat. Immunol. *15*, 88–97.

Maeda, K., Singh, S.K., Eda, K., Kitabatake, M., Pham, P., Goodman, M.F., and Sakaguchi, N. (2010). GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. J. Biol. Chem. *285*, 23945–23953.

Van Maldegem, F., Scheeren, F.A., Aarti Jibodh, R., Bende, R.J., Jacobs, H., and van Noesel, C.J.M. (2009). AID splice variants lack deaminase activity. Blood *113*, 1862–1864; author reply 1864.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99–103.

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.

Marabita, F., Almgren, M., Lindholm, M.E., Ruhrmann, S., Fagerström-Billai, F., Jagodic, M., Sundberg, C.J., Ekström, T.J., Teschendorff, A.E., Tegnér, J., et al. (2013). An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. Epigenetics *8*, 333–346.

Matthews, A.J., Zheng, S., DiMenna, L.J., and Chaudhuri, J. (2014). Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. (Elsevier Inc.).

McBride, K.M., Gazumyan, A., Woo, E.M., Schwickert, T.A., Chait, B.T., and Nussenzweig, M.C. (2008). Regulation of class switch recombination and somatic mutation by AID phosphorylation. J. Exp. Med. *205*, 2585–2594.

McHeyzer-Williams, M., Okitsu, S., Wang, N., and McHeyzer-Williams, L. (2011). Molecular programming of B cell memory. Nat. Rev. Immunol. *12*, 24–34.

Meyers, G., Ng, Y.-S., Bannock, J.M., Lavoie, A., Walter, J.E., Notarangelo, L.D., Kilic, S.S., Aksu, G., Debré, M., Rieux-Laucat, F., et al. (2011). Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. Proc. Natl. Acad. Sci. U. S. A. *108*, 11554–11559.

Minegishi, Y., Coustan-Smith, E., Wang, Y.H., Cooper, M.D., Campana, D., and Conley, M.E. (1998). Mutations in the human lambda5/14.1 gene result in B cell deficiency and agammaglobulinemia. J. Exp. Med. *187*, 71–77.

Minegishi, Y., Coustan-Smith, E., Rapalus, L., Ersoy, F., Campana, D., and Conley, M.E. (1999a). Mutations in Igalpha (CD79a) result in a complete block in B-cell development. J. Clin. Invest. *104*, 1115–1121.

Minegishi, Y., Rohrer, J., Coustan-Smith, E., Lederman, H.M., Pappu, R., Campana, D., Chan, A.C., and Conley, M.E. (1999b). An essential role for BLNK in human B cell development. Science *286*, 1954–1957.

Morgan, H.D., Dean, W., Coker, H. a, Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J. Biol. Chem. *279*, 52353–52360.

Mu, Y., Prochnow, C., Pham, P., Chen, X.S., and Goodman, M.F. (2012). A structural basis for the biochemical behavior of activation-induced deoxycytidine deaminase class-switch recombination-defective hyper-IgM-2 mutants. J. Biol. Chem. *287*, 28007–28016.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J. Biol. Chem. *274*, 18470–18476.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. Cell *102*, 553–563.

Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K., and Honjo, T. (2000). Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene. Genomics *68*, 85–88.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat. Chem. Biol. *8*, 751–758.

Nagaoka, H., Tran, T.H., Kobayashi, M., Aida, M., and Honjo, T. (2010). Preventing AID, a physiological mutator, from deleterious activation: regulation of the genomic instability that is associated with antibody diversity. Int. Immunol. *22*, 227–235.

Nieuwenhuis, P., and Opstelten, D. (1984). Functional anatomy of germinal centers. Am. J. Anat. *170*, 421–435.

Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. Annu. Rev. Biochem. *76*, 1–22.

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.

Odegard, V.H., Kim, S.T., Anderson, S.M., Shlomchik, M.J., and Schatz, D.G. (2005). Histone modifications associated with somatic hypermutation. Immunity *23*, 101–110.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Okazaki, I., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., and Honjo, T. (2003). Constitutive expression of AID leads to tumorigenesis. J. Exp. Med. *197*, 1173–1181.

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, E.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.

Parra, M. (2009). Epigenetic events during B lymphocyte development. Epigenetics *4*, 462–468.

Pasqualucci, L., Kitaura, Y., Gu, H., and Dalla-Favera, R. (2006). PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. Proc. Natl. Acad. Sci. U. S. A. *103*, 395–400.

Patenaude, A., and Noia, J.M. Di (2010). The mechanisms regulating the subcellular localization of AID. 325–331.

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.

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.

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. J. Biol. Chem. *283*, 17428–17439.

Piccolo, F.M., and Fisher, A.G. (2014). Getting rid of DNA methylation. Trends Cell Biol. *24*, 136–143.

Pone, E.J., Zhang, J., Mai, T., White, C.A., Li, G., Sakakura, J.K., Patel, P.J., Al-Qahtani, A., Zan, H., Xu, Z., et al. (2012). BCR-signalling synergizes with TLR-signalling for induction of AID and immunoglobulin class-switching through the non-canonical NF-κB pathway. Nat. Commun. *3*, 767.

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.

Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F., and Chen, X.S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature *445*, 447–451.

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.

Rakhmanov, M., Keller, B., Gutenberger, S., Foerster, C., Hoenig, M., Driessen, G., van der Burg, M., van Dongen, J.J., Wiech, E., Visentini, M., et al. (2009). Circulating CD21low B cells in common variable immunodeficiency resemble tissue homing, innate-like B cells. Proc. Natl. Acad. Sci. U. S. A. *106*, 13451–13456.

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.

Reth, M., and Nielsen, P. (2014). Signaling circuits in early B-cell development. (Elsevier Inc.).

Revy, P., Muto, T., Levy, Y., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Gennery, A., Tezcan, I., Ersoy, F., et al. (2000). Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2). *102*, 565–575.

Rodriguez, J., Vives, L., Jordà, M., Morales, C., Muñoz, M., Vendrell, E., and Peinado, M. a (2008). Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells. Nucleic Acids Res. *36*, 770–784.

Rodriguez-Cortez, V.C., Hernando, H., de la Rica, L., Vento, R., and Ballestar, E. (2011). Epigenomic deregulation in the immune system. Epigenomics *3*, 697–713.

Salzer, E., Santos-Valente, E., Klaver, S., Ban, S.A., Emminger, W., Prengemann, N.K., Garncarz, W., Müllauer, L., Kain, R., Boztug, H., et al. (2013). B-cell deficiency and severe autoimmunity caused by deficiency of protein kinase C δ. Blood *121*, 3112–3116.

Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. Nat. Rev. Genet. *9*, 129–140.

Schena, F., Volpi, S., Faliti, C.E., Penco, F., Santi, S., Proietti, M., Schenk, U., Damonte, G., Salis, A., Bellotti, M., et al. (2013). Dependence of immunoglobulin class switch recombination in B cells on vesicular release of ATP and CD73 ectonucleotidase activity. Cell Rep. *3*, 1824–1831.

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callén, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I. a, 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.

Schwickert, T.A., Lindquist, R.L., Shakhar, G., Livshits, G., Skokos, D., Kosco-Vilbois, M.H., Dustin, M.L., and Nussenzweig, M.C. (2007). In vivo imaging of germinal centres reveals a dynamic open structure. Nature *446*, 83–87.

Shaffer, A.L., Yu, X., He, Y., Boldrick, J., Chan, E.P., and Staudt, L.M. (2000). BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. Immunity *13*, 199–212.

Shaknovich, R., Cerchietti, L., Tsikitas, L., Kormaksson, M., De, S., Figueroa, M.E., Ballon, G., Yang, S.N., Weinhold, N., Reimers, M., et al. (2011). DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. Blood *118*, 3559–3569.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646–650.

Sims, R.J., Nishioka, K., and Reinberg, D. (2003). Histone lysine methylation: a signature for chromatin function. Trends Genet. *19*, 629–639.

Southall, S.M., Cronin, N.B., and Wilson, J.R. (2014). A novel route to product specificity in the Suv4-20 family of histone H4K20 methyltransferases. Nucleic Acids Res. *42*, 661–671.

Stavnezer, J. (2011). Complex regulation and function of activation-induced cytidine deaminase. Trends Immunol. *32*, 194–201.

Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. *26*, 261–292.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41–45.

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.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity *28*, 621–629.

Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nat. Publ. Gr. *15*, 703–708.

Thambirajah, A. a, Ng, M.K., Frehlick, L.J., Li, A., Serpa, J.J., Petrotchenko, E. V, Silva-Moreno, B., Missiaen, K.K., Borchers, C.H., Adam Hall, J., et al. (2012). MeCP2 binds to nucleosome free (linker DNA) regions and to H3K9/H3K27 methylated nucleosomes in the brain. Nucleic Acids Res. *40*, 2884–2897.

Thiel, J., Kimmig, L., Salzer, U., Grudzien, M., Lebrecht, D., Hagena, T., Draeger, R., Voelxen, N., Völxen, N., Bergbreiter, A., et al. (2012). Genetic CD21 deficiency is associated with hypogammaglobulinemia. J. Allergy Clin. Immunol. *129*, 801–810.e6.

Tran, T.H., Nakata, M., Suzuki, K., Begum, N. a, Shinkura, R., Fagarasan, S., Honjo, T., and Nagaoka, H. (2010). B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. Nat. Immunol. *11*, 148–154.

Traver, D., and Akashi, K. (2004). Lineage commitment and developmental plasticity in early lymphoid progenitor subsets. Adv. Immunol. *83*, 1–54.

Tsang, L.W.K., Hu, N., and Underhill, D.A. (2010). Comparative analyses of SUV420H1 isoforms and SUV420H2 reveal differences in their cellular localization and effects on myogenic differentiation. PLoS One *5*, e14447.

Turelli, P., Mangeat, B., Jost, S., Vianin, S., and Trono, D. (2004). Inhibition of hepatitis B virus replication by APOBEC3G. Science *303*, 1829.

Tuscano, J., Druey, K., and Riva, A. (1996). Bcl-x rather than Bcl-2 mediates CD40-dependent centrocyte survival in the germinal center. … 1359–1364.

Victora, G.D., and Nussenzweig, M.C. (2012). Germinal centers. Annu. Rev. Immunol. *30*, 429–457.

Vuong, B.Q., Herrick-Reynolds, K., Vaidyanathan, B., Pucella, J.N., Ucher, A.J., Donghia, N.M., Gu, X., Nicolas, L., Nowak, U., Rahman, N., et al. (2013). A DNA break- and phosphorylation-dependent positive feedback loop promotes immunoglobulin class-switch recombination. Nat. Immunol. *14*, 1183–1189.

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.

Warnatz, K., Denz, A., Dräger, R., Braun, M., Groth, C., Wolff-Vorbeck, G., Eibel, H., Schlesier, M., and Peter, H.H. (2002). Severe deficiency of switched memory B cells (CD27(+)IgM(-)IgD(-)) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease. Blood *99*, 1544–1551.

Woo, C.J., Martin, A., and Scharff, M.D. (2003). Induction of somatic hypermutation is associated with modifications in immunoglobulin variable region chromatin. Immunity *19*, 479–489.

Wu, X., Darce, J.R., Chang, S.K., Nowakowski, G.S., and Jelinek, D.F. (2008). Alternative splicing regulates activation-induced cytidine deaminase (AID): implications for suppression of AID mutagenic activity in normal and malignant B cells. Blood *112*, 4675–4682.

Xu, Z., Pone, E.J., Al-Qahtani, A., Park, S.-R., Zan, H., and Casali, P. (2007). Regulation of aicda expression and AID activity: relevance to somatic hypermutation and class switch DNA recombination. Crit. Rev. Immunol. *27*, 367–397.

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.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. Nat. Immunol. *12*, 62–69.

Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P., et al. (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. Nature *449*, 478–482.

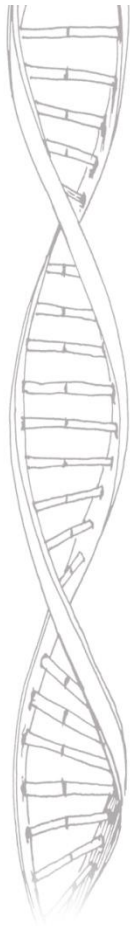
De Yébenes, V.G., Belver, L., Pisano, D.G., González, S., Villasante, A., Croce, C., He, L., and Ramiro, A.R. (2008). miR-181b negatively regulates activation-induced cytidine deaminase in B cells. J. Exp. Med. *205*, 2199–2206.

Yel, L., Minegishi, Y., Coustan-Smith, E., Buckley, R.H., Trübel, H., Pachman, L.M., Kitchingman, G.R., Campana, D., Rohrer, J., and Conley, M.E. (1996). Mutations in the mu heavy-chain gene in patients with agammaglobulinemia. N. Engl. J. Med. *335*, 1486–1493.

Zahn, A., Eranki, A.K., Patenaude, A.-M., Methot, S.P., Fifield, H., Cortizas, E.M., Foster, P., Imai, K., Durandy, A., Larijani, M., et al. (2014). Activation induced deaminase C-terminal domain links DNA breaks to end protection and repair during class switch recombination. Proc. Natl. Acad. Sci. U. S. A. *111*, E988–E997.

Zelm, M. van, and Smet, J. (2010). CD81 gene defect in humans disrupts CD19 complex formation and leads to antibody deficiency. J. Clin. Invest. *120*.

Van Zelm, M.C., Reisli, I., van der Burg, M., Castaño, D., van Noesel, C.J.M., van Tol, M.J.D., Woellner, C., Grimbacher, B., Patiño, P.J., van Dongen, J.J.M., et al. (2006). An antibody-deficiency syndrome due to mutations in the CD19 gene. N. Engl. J. Med. *354*, 1901–1912.



SUPPLEMENTARY INFORMATION

8. SUPPLEMENTARY INFORMATION

In the following section, I enclose two publications in which I participated. The first one is a review published in *Epigenomics* in 2011. The review is entitled: **"Epigenomic deregulation in the immune system"**. In this publication, we presented the relevant findings that linked epigenetics mechanisms with immunological disorders and we highlighted the importance of the epigenetic analysis to understand the molecular mechanisms underlying immune system associated diseases.

In the second publication included in this section, I participated as a co-author. This was a study from my group lab in which we determined the DNA methylation changes associated with immortalization of B cells after the infection with EBV. This work was published in *Genome Biology* in 2013. The title of this publication is **"The B cell transcription program mediates hypomethylation and overexpression of key genes in Epstein-Barr virus-associated proliferative conversion"**. In addition to the characterization of DNA methylation changes associated with the immortalization of B cells by EBV, we determined that these changes were mainly associated with genes regulated by B cell specific transcription factors. Thus, we proposed that the B cell transcription program drives the epigenetic changes that allows the expression of relevant genes during immortalization with EBV.

Epigenomic deregulation in the immune system

Proper immune function is the result of multiple cell commitment and differentiation steps, and adequate control of activation mechanisms. Deregulation of transcriptional programs in immune cells leads to the development of hematological malignancies, autoimmune diseases or immunodeficiencies. In this sense, epigenetic control of gene expression plays an essential role in the correct function of the immune system and the integrity of identity of relevant cell types. Epigenetic deregulation can result as a consequence of genetic changes in transcription factors, elements of signaling pathways or epigenetic enzymes, or as an effect of a variety of environmental factors. On top of genetic predisposition, viral infection and other external factors influence the development of immune-related diseases. In recent years, major strides have been made towards understanding the contribution of genetics in these immune disorders. Less progress has been made in dissecting the contribution of epigenetic factors in their etiology. Herein, it is presented what is currently known about epigenetic alterations in immune system associated disorders. It is also discussed how epigenomic analysis can help to understand the molecular basis of these diseases and how this information can be used in the clinical setting.

KEYWORDS: autoimmune ▪ DNA methylation ▪ epigenetics ▪ immune system ▪ immunodeficiency ▪ lymphocyte ▪ miRNAs

Virginia C Rodriguez-Cortez, Henar Hernando, Lorenzo de la Rica, Roser Vento & Esteban Ballestar*

Chromatin & Disease Group, Cancer Epigenetics & Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Avda. Gran Via 199–203, 08907 L'Hospitalet de Llobregat, Barcelona, Spain
**Author for correspondence:*
Tel.: +34 932 607 133
Fax: +34 932 607 219
eballestar@idibell.org

The immune system encompasses different cells and molecules that protect the host organism by recognizing and eliminating potentially harmful internal or external agents. The immune response comprises both innate and adaptive mechanisms that are integrated and coordinated to provide specific lines of defense in function of the threatening pathogen or situation. Innate immune cells (chiefly, macrophages, neutrophils, dendritic cells and natural killer lymphocytes) provide a more rapid but less specific response to microbial infections than do adaptive immune cells (primarily B and T lymphocytes), which provide a highly antigen-specific response and immunological memory [1,2].

Innate immune cells recognize highly conserved molecules present in a variety of microorganisms through a set of germline-encoded receptors, known as pattern recognition receptors that are not modified through recombination after fertilization [3]. Interactions between microorganisms and pattern recognition receptors lead to activation of different innate mechanisms, including phagocytosis, release of inflammatory mediators, activation of complement system proteins, and synthesis of acute phase proteins, cytokines and chemokines [4]. Unlike the pattern recognition receptors of innate cells, the receptors of adaptive cells are generated through somatic recombination of segments of their encoding sequences, and subsequently, selected for clonal expansion to provide a wide repertoire of specificities, not only against pathogens but also against virtually any antigen [5]. However, the innate and adaptive mechanisms are not separate: communication between the two components is mediated by antigen-presenting cells, mainly represented by dendritic cells. Antigen-presenting cells instruct B and T lymphocytes to develop a particular effector response through the costimulatory signals [6].

Differentiation and activation of immune cells are very tightly regulated; they are controlled at different stages of development to ensure an efficient immune response that maintains the delicate balance between self-recognition and attack of foreign molecules or microorganisms, and that discriminates between injured and normal host cells. Control of these processes greatly depends on the timely expression of specific sets of genes. It is well known that during immune cell differentiation, complex networks of transcription factors orchestrate activation and repression of thousands of genes that confer proper cell identity [7], and that during B or T cell activation, several key genes are expressed to enhance the specificity of the immune response [8]. There is an intimate functional link between transcription factors and epigenetic modifications in the regulation of gene expression, and as we discuss

future medicine part of fsg

later, they are codependent. Proper differentiation and activation partly depends on the genetic integrity of the genes encoding for these key transcription factors. Myriad genetic alterations or polymorphisms have been identified, which has contributed to identify factors that are crucial in immune cell differentiation and in immune regulation. These alterations or variants have been associated with various immune-related disorders, including autoimmune diseases and immunodeficiencies. Genetic defects in hematopoietic transcription factors are also key in the development of leukemias and lymphomas, although these will not be discussed in this article. Classical genetic analysis and, more recently, genome-wide association studies have enabled identification of polymorphisms associated with several autoimmune diseases in many genes, including several in the MHC region, *STAT4*, *PTPN22*, *IRF5* and others [9]. Regarding immunodeficiencies, several genes (e.g., CD40 ligand [*CD154*]) exhibit mutations that lead to impaired class switch recombination, and these are often associated with hyper-IgM syndrome [10].

Given the interdependence between transcription factors and elements of the epigenetic machinery, defects or variations in the sequence or structure of transcription factors can affect the setting of epigenetic modifications, which ultimately have an influence on transcription profiles. On the other hand, the cell environment, including viral infection, exposure to chemicals or hormones can also influence the epigenetic profile of cells, and therefore, have a direct effect on gene-expression profiles. For instance, hormones and other small chemicals bind their corresponding receptors and binding can modify their binding properties of these receptors to their target genomic sites or the subunit composition (including epigenetic enzymes) of the nuclear complexes where those receptors are embedded. Another example is provided by viral infection, where viral proteins can use the nuclear machinery of the host cell and directly target and modify the epigenetic status at specific genomic sites. Epigenetic modifications are essential for gene function as they are determinant in the establishment of specific transcription profiles, and ultimately, determine cell identity and function. Indeed, there is an ever increasing interest in epigenetic modifications, however most of the data come from the analysis of a few genes. The availability of high-throughput epigenome mapping techniques is opening new possibilities to address the molecular basis of these diseases. Herein, we report on some of the most relevant findings on epigenetic alterations in the context of immune system-related disorders using autoimmune diseases and immunodeficiencies as prime examples. We will also discuss the potential effects of viral infection on the epigenetic profiles of immune cells.

Epigenetic regulation

Proper cell identity and function requires the precise expression or repression of particular sets of genes. The gene sets that are specifically expressed in different cell types are determined by many factors, including which signaling pathways are activated, local DNA structure and chromatin organization, the availability and binding of transcription factors and other nuclear factors, and the distribution and dynamics of epigenetic modifications. All these factors are interconnected. In this article, we will mainly focus on the role of epigenetic modifications.

Epigenetic modifications exist in two major forms: DNA methylation and histone post-translational modifications. Both groups of modifications have effects on gene transcription, especially when occurring at or near a promoter sequence (Figure 1).

In vertebrates, DNA methylation mainly takes place at CpG dinucleotides. CpG sites tend to concentrate in the genome in various repetitive elements (e.g., satellite DNA, short and long interspersed nuclear elements) and in CpG islands, CG-rich regions coincident with the promoters of a high proportion of RNA polymerase II-transcribed genes. In general, promoter methylation leads to transcriptional repression. In normal cells, most promoter CpG islands remain unmethylated and methylation is limited to a relatively small proportion of genes including imprinted genes, X-linked genes in females, and some tissue-specific genes. The general repressive role of promoter methylation is very well characterized in cancer cells, where large groups of genes become hypermethylated at their promoter CpG islands and are therefore repressed [11]. It has also been shown that regions adjacent to CpG islands, termed CpG island shores, can also undergo changes in DNA methylation and gene expression [12]. Regarding CpG sites in repetitive sequences, these are frequently methylated in normal cells, where these elements constitute the major source of 5-methylcytosine. In cancer, these sequences become hypomethylated. Hypomethylation of repetitive elements is generally associated with chromosomal instability and has been proposed to participate in increased mutation events [11].

Methylation of CpGs is catalyzed by DNA methyltransferases (DNMTs) [13], which use the methyl group from *S*-adenosylmethionine. DNA demethylation can occur as a result of pharmacological inhibition of DNMTs. This process is the molecular basis for several compounds used to decrease aberrant levels of hypermethylation in cancer [14]. Alternatively, active demethylation occurs in a limited specific physiological contexts, including cell differentiation and reprogramming, and during lymphocyte activation [15,16]; however, the molecular pathways involved remain less well known. Compelling evidence indicates that active demethylation may be dependent on the activity of cytosine deaminases, such as activation-induced cytidine deaminase (AICDA) [17,18]. It has been reported that AICDA-dependent demethylation would require the participation of the G/T mismatch-specific thymine DNA glycosylase, and the damage response factor GADD45α [19]. Recently, conversion of 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine by TET proteins has also been described, and the possibility that DNA demethylation might occur as a TET-mediated mechanism is currently explored [20–22].

Another major source of epigenetic information comes from post-translational modification of histones [23]. There are at least 60 amino acid residues that can undergo post-translational modifications [23]. These include lysine acetylation, lysine or arginine (mono-, di- or tri-) methylation, serine or threonine phosphorylation, and other, less frequent modifications. Specific combinations of histone modifications constitute a sort of code associated with different transcriptional states, competence for DNA replication or DNA repair. Chromatin immunoprecipitation on individual genes, or in combination with microarrays or high-throughput sequencing, is helping to establish associations between histone modifications and functional situations. For instance, trimethylation of histone H3 at lysines 9 and 27 (H3K9me3; H3K27me3) generally associate with gene repression; whereas hyperacetylation of histones H3 and H4 (H3Ac; H4Ac) and trimethylation of lysine 4 of H3 (H3K4me3) are characteristic of many active genes. Different enzymes, including histone lysine and arginine methyltransferases and demethylases, and histone lysine acetyltransferases and deacetylases, maintain the balance of histone modifications [24]. These enzymes constitute a potential source of new targets for drug development [25].

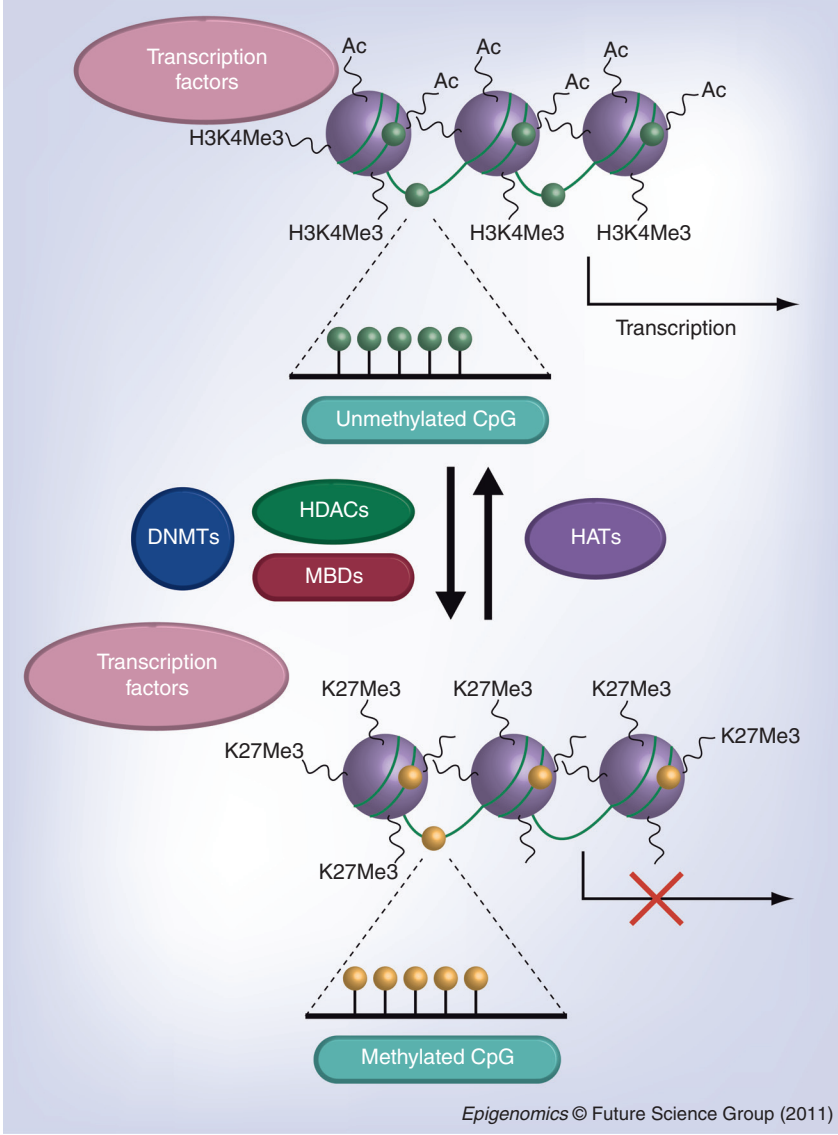


Figure 1. Relationship of major epigenetic modifications with transcriptional activity. This image shows the transition of the active chromatin state into the inactive one. Histone octamers are represented by purple circles. DNA is represented as a green line. Histone tails are protruding lines from octamers. Methyl-CpG dinucleotides are represented by yellow circles, whereas unmethylated CpGs are represented as green circles. In the first state (top), chromatin is relaxed and transcriptionally competent: DNA is unmethylated, N-terminal tails of histones are Ac by the action of the HATs and histone H3 tails are methylated at K4 (H3K4me3). This situation allows the access of transcription factors and leads to gene transcription. In the second situation (bottom), histone acetylation is lost by the action of the HDACs, and DNA is methylated by DNMTs. Methylated regions are enriched with MBDs that are also associated with corepressor complexes. In this inactive chromatin, transcription factors are released and the gene is not expressed. Ac: Acetylated; DNMT: DNA methyltransferase; HAT: Histone acetyltransferase; HDAC: Histone deacetylase; MBD: Methyl-CpG binding domain protein.

Several mechanisms interconnect DNA methylation and histone modifications. These include factors such as methyl-CpG binding domain proteins that associate with methylated DNA and are subunits of multiprotein complexes that contain histone modification enzymes [26]. Polycomb group proteins, such as

EZH2, constitute another link between DNA methylation and histone modifications [27].

As indicated, epigenetic modifications are key to regulating gene expression. The regulatory role of epigenetic modifications is amplified when this associates with miRNA regulation. miRNAs are short (21–24 nucleotides in length), ncRNA species that downregulate gene expression post-transcriptionally. Although miRNAs cannot strictly be considered epigenetic factors themselves, there are complex networks that interconnect them with epigenetic modifications. Several studies have shown that epigenetic mechanisms that regulate miRNA expression. Conversely, different subsets of miRNAs control the levels of important epigenetic enzymes [28].

Cell identity and function depends on the precise expression of thousands of genes and the adequate stability of noncoding sequences. Transcription factors, miRNAs and epigenetic modifications will then be essential to maintain the adequate balance. Different cell types and specific functional situations are then characterized by specific epigenomic profiles and transcriptomes. It is then essential to obtain and integrate these data to get a complete understanding of the epigenetic control of gene regulation in the immune system. Furthermore, the epigenome acts as an interface between the genome and the environment that may be considered at two levels: the cellular environment and the environment surrounding an entire organism.

Much knowledge has been acquired on epigenetic alterations in immune system-related diseases over the past few years. In hematological malignancies, the epigenetic switch at many genomic sites is also commonly recognized. However, a better understanding of the causes of epigenetic deregulation in other groups of disorders, such as autoimmune diseases and immunodeficiencies, is needed. Mapping epigenomic changes at the DNA methylation, the histone modification and the factor binding levels, as well as knowing their direct associations with gene expression in immune system cells, will surely provide a solid base from which to address these issues.

Epigenetic deregulation in autoimmune disease

During the past few years, the study of epigenetic alterations in the pathogenesis of autoimmune diseases is raising unprecedented attention among clinicians and researchers in the field although our knowledge is still limited. Autoimmune diseases are multifactorial illnesses caused by aberrant activation of T and/or B cells, in the absence of an ongoing infection or other discernible cause, that attack the self-molecules or components of the organism [29]. To prevent harming the organism itself, the system must have limited recognition of self-molecules. This is achieved through immunological tolerance, which ensures that few self-reactive lymphocytes are generated, and that these remain inactive [30]. Although autoreactive lymphocytes and antibodies are found in the general population, certain individuals suffer from a combination of genetic susceptibility and exposure to environmental risk factors that lead to disruption of this tolerance, which ultimately provokes immunologic attack of their own cells and tissue.

Autoimmune reactions can occur against one organ (organ-specific autoimmunity) or against various organs (systemic autoimmunity). Organ-specific autoimmune diseases include diabetes mellitus Type 1, primary biliary cirrhosis and celiac disease; examples of systemic autoimmune diseases are systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and multiple sclerosis (MS). Most of the currently available epigenetics data on autoimmune disorders has focused on two rheumatic diseases, SLE and RA. In this section, we will mainly discuss these two diseases.

The etiology of autoimmune diseases remains largely unknown. Although several susceptibility loci, including HLA genes, *PTPN22*, *IRF5* and *STAT4*, have been identified, the onset of these conditions cannot be explained by genetics alone [31,32]. In fact, widely incomplete concordance rates in monozygotic (MZ) twins [33–35] strongly support the existence of environmental factors in the development of autoimmunity. Several environmental risk factors are known to be involved, including exposure to chemicals or hormones, and viral infection. The presence of these molecules in the cell environment can have effects on cell signaling pathways or directly result in their binding with different ligand-binding receptors that can translocate to the cell nucleus and, ultimately, affect epigenetic modifications and consequently, alter the activity of genes relevant to immune function.

The first clues of epigenetic deregulation in autoimmune disease arose in the mid-1980s in the context of SLE. Treatment of $CD4^+$ T lymphocytes with DNMT inhibitors such as 5-azacytidine, hydralazine or procainamide, which potentially result in passive demethylation, made these cells autoreactive: they could be

activated by autologous macrophages by losing the requirement of an antigen [36,37]. The observed phenotype of the DNMT inhibitor-treated CD4+ T cells was similar to that of CD4+ T cells isolated from SLE patients [38]. These results, together with the finding that CD4+ T lymphocytes isolated from both SLE and RA patients had lower levels of 5-methylcytosine, led to implication of DNA methylation changes in the pathogenesis of autoimmune rheumatic diseases [39]. In the case of SLE, the symptomatology directly correlated with a decrease of 5-methylcytosine [40,41]. Moreover, some researchers then reported that administration of these drugs to humans caused drug-induced lupus [42]. Also, T cells isolated from healthy individuals who had been treated with hypomethylating drugs were subsequently injected into syngeneic mice, they caused a lupus-like disease [43]. Hypomethylation can potentially have different effects: hypomethylation of repetitive elements could result in increased chromosomal instability. When associated with promoters, hypomethylation generally associates with increased expression.

Analysis of selected individual genes, by using methylation-sensitive restriction enzymes or bisulfite genomic sequencing, led to the identification of several genes that are hypomethylated at their promoters and overexpressed in SLE. Examples of specific genes that exhibit altered DNA mehylation and increased expression levels in SLE, and their phenotypic implications, along with methylation changes in other autoimune diseases are presented in TABLE 1. Most DNA methylation analyses have employed candidate gene strategies, and the vast majority of them have been performed by the group of Bruce Richardson during the past two decades.

More recently, high-throughput approaches have been used to obtain DNA methylation profiles of individual SLE patients. In one study, researchers performed DNA methylation profiling of MZ twins discordant for SLE by using methylation bead arrays and white blood cells. This comparison enabled identification of roughly 50 gene promoters that were differentially methylated between each SLE patient and their corresponding healthy twin. Most of the genes showed increased expression levels associated with their hypomethylated status in SLE twins with respect to healthy siblings. This study also showed that the 18S and 28S segments of ribosomal RNA genes are also hypomethylated and overexpressed in SLE twins with respect to normal healthy twins. The use of MZ twins also enabled screening to be done without the existence of genetic differences between individuals [44]. Interestingly, in a different study, on CD4+ T cells, investigators identified another set of differentially methylated genes in SLE [45], some of which had also been reported in the study with MZ twins. The study with CD4+ cells allowed for the conformation of some of the results obtained for white blood cells derived from discordant MZ twins. Candidate gene studies and these first high-throughput analyses have unveiled several functional targets for epigenetic deregulation in SLE (FIGURE 2A). These studies reinforce the notion that further high-throughput analysis need to be done with other cell types relevant to immune function, particularly with different B cell subsets.

The need for studying DNA methylation changes in specific cell types is highlighted for instance by the results obtained for RA. High-throughput methylation profiling of white blood cells from MZ twins discordant for RA has not shown significant differences at the gene promoter level [44]. In contrast, there is accumulated evidence that synovial fibroblasts derived from RA are epigenetically altered. Synovial fibroblasts are major effector cells of joint degradation in RA. These cells are more aggressive in RA than their healthy counterparts, as they overexpress metalloproteinases, cytokines and show tumor-like behavior (i.e., invasiveness, resistance to apoptosis and anchorage-independent growth), as reviewed in [46]. Many of the phenotypic features of RA synovial fibroblasts can be explained by their epigenetic deregulation. Similarly to T cells from SLE or RA, synovial fibroblasts from RA patients show lower levels of global DNA methylation and expression of DNMT1 is decreased [47,48]. Hypomethylation is related to the cells' aggressiveness, as demonstrated by the fact that healthy fibroblasts isolated from healthy joints and subsequently treated with DNA hypomethylating agents, resemble their pathologic counterparts [49]. Global hypomethylation of RA synovial fibroblasts has been found to be associated with hypomethylation and increased expression of the retrotransposon LINE-1 [50]. Specifically, hypomethylation of several CpG sites upstream of open reading frame-associated LINE-1 elements is associated with increased levels of several genes [51] related to both invasiveness and the tumor phenotype [48,52]. The promoters of several miRNA genes are also hypomethylated, and expression of these genes is increased. For example, hsa-miRNA-203 is hypomethylated and upregulated in RA synovial fibroblasts. Transfection of normal

Table 1. DNA methylation alterations in autoimmune diseases.

Disease	Cell type/tissue	Sequence name	Effect	Event	Ref.
Systemic					
SLE	$CD4^+$ T-cells	*ITGAL/CD11a*	Cell–cell adhesion	Hypometh.	[128]
		CD40LG	Stimulates B-cell IgG overproduction	Hypometh.	[129]
		CD70	Required for T-cell proliferation	Hypometh.	[130]
		PRF1	Autoreactive killing	Hypometh.	[131]
		CD9	Cell adhesion and migration	Hypometh.	[45]
		MMP9	Extracellular matrix destruction	Hypometh.	[45]
		Repeats		Hypometh.	[43]
	PBMCs	*IFNGR2*	IFN-γ receptor	Hypometh.	[44]
		CD9	Cell adhesion and migration	Hypometh.	[44]
		MMP9	Extracellular matrix destruction	Hypometh.	[44]
		MMP14	Extracellular matrix destruction	Hypometh.	[44]
		LCN2	Iron transporter	Hypometh.	[44]
		rDNA (18S, 28S)	Assembly in ribosomal particles	Hypometh.	[44]
		Repeats		Hypometh.	[44]
	$CD19^+$ B-cells	*CD5-E1B*	Interleukin production	Hypometh.	[132]
RA	$CD4^+$ T-cells	Repeats		Hypometh.	[43]
	Synovial fibroblasts	LINE-1	Gene-expression fine-tuning	Hypometh.	[50]
		MAPK1	Proliferation signaling	Hypometh.	[52]
		miRNA-203	Induction of MMPs and IL-6	Hypometh.	[53]
		MET	Proto-oncogene	Hypometh.	[52]
		DR3	Apoptosis	Hypermeth.	[54]
		Repeats		Hypometh.	[49]
	PBMCs	*IL-6*	B-cell response	Hypometh.	[133]
MS	Schwann cells, oligodendrocytes	*PAD2*	Citrullination of MBP, favoring demyelination	Hypometh.	[134]
SjS	Labial salivary glands	*BP230*	Involved in basal lamina anchorage	Hypermeth.	[135]
SD	$CD4^+$ T-cells	Repeats		Hypometh.	[136]
	SD fibroblasts	*FL1*	Increased production of collagen	Hypermeth.	[137]
	PBMCs	X chromosome	SXCIM		[138]
Organ specific					
Ps	Skin cells	*p16* (*CDKN2A*)	Tumor suppressor	Hypermeth.	[139]
		Repeats			[140]
		PTPN6	Cell proliferation and signaling	Hypometh.	[141]
	Hematopoietic cells	*p15* (*CDKN2B*)	Cell cycle progression	Hypometh.	[142]
		p16 (*CDKN2A*)	Tumor suppressor	Hypometh.	[143]
		p21 (*CDKN1A*)	Cell cycle progression	Hypometh.	[142]
	PBMCs	Repeats		Hypermeth.	[140]
		p14 (*CDKN2B*)	Cell cycle progression	Hypermeth.	[140]
UC	Rectum mucosa	Global hypomethylation		Hypometh.	[144]
		p16 (*CDKN2A*)	Tumor suppressor	Hypermeth.	[145]
		p14 (*CDKN2B*)	Cell cycle progression	Hypermeth.	[146]
		MDR1	Transmembrane drug efflux pump	Hypermeth.	[147]
AITDs	PBMCs	X chromosome	SXCIM		[148]

AITD: Autoimmune thyroid disease; Hypermeth.: Hypermethylation; Hypometh.: Hypomethylation; MBP: Myelin basic protein; MS: Multiple sclerosis; PBMC: Peripheral blood mononuclear cell; Ps: Psoriasis; RA: Rheumatoid arthritis; SD: Scleroderma or progressive systemic sclerosis; SjS: Sjögren's syndrome; SLE: Systemic lupus erythematosus; SXCIM: Skewed X-chromosome inactivation mosaicism; UC: Ulcerative colitis.

fibroblasts with this miRNA leads to higher levels of matrix metalloproteinase-1 and IL-6 proteins, which in turn, contribute to the proinflammatory microenvironment of rheumatic joints (Figure 2B) [53].

In contrast with the data for SLE, and as observed in cancer [11], specific promoter hypermethylation has also been found in RA synovial cells. For example, the promoter of the death receptor 3 gene is hypermethylated

(FIGURE 2B), and therefore, the corresponding mRNA levels are downregulated, which explains why these cells exhibit enhanced resistance to apoptosis [54]. As indicated above, the currently available data highlights the need for systematic profiling of DNA methylation alterations for specific cell types in these diseases.

Apart from DNA methylation profile alterations, altered histone modifications have also been found for several SLE and RA. For instance, SLE T cells display aberrant acetylation of histones H3 and H4 [55], a phenomenon which correlates with alterations in gene expression including overexpression of *IL-10* and *CD154* and downregulation of *IFNG* [56]. However, to date, histone modification-mapping data from SLE individuals is scarce. Regarding alteration of histone modifications in RA, research based on histone deacetylase (HDAC) inhibitors has suggested roles for certain histone marks. Interestingly, RA synovial fibroblasts show higher amounts of HDAC1 [57], and depletion of HDACs in RA synovial fibroblasts leads to decreased cell proliferation and to inhibition of TNF-α cytokines thus, it mitigates the aggressive phenotype [58]. Moreover, treatment of various mouse models of RA with HDAC inhibitors ameliorates RA pathology, by inhibiting RA synovial fibroblasts proliferation, downregulating angiogenesis and proinflammatory cytokines, preventing synovial hyperplasia and joint destruction [59–63], and induces apoptosis via the Fas receptor and TNF-related apoptosis-inducing ligand [64,65]. HDAC inhibitors can also suppress the severity of arthritis by modulating blood immune cells: they induce apoptosis, inhibit cell proliferation, and, in autoreactive T cells, block release of IFN-γ [66]. Furthermore, these drugs induce osteoblast proliferation and matrix mineralization [67] and inhibit osteoclastogenesis [68]. These features have made HDAC inhibitors prime candidates for pharmacological treatment of RA [69]. Recent data indicates that EZH2, a Polycomb group protein with histone H3K27 methyltransferase activity, is overexpressed in RA synovial fibroblasts. Analysis of several targets of this factor, for example, the secreted frizzled-related protein 1 gene, has shown occupation at the promoter and change in histone methylation in these cells [70]. Further studies on histone modifications may shed light on the role of these changes in aberrant expression patterns in the cells.

As indicated above, most of the epigenetics studies in autoimmune disease have focused on SLE or RA. Only a few epigenomic analyses have been performed and little data is available for other autoimmune diseases (TABLE 1). Researchers recently performed a high-throughput analysis of CD4+ T cells obtained from MZ twins discordant for MS [71]. Remarkably, they found very few differences at the DNA methylation level. However, the small number of samples and the fact that only one cell type was studied highlight the need for greater research efforts in this area.

Epigenetics & immunodeficiency

Immunodeficiencies represent an opposite pathological scenario to that of autoimmune diseases. In these disorders, one or several components of the immune system are defective, which leaves the organism at greater susceptibility to recurrent infections. Immunodeficiencies are classified as either primary, when caused by genetic defects in the immune cells, or secondary, when caused by external factors (e.g., infectious agents, drugs, metabolic diseases and environmental conditions). Primary immunodeficiency diseases (PIDs) arise from

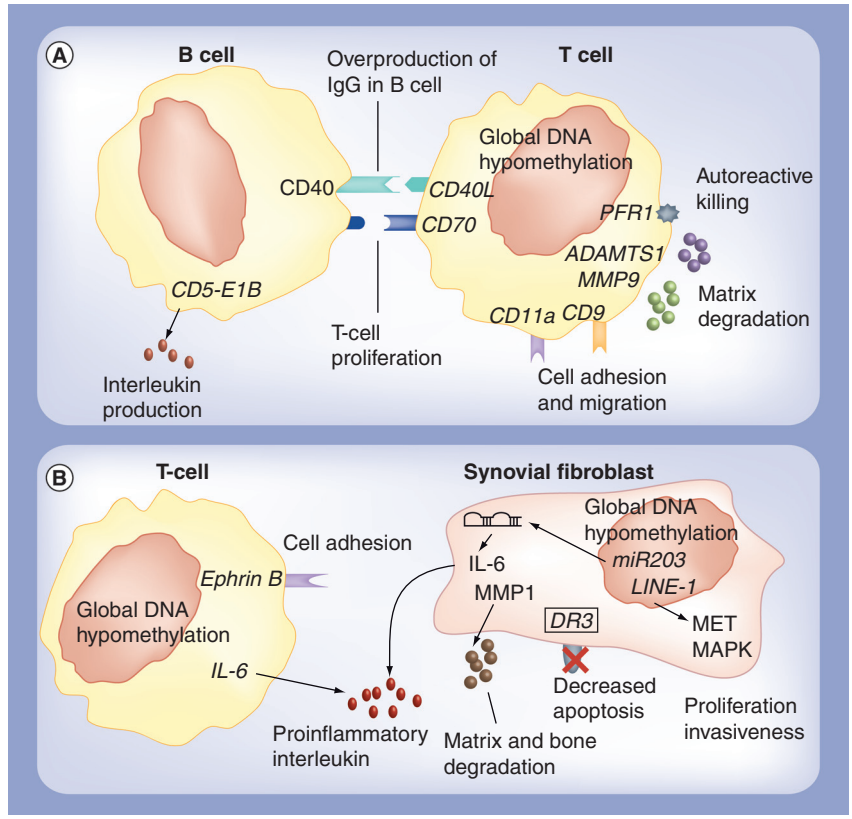


Figure 2. Functional effects on specific cell types of DNA methylation changes in systemic lupus erythematosus and rheumatoid arthritis.
(A) Some examples of hypomethylated and overexpressed genes in T and B cells from systemic lupus erythematosus patients are shown. Their potential functional implications are indicated. **(B)** Examples of some hypomethylated (italic) or hypermethylated (italic in a square) in B cells and rheumatoid arthritis synovial fibroblasts, and the functional consequences of epigenetic deregulation. In synovial fibroblasts, miRNA-203 is hypomethylated and upregulated, causing IL-6 and MMP1 overexpression. LINE-1 repetitive sequences are also hypomethylated and overexpressed, affecting MET and MAPK overexpression.

over 100 genetic alterations in different factors of the innate and adaptive immunities, and PID patients show distinct susceptibilities to various types of pathogens according to their immune defect [72–74].

Albeit PIDs strongly depend on genetic alterations, there is an ever-growing body of evidence implicating additional factors in their etiology. It has been suggested that epigenetic deregulation participates in development of PIDs. As indicated above, epigenetic deregulation can be the result of defects in transcription factors or other elements that ultimately interact with the epigenetic machinery. Also, external factors can modulate the epigenetic profile in individuals with a genetic susceptibility to develop PID. Despite the proposed participation of epigenetic deregulation events in PIDs, for most of them, no data are currently available. In contrast, one specific type of PID constitutes the most remarkable example of the impact of epigenetic defects in an immune system-associated disease: immunodeficiency, centromeric instability and facial anomalies (ICF) syndrome. ICF syndrome is a rare, autosomal recessive disease characterized by agammaglobulinemia in the presence of normal B and T lymphocytes counts, leading to several recurrent infections that cause high mortality in early childhood. Interestingly, ICF syndrome patients are characterized by the presence of mutations in the *DNMT3B* gene, which is responsible for *de novo* DNA methylation during embryogenesis and germ cell development [75,76].

At the chromosomal level, ICF syndrome patients' cells are characterized by DNA hypomethylation in satellite 2 and 3 repeats and by chromatin instability, mostly restricted to mitogen-stimulated lymphocytes, due to alterations in the pericentromeric heterochromatin of chromosomes 1, 9 and 16 (Figure 3) [77]. Interestingly, DNA hypomethylation in ICF syndrome is apparently restricted to the aforementioned regions. Analysis of the CpG islands of genes whose expression is specifically upregulated or downregulated in lymphoblastoid cell lines derived from ICF syndrome patients has shown that these genes do not undergo the DNA methylation changes at their promoter regions [78,79]. Therefore, it has been proposed that hypomethylation in the normally constitutively heterochromatic regions could have a *trans* effect on gene expression by altering the normal sequestration of DNA-sequence specific proteins to the heterochromatin. [79]. Furthermore, it has been proposed that the gene-expression changes observed in ICF syndrome samples might depend on other epigenetic mechanisms, such as changes in histone modifications or in the expression of ncRNAs. Indeed, histone modification changes in the promoter region of five upregulated genes with subtle or absent hypomethylation have been reported in ICF cells. Chromatin immunoprecipitation analysis has revealed loss of repressive H3K27me3 and gain of transcriptionally activating H3K4me3, which correlate with the increased expression of these genes [78]. Alternatively, genome-wide miRNA expression profiling in ICF syndrome lymphoblastoid cell lines has demonstrated deregulation at the miRNA level. Although no changes in the CpG islands associated with the deregulated miRNAs were found, changes were observed in the histone marks, in accordance with the deregulation observed in the miRNAs, suggesting that expression of these miRNAs is controlled through this epigenetic mechanism [80]. Interestingly, most of the analyzed miRNAs and genes were related to pathogenesis of ICF syndrome.

Although the mechanism underlying agammaglobulinemia in ICF syndrome is not fully understood, epigenetic phenomena are known to be strongly influential. Moreover, ICF syndrome underscores the functional relevance of DNA methylation changes in genomic sequences beyond promoter regulation and embodies how specific changes in the methylation patterns of heterochromatin regions can significantly modulate immune response.

Another interesting PID, from the epigenetics perspective, is hyper-IgM 2 (HIGM2) syndrome, an autosomal recessive or dominant disorder that is caused by mutations in the *AICDA* gene, which encodes the enzyme AID [81]. AID is the key enzyme in class switch recombination and in somatic hypermutation, two events necessary for antibody affinity maturation during B lymphocyte activation [82].

AID has recently been linked to the active DNA demethylation process. The first evidence came from *in vitro* assays that demonstrated that AID deaminates 5mC in ssDNA to generate thymidine. This led to the proposal of an equivalent pathway for demethylation of DNA, whereby 5mC deamination could be followed by T:G base excision repair by glycosylases such as TDG or MBD4 [83]. Evidence of a relationship between AID and DNA demethylation in mammals came from two different systems: primordial germ cells from mice, and heterokaryons

(cell fusions) of mouse embryonic stem cells and human fibroblasts. With the first system, researchers observed that primordial germ cells from *AICDA*$^{-/-}$ mice exhibited higher methylation levels than did those of the their wild-type counterparts, suggesting that AID is involved in demethylation during epigenetic reprogramming [84]. In the second system, the heterokaryons were used as an *in vitro* reprogramming model to study the concomitant demethylation process. Transcript levels of the pluripotency markers OCT4 and NANOG were downregulated through promoter methylation after AID knockdown, thus inhibiting the reprogramming process [85].

All of these findings strongly suggest that AID has a role in active DNA demethylation. Assuming that this role also extended to the B lymphocyte context, then epigenetic mechanisms – specifically, aberrant DNA methylation changes – might play a role in pathogenesis of HIGM2 syndrome, in which AID function is impaired. Our knowledge on epigenetic deregulation in primary and secondary immunodeficiencies is still very limited and the few examples mentioned above are rare. Systematic epigenomic profiling of different immune cell types derived from patients will contribute to delimitate the role of epigenetic deregulation events in this group of diseases.

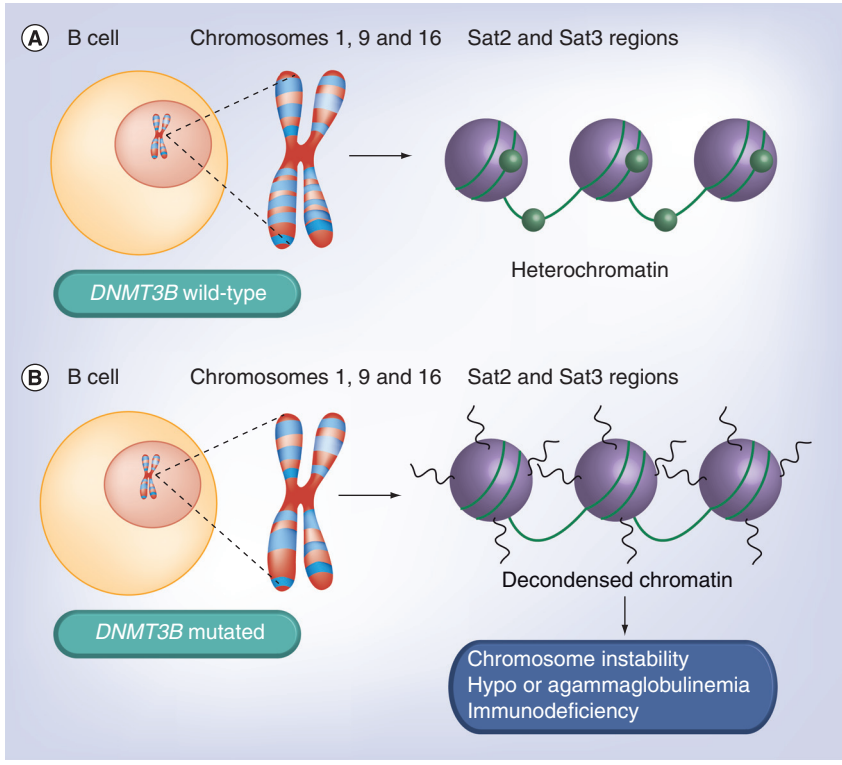


Figure 3. Main chromatin alterations described for immunodeficiency, centromeric instability and facial anomalies syndrome. (A) Sat2 and Sat3 regions are methylated in the presence of normal DNMT3B enzyme. **(B)**. Mutations in *DNMT3B* gene are associated with chromatin decondensation in Sat2 and Sat3 regions of chromosomes 1, 9 and 16. Hypomethylation in Sat regions is detected in several cell types from immunodeficiency, centromeric instability and facial anomalies syndrome patients, but chromosomal abnormalities are mostly detected in mitogen-activated lymphocytes.
Sat: Satellite.

Virally driven epigenetic deregulation in the immune system

Viral infection is a major external factor known to influence multiple immune-related disorders. In fact, different sorts of viral infection are etiological factors for both autoimmune diseases and secondary immunodeficiencies. Several human retroviruses (e.g., HIV) and some human herpes viruses (e.g., Epstein–Barr virus [EBV]) target and infect immune system cells. The viruses alter the function of these cells, ultimately causing immune system defects. Chronic HIV infection causes severe immunosuppression that leads to persistent viremia and progressive clinical disease. In contrast, EBV is immunologically contained after acute symptomatic infection [86]. Some researchers have proposed that cooperation among different viruses might be important in the pathogenesis of virally induced malignancies [87].

Primary infection of B cells by EBV normally occurs in childhood and is asymptomatic, but when it occurs in adolescence, it causes infectious mononucleosis. Some EBV infected B cells are eliminated by cytotoxic T lymphocytes, whereas others evade the immune response. After primary infection, EBV remains latent in the host B cells throughout life [88]. The EBV has two main immune-escape mechanisms during this latent infection: maintaining limited expression of viral proteins and maintaining a low number of viral genomes [89]. Latent EBV infection is related to several autoimmune diseases and to different types of cancer; in this review, we focus on the former.

Many of the autoimmune diseases associated with EBV infection are characterized by the presence of elevated levels of anti-EBV antigen antibodies. These include SLE, RA, MS, Sjögren's syndrome, autoimmune thyroiditis, autoimmune hepatitis, cryptogenic fibrosing alveolitis and pure red aplasia [90], the most prevalent of which are SLE, RA and MS. In the case of SLE and RA, patients exhibit higher titers of anti-EBV antibodies, impaired T cell responses to EBV antigens, and a higher viral load in peripheral blood mononuclear cells [91,92]. The contribution of EBV to MS is demonstrated by

the presence of EBV-specific antibodies in MS patients and by the detection of EBV markers in postmortem brains tissue from patients that had suffered a secondary, progressive phase of MS [93].

EBV latency is epigenetically regulated. Interestingly, EBV genomes, which are packaged into virions, are unmethylated [94]. When EBV infects B cells, its genome is circularized into an episome and organized into chromatin [95], and at certain CpG sites the EBV genome become increasingly methylated, resulting in repression of the associated sequences [94]. Methylation of the EBV genome is driven by cellular DNMTs and occurs in a cell-type-specific manner, signifying that a viral genome may contain distinct epigenetic marks in function of the phenotype of the infected host cell [96]. On the other hand, it is likely that EBV influences the epigenetic status of host cells and this participates in phenotypic changes associated with infected B cells (Figure 4A). For instance, in the latency III program where resting B cells become proliferating lymphoblastoid cells, six EBV nuclear antigens (EBV nuclear antigen [EBNA] -1, -2, -3A, -3B and –LP) and three latent proteins (latent membrane protein-1, -2A and -2B) are expressed. EBNA proteins influence cellular transcription with different elements of the host signaling and transcription machineries. Recent data has shown that EBNA3C and EBNA3A can repress *p16*(*INK4A*) and *p14*(*ARF*) by directly changing their H3K27me3, H3K4me3 and H3-acetylation levels relative to resting B cells, although without altering CpG methylation [97]. In other genes, EBNA3A and EBNA3C also influence the DNA methylation profile at their promoters [98]. Obviously, genome-wide information is required to better understand how the host epigenomic profile is influenced by viral proteins. In this sense, preliminary high-throughput methylation analysis of EBV-transformed lymphoblastoid cell lines suggest that EBV can induce dramatic changes in the DNA methylome of B cells [99].

miRNA deregulation is also associated with EBV infection. Both viral [100] and cellular miRNAs participate in this process. It is also likely that miRNA deregulation is epigenetically deregulated following EBV infection. Viral miRNAs can be involved in EBV infection in two ways: directly, by targeting viral transcripts [101–103]; or indirectly, by targeting cellular factors involved in the EBV infection process [104–106], which can themselves influence on the setting of epigenetic modifications and affect gene expression. Furthermore, EBV infection provokes deregulation of the expression of cellular miRNAs, as has been demonstrated in several studies, most of which have focused on changes in cellular profile of miRNAs during immortalization induced by EBV infection [107–109]. miRNA-155 and miRNA-146a have been described as cellular miRNAs involved in immortalization, and both of them are regulated by the viral protein latent membrane protein-1 [108,110–112]. miRNA-155 targets certain cellular proteins involved in immortalization process (e.g., the bone morphogenetic protein) [113], whereas miRNA-146a is a regulator of interferon-responsive genes [110].

Another fundamental example of viral infection and immune function is HIV. HIV type 1 (HIV-1) is a lentivirus that causes AIDS. This virus establishes latent reservoirs during the early stages of infection, primarily in memory $CD4^+$ T cells, in which proviral DNA integrates into the host genome but is transcriptionally inactive [114]. HIV can also establish chronic, nonproductive, or low level persistent reservoirs in monocyte–macrophage lineage cells, dendritic cells [115], various tissue types, such as the male urogenital tract [116] and the CNS [117], and $CD34^+/CD38^+$ bone marrow progenitor cells [118]. Insertion of the proviral DNA template into the host cell DNA, or integration, is fundamental in the viral lifecycle of HIV-1: it is needed for efficient synthesis of viable progeny [119]. The location of viral integration sites is crucial to the control of transcriptional activity of the integrated proviral genome. When actively replicating, the viral DNA preferentially integrates into active host genes [120].

HIV-1 activity is also modulated by the host epigenetic machinery. Conversely, it has been reported that HIV-1 also influences the activity of infected cells (Figure 4B). The HIV-1 promoter is methylated under latency conditions; thus, DNA methylation has been shown to be crucial for transcriptional regulation of HIV-1 latency [121]. Two CpG islands have been identified in the 5′ region of the HIV provirus. The methylated DNA in these regions is bound to methyl-CpG-binding domain protein 2. DNA methylation seems to happen early on during silencing and occurs during the first 48 h postinfection, which suggests that methylation of proviral DNA is a principal driver of transcriptional silencing of the provirus [115]. HIV proviral DNA is epigenetically regulated

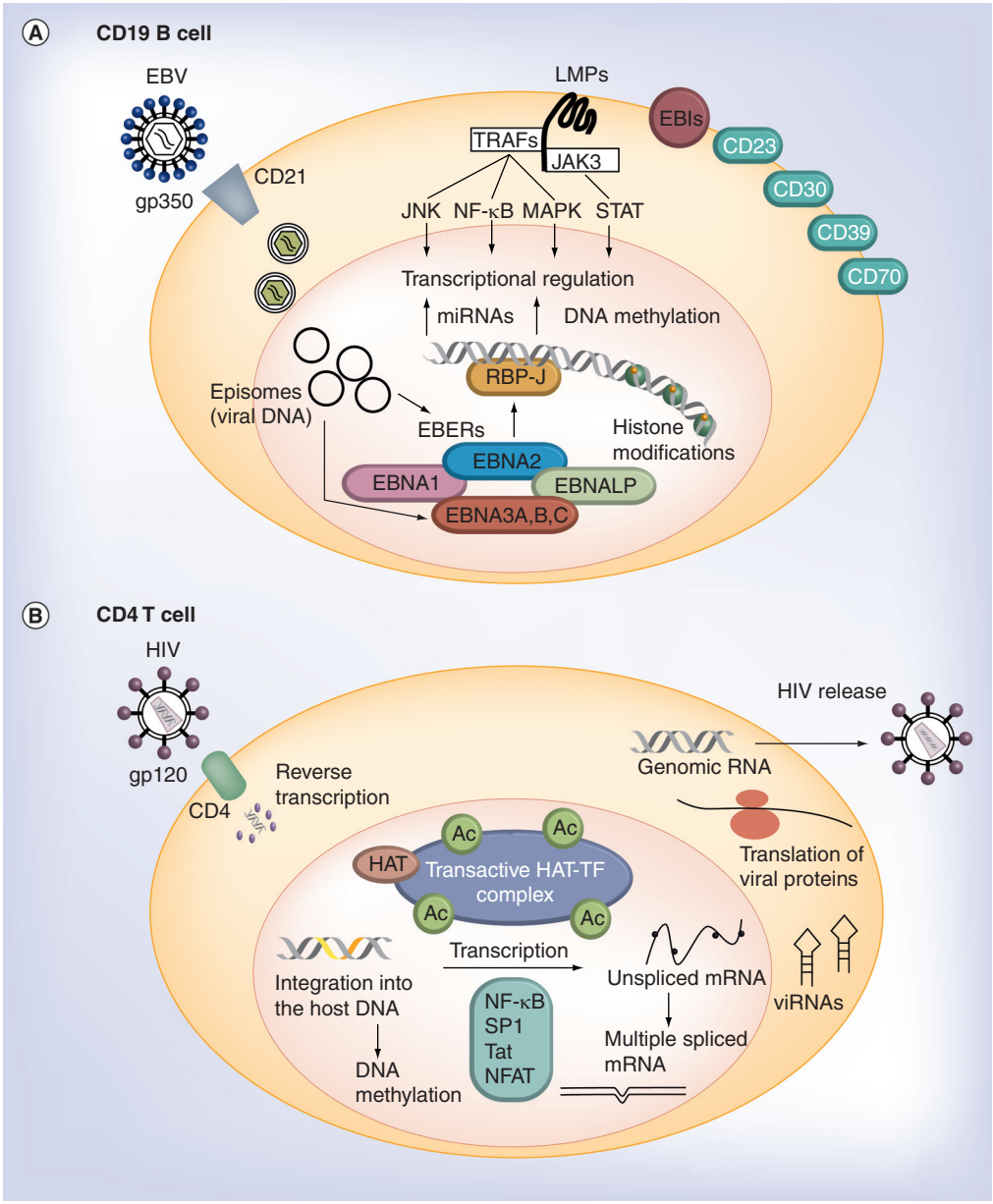


Figure 4. Epstein–Barr virus and HIV infection of B and T lymphocytes. (A) EBV infects B lymphocytes through the binding of the viral gp350 to the CD21 receptor. EBV remains in the B cells as an extrachromosomal viral episome and constitutively expresses a limited set of viral gene products. In the case of the latency III program, six EBV nuclear antigens (EBNAs 1, 2, 3A, 3B, 3C and -LP) and three latent membrane proteins (LMPs 1, 2A and 2B) are expressed. In addition to the latent proteins, two small ncRNAs, EBERs 1 and 2, are expressed. EBNAs influence both viral and cellular transcription through interaction with the DNA-binding protein, RBP-Jk. The cytoplasmic domain of LMPs interacts with TRAFs and with JAK3, resulting in activation of different signaling pathways: NF-κB, MAPK, JNK and STAT. Viral proteins are also able to regulate the gene expression by epigenetic mechanisms such as DNA methylation, histone modifications and miRNAs. **(B)** In the case of HIV, the HIV glycoprotein gp120 binds to the CD4 receptor of T cells leading to the entry of the virus into the host cell. Viral RNA is retrotranscribed into proviral dscDNA, which is transported into the nucleus and is integrated into a host cell chromosome. After integration, the provirus remains quiescent, exists in a permanent postintegration latent state. HIV infection increases the cellular capacity to methylate genes. The recruitment of HATs makes this chromatin permissive for viral genome transcription by the synergic interaction of cellular transcription factors (NF-κB, NFAT and SP1, and the viral transactivators Tat and Rev). New virions are assembled and released through the cell membrane. Viral genomes also produce miRNAs, called viRNAs, which can target viral RNAs and cellular mRNAs or miRNAs.
Ac: Acetylated; EBV: Epstein–Barr virus; gp: glycoprotein; HAT: histone acetyltransferase.

by enzyme complexes, through the covalent modification of the amino termini of core histones and by ATP-dependent enzyme complexes, which control the interactions of DNA with these histones [122]. CpG methylation also controls HIV reactivation from latency [123]. It has also been shown that infection by HIV-1 also increases *de novo* methylation of lymphoid cells in genes such as *p16(INK4A)* [124]. Both observations suggest that drugs that influence DNA methylation could potentially modulate the HIV-1 cycle. HIV-1 has also been demonstrated to influence other elements of the host epigenetic machinery. For instance, HIV-1 Tat protein inhibits the SIRT1 deacetylase and induces T cell hyperactivation [125]. As for most examples referenced in this review, despite the recognition of the participation of epigenetic mechanisms in HIV-1 cycle there is no available data on the effects of HIV-1 integration or reactivation on epigenomic profiles of $CD4^+$ T cells.

HIV-I infection also associates with miRNA deregulation and, in some cases, interaction between miRNAs and epigenetic factors has been identified. For instance, HIV-1 infection results in downregulation of the cellular cluster miRNA-17/92, leading to upregulation of its targets including P300/CBP-associated factor [126], one of the most common histone acetyltransferases, resulting in more efficient viral replication. It is likely that this pathway is involved in deregulation of histone acetylation by the P300/CBP-associated factor. However, there are no available epigenomic data where the effects on distribution of histone acetylation associated with upregulation of the P300/CBP-associated factor under these conditions have been examined. Another example is provided by the HIV-1 TAR element, a short hairpin structure located at the 5′ of HIV viral mRNA. This element is processed by Dicer and produces a viral miRNA that has been proposes to participate in the targeting of HDAC1 to viral long terminal repeat [127]. It is likely that many other HIV-1 encoded miRNAs and cellular miRNAs participate in the deregulation of factors that participate in the establishment of epigenetic marks; however no results are yet available.

As with other factors and diseases discussed in this review, systematic screening of histone modifications, DNA methylation and the resulting effects on protein-coding genes, and miRNA expression will be essential to understand the interactions between these pathogens, hosts and the participation of these elements in disease development.

Conclusion

Compelling evidence indicates a fundamental role for epigenetic deregulation processes in immune-related diseases: autoimmune diseases and immunodeficiencies. Epigenetic deregulation can result as a consequence of genetic defects in key factors that are part or are in close association with elements of the epigenetic machinery. Alternatively, external factors including viral infection can influence epigenetic profiles that ultimately result in transcriptional deregulation. Epigenetic alterations can be used as clinical markers of response or disease progression. Furthermore, epigenetic alterations can be pharmacologically reverted. This potential to tackle aberrant changes opens the possibility of developing novel therapies, some which are already used for the clinical treatment of hematological malignancies. Epigenomic mapping in immune-related disease is still at its beginnings. In autoimmune diseases, most efforts to identify the epigenetic alterations that occur in autoimmune rheumatic disease have focused on SLE and RA, and have served to identify both global and sequence-specific hypomethylation and overexpression of key genes in immune function. Recent high-throughput analyses have allowed the identification of larger sets of genes. However, efforts to identify both DNA methylation and histone modifications changes in specific cell types and particular situations within the course of the disease need to be made. In immunodeficiencies, our knowledge on the participation of epigenetic alterations in their pathogenesis is even more scarce. Interestingly, two rare immunodeficiencies, the ICF syndrome and the HIGM2 syndrome are relevant from the epigenetic point of view. The first one is characterized by mutations in a key enzyme of the epigenetic machinery, DNMT3B. In the case of HIGM2 syndrome, patients are characterized by mutations in activation-induced deaminase, which has recently been reported to participate in active DNA demethylation. The role of viral infection in inducing epigenetic alterations is also relevant, as different viruses such as EBV or HIV are directly associated with the pathogenesis of both autoimmune diseases and immunodeficiency, and their pathogenic activity involves interference and user of the host epigenetic machinery.

Future perspective

Our understanding of the role of epigenetic alterations in immune-related diseases is still at its beginnings. The availability of high-throughput techniques to investigate both DNA methylation and histone modification profiles opens up a number of new possibilities. In the next few years we will witness the systematic generation of DNA methylation and histone modification profiles in a high number of samples in different stages of the course of disease, a variety of tissue types and the integration of these massive amount of data with expression data and with genome-wide association studies. The potential of these analyses will serve to track all potential specific cell types relevant to disease pathogenesis, and a more prominent use of this information in a clinical setting.

Financial & competing interests disclosure

E Ballestar is supported by PI081346 (FIS) grant from the Spanish Ministry of Science and Innovation (MICINN) and 2009SGR184 grant from AGAUR (Catalan Government). ***The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.***

No writing assistance was utilized in the production of this manuscript.

Executive summary

- The immune system encompasses different cells and molecules that protect the host organism by recognizing and eliminating potentially harmful internal or external agents.
- Differentiation and activation of immune cells are very tightly regulated and they are controlled at different stages of development to ensure an efficient immune response that maintains the delicate balance between self-recognition and attack of foreign molecules or microorganisms.
- Proper differentiation and activation partly depends on the genetic integrity of the genes encoding for these key factors. It is also associated with an environmental component and viral infection, exposure to chemicals or hormones influence their development.
- Environmental factors can influence the epigenetic profile of cells, and therefore, have a direct effect on gene-expression profiles.
- Mutations or specific polymorphisms in transcription factors, elements of the epigenetic machinery or cell signaling pathways can also influence the epigenome of immune cells.

Epigenetic modification regulates gene function

- DNA methylation and histone modifications are associated with gene-expression profiles.
- There are intimate connections between epigenetic modifications, transcription factors and cell signaling pathways.

Epigenetic deregulation in autoimmune disease

- Autoimmune diseases are genetically complex diseases that arise in a given genetic background exposed to specific environmental factors.
- Systemic lupus erythematosus and rheumatoid arthritis are the best studied autoimmune diseases from the epigenetic point of view. However, our knowledge is still limited. Only a few candidate gene-based analyses and high-throughput methylation studies have been performed. In systemic lupus erythematosus and rheumatoid arthritis, global hypomethylation has been determined in different lymphocyte sets. Also, hypomethylation is the most common feauture for gene promoters.
- In rheumatoid arthritis, epigenetic deregulation in synovial fibroblasts participates in the generation of their aberrant phenotype.

Epigenomics & immunodeficiency

- Primary immunodeficiencies have both a genetic and an environmental component.
- Immunodeficiency, centromeric instability and facial anomalies syndrome is a rare primary deficiency disease caused by mutations in DNA methyltransferase-3B and changes in the methylation status of pericentromeric chromatin. It is likely that changes in methylation patterns in immunodeficiency, centromeric instability and facial anomalies syndrome patients is responsible for chromosomal instability.
- Hyper-IgM 2 syndrome is associated with mutations in the activation-induced cytidine deaminase gene, encoding activation-induced deaminase, an enzyme with putative demethylating activity.

Virally driven epigenetic deregulation in the immune system

- HIV and Epstein–Barr virus infection are critical in the development of several immune-related disorders.
- HIV and Epstein–Barr virus latency and reactivation cycles are influenced by the epigenetic activity of host cells.
- HIV and Epstein–Barr virus influence the epigenome of host cells and affect their function.

Conclusion

- The existence of a wide range of epigenetic alterations in immune-associated disease is well established.

Future perspective

- Availability of high-throughput techniques for different epigenetic modifications in different cell types, stages or courses of disease and large numbers of samples will allow dissection of the specific role of these alterations in disease.
- Novel compounds modulating the activity of elements of the epigenetic machinery will be used for the reversion of epigenetic alterations.

References

Papers of special note have been highlighted as:
▪▪ of considerable interes

1 Abbas A, Janeway CA Jr. Immunology: improving on nature review in the twenty-first century. *Cell* 100(1), 129–138 (2000).

2 Medzhitov R, Janeway CA Jr. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91(3), 295–298 (1997).

3 Medzhitov R, Janeway CA Jr. Innate Immunity. *N. Engl. J. Med*. 343, 338–344 (2000).

4 Cruvinel Wde M, Mesquita D Jr, Araújo JA *et al.* Immune system – part I. Fundamentals of innate immunity with emphasis on molecular and cellular mechanisms of inflammatory response. *Rev. Bras. Reumatol.* 50(4), 434–461 (2010).

5 Bonilla F, Oettgen HC. Adaptive immunity. *J. Allergy.Clin. Immunol*. 125(2), S33–S40 (2010).

6 Medzdhitov R, Janeway CA Jr. Innate immunity: impact on the adaptive immune response. *Curr. Opin. Immunol.* 9(1), 4–9 (1997).

7 Orkin SH. Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev*. 1(1), 57–64 (2000).

8 Hewagama A, Richardson B. The genetics and epigenetics of autoimmune diseases. *J. Autoimmunity* 33(1), 3–11 (2009).

9 Delgado-Vega A, Sánchez E, Löfgren S, Castillejo-López C, Alarcón-Riquelme ME. Recent findings on genetics of systemic autoimmune diseases. *Curr. Opin. Immunol*. 22(6), 698–705 (2010).

10 Conley ME, Dobbs AK, Farmer DM *et al.* Primary B cell immunodeficiencies: comparisons and contrasts. *Ann. Rev. Immunol*. 27, 199–227 (2009).

11 Esteller M. Epigenetics in cancer. *N. Engl. J. Med.* 358(11), 1148–1159 (2008).

12 Irizarry RA, Ladd-Acosta C, Wen B *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet*. 41(2), 178–186 (2009).

13 Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem*. 74, 481–514 (2005).

14 Garcia-Manero G. Demethylating agents in myeloid malignancies. *Curr. Opin. Oncol.* 20(6), 705–710 (2008).

15 Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463(7284), 1042–1047 (2010).

16 Bruniquel D, Schwartz RH. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. *Nat. Immunol.* 4(3), 235–240 (2003).

17 Ooi SK, Bestor TH. The colorful history of active DNA demethylation. *Cell* 133(7), 1145–1148 (2008).

18 Fritz EL, Papavasiliou FN. Cytidine deaminases: AIDing DNA demethylation? *Genes Dev*. 24(19), 2107–2114 (2010).

19 Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135(7), 1201–1212 (2008).

20 Tahiliani M, Koh KP, Shen Y *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324 (5929), 930–935 (2009).

21 Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466(7310), 1129–1133 (2010).

22 Ito S, Shen L, Dai Q *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333(6047), 1300–1303 (2011).

23 Kouzarides T. Chromatin modifications and their function. *Cell* 128(4), 693–705 (2007).

24 Allis CD, Berger SL, Cote J *et al.* New nomenclature for chromatin-modifying enzymes. *Cell* 131(4), 633–636 (2007).

25 Biancotto C, Frigè G, Minucci S. Histone modification therapy of cancer. *Adv. Genet.* 70, 341–386 (2010).

26 Ballestar E, Esteller M. The impact of chromatin in human cancer. linking DNA methylation to gene silencing. *Carcinogenesis* 23, 1103–1109 (2002).

27 Müller J, Verrijzer P. Biochemical mechanisms of gene regulation by Polycomb group protein complexes. *Curr. Opin. Genet. Dev.* 19, 150–158 (2009).

28 Fabbri M, Calin GA. Epigenetics and miRNAs in human cancer. *Adv. Genet.* 70, 87–99 (2010).

29 Davidson A, Diamond B. Autoimmune diseases. *N. Engl. J. Med.* 345(5), 340–350 (2001).

30 Goodnow CC, Sprent J, Fazekas de St Groth B, Vinuesa CG. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* 435(7042), 590–597 (2005).

31 Lie BA, Thorsby E. Several genes in the extended human MHC contribute to predisposition to autoimmune diseases. *Curr. Opin. Immunol.* 17(5), 526–531 (2005).

32 Delgado-Vega A, Sanchez E, Lofgren S, Castillejo-Lopez C, Alarcon-Riquelme ME. Recent findings on genetics of systemic autoimmune diseases. *Curr. Opin. Immunol.* 22(6), 698–705 (2010).

33 Silman AJ, Macgregor AJ, Thomson W *et al.* Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *Br. J. Rheumatol.* 32(10), 903–907 (1993).

34 Grennan DM, Parfitt A, Manolios N *et al.* Family and twin studies in systemic lupus erythematosus. *Dis. Markers* 13(2), 93–98 (1997).

35 Aho K, Koskenvuo M, Tuominen J, Kaprio J. Occurrence of rheumatoid arthritis in a nationwide series of twins. *J. Rheumatol.* 13(5), 899–902 (1986).

36 Richardson B. Effect of an inhibitor of DNA methylation on T cells. II. 5-azacytidine induces self-reactivity in antigen-specific $T4^+$-cells. *Hum. Immunol.* 17(4), 456–470 (1986).

37 Cornacchia E, Golbus J, Maybaum J, Strahler J, Hanash S, Richardson B. Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity. *J. Immunol.* 140(7), 2197–2200 (1988).

38 Richardson BC, Strahler JR, Pivirotto TS *et al.* Phenotypic and functional similarities between 5-azacytidine-treated T cells and a T cell subset in patients with active systemic lupus erythematosus. *Arthritis Rheum*. 35(6), 647–662 (1992).

39 Ballestar E. Epigenetic alterations in autoimmune rheumatic diseases. *Nat. Rev. Rheumatol*. 7(5), 263–271 (2011).

40 Richardson B, Scheinbart L, Strahler J, Gross L, Hanash S, Johnson M. Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Rheum.* 33(11), 1665–1673 (1990).

41 Harmon CE, Portanova JP. Drug-induced lupus: clinical and serological studies. *Clin. Rheum. Dis.* 8(1), 121–135 (1982).

42 Lei W, Luo Y, Lei W *et al.* Abnormal DNA methylation in $CD4^+$ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. *Scand. J. Rheumatol.* 38(5), 369–374 (2009).

43 Quddus J, Johnson KJ, Gavalchin J *et al.* Treating activated $CD4^+$ T cells with either of two distinct DNA methyltransferase inhibitors, 5-azacytidine or procainamide, is sufficient to cause a lupus-like disease in syngeneic mice. *J. Clin. Invest.* 92(1), 38–53 (1993).

44 Javierre BM, Fernandez AF, Richter J *et al.* Changes in the pattern of DNA methylation associate with twin discordance in systemic

lupus erythematosus. *Genome Res.* 20(2), 170–179 (2010).

▪▪ **Represents the first high-throughput DNA methylation analysis on systemic lupus erythematosus. The comparison of monozygotic twins allows to identify DNA methylation differences without interference from genetic variation.**

45 Jeffries MA, Dozmorov M, Tang Y, Merrill JT, Wren JD, Sawalha AH. Genome-wide DNA methylation patterns in CD4$^+$ T cells from patients with systemic lupus erythematosus. *Epigenetics* 6(5), 593–601 (2011).

▪▪ **Represents the first high-throughput DNA methylation analysis on a separated cell population in systemic lupus erythematosus, specifically CD4$^+$ cells.**

46 Neumann E, Lefevre S, Zimmermann B, Gay S, Muller-Ladner U. Rheumatoid arthritis progression mediated by activated synovial fibroblasts. *Trends Mol. Med.* 16(10), 458–468 (2010).

47 Corvetta A, Della Bitta R, Luchetti Mm, Pomponio G. 5-methylcytosine content of DNA in blood, synovial mononuclear cells and synovial tissue from patients affected by autoimmune rheumatic diseases. *J. Chromatogr.* 566(2), 481–491 (1991).

48 Neidhart M, Rethage J, Kuchen S *et al.* Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. *Arthritis Rheum.* 43(12), 2634–2647 (2000).

▪▪ **Role of hypomethylation in the phenotype of rheumatoid arthritis synovial fibroblasts.**

49 Karouzakis E, Gay RE, Michel BA, Gay S, Neidhart M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum.* 60(12), 3613–3622 (2009).

50 Neidhart M *et al.* Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue. association with genomic DNA hypomethylation and influence on gene expression. *Arthritis Rheum.* 43, 2634–2647 (2000).

51 Karouzakis E, Gay RE, Michel BA, Gay S, Neidhart M. DNA hypomethylation in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum.* 60, 3613–3622 (2009).

52 Kuchen S, Seemayer CA, Rethage J *et al.* The L1 retroelement-related p40 protein induces p38δ MAP kinase. *Autoimmunity* 37(1), 57–65 (2004).

53 Stanczyk J, Ospelt C, Karouzakis E *et al.* Altered expression of microRNA-203 in rheumatoid arthritis synovial fibroblasts and its role in fibroblast activation. *Arthritis Rheum.* 63(2), 373–381 (2011).

54 Takami N, Osawa K, Miura Y *et al.* Hypermethylated promoter region of DR3, the death receptor 3 gene, in rheumatoid arthritis synovial cells. *Arthritis Rheum.* 54(3), 779–787 (2006).

55 Hu N, Qiu X, Luo Y *et al.* Abnormal histone modification patterns in lupus CD4$^+$ T cells. *J. Rheumatol.* 35(5), 804–810 (2008).

56 Mishra N, Brown DR, Olorenshaw IM, Kammer GM. Trichostatin A reverses skewed expression of CD154, interleukin-10, and interferon-γ gene and protein expression in lupus T cells. *Proc. Natl Acad. Sci.* USA 98(5), 2628–2633 (2001).

57 Kawabata T, Nishida K, Takasugi K *et al.* Increased activity and expression of histone deacetylase 1 in relation to tumor necrosis factor-α in synovial tissue of rheumatoid arthritis. *Arthritis Res. Ther.* 12(4), R133 (2010).

58 Horiuchi M, Morinobu A, Chin T, Sakai Y, Kurosaka M, Kumagai S. Expression and function of histone deacetylases in rheumatoid arthritis synovial fibroblasts. *J. Rheumatol.* 36(8), 1580–1589 (2009).

59 Chung YL, Lee MY, Wang AJ, Yao LF. A therapeutic strategy uses histone deacetylase inhibitors to modulate the expression of genes involved in the pathogenesis of rheumatoid arthritis. *Mol. Ther.* 8(5), 707–717 (2003).

60 Nishida K, Komiyama T, Miyazawa S *et al.* Histone deacetylase inhibitor suppression of autoantibody-mediated arthritis in mice via regulation of p16INK4a and p21(WAF1/Cip1) expression. *Arthritis Rheum.* 50(10), 3365–3376 (2004).

61 Lin HS, Hu CY, Chan HY *et al.* Antirheumatic activities of histone deacetylase (HDAC) inhibitors *in vivo* in collagen-induced arthritis in rodents. *Br. J. Pharmacol.* 150(7), 862–872 (2007).

62 Nasu Y, Nishida K, Miyazawa S *et al.* Trichostatin A, a histone deacetylase inhibitor, suppresses synovial inflammation and subsequent cartilage destruction in a collagen antibody-induced arthritis mouse model. *Osteoarthritis Cartilage* 16(6), 723–732 (2008).

63 Choo QY, Ho PC, Tanaka Y, Lin HS. Histone deacetylase inhibitors MS-275 and SAHA induced growth arrest and suppressed lipopolysaccharide-stimulated NF-κB p65 nuclear accumulation in human rheumatoid arthritis synovial fibroblastic E11 cells. *Rheumatology (Oxford)* 49(8), 1447–1460 (2010).

64 Jüngel A, Baresova V, Ospelt C *et al.* Trichostatin A sensitizes rheumatoid arthritis synovial fibroblasts for TRAIL-induced apoptosis. *Ann. Rheum. Dis.* 65(7), 910–912 (2006).

65 Morinobu A, Wang B, Liu J, Yoshiya S, Kurosaka M, Kumagai S. Trichostatin A cooperates with Fas-mediated signal to induce apoptosis in rheumatoid arthritis synovial fibroblasts. *J. Rheumatol.* 33(6), 1052–1060 (2006).

66 Zhou X, Hua X, Ding X, Bian Y, Wang X. Trichostatin differentially regulates Th1 and Th2 responses and alleviates rheumatoid arthritis in mice. *J. Clin. Immunol.* 31(3), 395–405 (2011).

67 Schroeder TM, Nair AK, Staggs R, Lamblin AF, Westendorf JJ. Gene profile analysis of osteoblast genes differentially regulated by histone deacetylase inhibitors. *BMC Genomics* 8, 362 (2007).

68 Nakamura T, Kukita T, Shobuike T *et al.* Inhibition of histone deacetylase suppresses osteoclastogenesis and bone destruction by inducing IFN-β production. *J. Immunol.* 175(9), 5809–5816 (2005).

69 Vojinovic J, Damjanov N, D'urzo C *et al.* Safety and efficacy of an oral histone deacetylase inhibitor in systemic-onset juvenile idiopathic arthritis. *Arthritis Rheum.* 63(5), 1452–1458 (2011).

70 Trenkmann M, Brock M, Gay RE *et al.* Expression and function of EZH2 in synovial fibroblasts. Epigenetic repression of the Wnt inhibitor SFRP1 in rheumatoid arthritis. *Ann. Rheum. Dis.* 70, 1482–1488 (2011).

71 Baranzini SE, Mudge J, van Velkinburgh JC *et al.* Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* 464(7293), 1351–1356 (2010).

▪▪ **First high-throughput DNA methylation, expression and SNP analysis on separated CD4$^+$ cells in multiple sclerosis discordant twins.**

72 Maródi L, Notarangelo LD. Immunological and genetic bases of new primary immunodeficiencies. *Nat. Rev. Immunol.* 7(11), 851–861 (2007).

73 Chinen J, Shearer WT. Secondary immunodeficiencies, including HIV infection. *J. Allergy Clin. Immunol.* 125(2), S195–S203 (2010).

74 Notarangelo LD. Primary immunodeficiencies. *J. Allergy Clin. Immunol.* 125(2), S182–S194 (2010).

75 Hansen RS, Wijmenga C, Luo P *et al.* The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl Acad. Sci. USA* 96, 14412–14417 (1999).

▪▪ **Seminal study where immunodeficiency, centromeric instability and facial anomalies syndrome was associated with mutations in DNA methytransferase-3B, a key epigenetic enzyme responsible for *de novo* methylation.**

76 Ehrlich M, Sanchez C, Shao C *et al.* ICF, an immunodeficiency syndrome: DNA methyltransferase-3B involvement, chromosome anomalies, and gene dysregulation. *Autoimmunity* 41, 253–271 (2008).

77 Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514 (2005).

78 Jin B, Tao Q, Peng J. DNA methyltransferase-3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum. Mol. Genet.* 17(5), 690–709 (2008).

79 Erlich M, Buchanan F, Tsien G *et al.* DNA methyltransferase-3B mutations linked to the ICF syndrome cause dysregulation of lymphocyte migration, activation and survival genes. *Hum. Mol. Genet.* 10(25), 2917–2931 (2001).

80 Gatto S, Ragione FD, Cimmino A *et al.* Epigenetic alteration of microRNAs in DNMT3B-mutated patients of ICF syndrome. *Epigenetics* 5(5), 427–443 (2010).

81 Revy P, Muto E, Levy Y *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). *Cell* 102(5), 565–575 (2000).

82 Muramatsu M, Sankaranand VS, Anant S *et al.* Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* 274(26), 18470–18476 (1999).

83 Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues. *J. Biol. Chem.* 279(50), 52353–52360 (2004).

84 Popp C, Dean W, Feng S *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by aid deficiency. *Nature* 463(7284), 1101–1105 (2010).

85 Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SY, Blau HM. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463(7284), 1042–1047 (2010).

▪▪ **Role of activation-induced cytidine deaminase as a demethylating activity during reprogramming.**

86 Deeks SG, Walker BD. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* 27(3), 406–416 (2007).

87 Da Silva SR, De Oliveira DE. HIV, EBV and KSHV: viral cooperation in the pathogenesis of human malignancies. *Cancer Lett.* 305(2), 175–185 (2011).

88 Young LS, Rickinson AB. Epstein–Barr virus: 40 years on. *Nat. Rev. Cancer.* 4(10), 757–768 (2004).

89 Yin Y, Manoury B, Fahraeus R. Self-inhibition of synthesis and antigen presentation by Epstein–Barr virus-encoded EBNA1. *Science* 301(5638), 1371–1374 (2003).

90 Pender MP. Infection of autoreactive B lymphocytes with EBV, causing chronic autoimmune diseases. *Trends Immunol.* 24(11), 584–588 (2003).

91 James JA, Neas BR, Moser KL *et al.* Systemic lupus erythematosus in adults is associated with previous Epstein–Barr virus exposure. *Arthritis Rheum.* 44(5), 1122–1126 (2001).

92 Poole BD, Scofield RH, Harley JB, James JA. Epstein–Barr virus and molecular mimicry in systemic lupus erythematosus. *Autoimmunity* 39(1), 63–70 (2006).

93 Serafini B, Severa M, Columba-Cabezas S *et al.* Epstein–Barr virus latent infection and *BAFF* expression in B cells in the multiple sclerosis brain: implications for viral persistence and intrathecal B cell activation. *J. Neuropathol. Exp. Neurol.* 69(7), 677–693 (2010).

94 Kintner C, Sugden B. Conservation and progressive methylation of Epstein–Barr viral DNA sequences in transformed cells. *J. Virol.* 38(1), 305–316 (1981).

95 Dyson PJ, Farrell PJ. Chromatin structure of Epstein–Barr virus. *J. Gen. Virol.* 66(Pt 9), 1931–1940 (1985).

96 Minarovits J, Minarovits-Kormuta S, Ehlin-Henriksson B, Falk K, Klein G, Ernberg I. Host cell phenotype-dependent methylation patterns of Epstein–Barr virus DNA. *J. Gen. Virol.* 72(Pt 7), 1591–1599 (1991).

97 Maruo S, Zhao B, Johannsen E, Kieff E, Zou J, Takada K. Epstein–Barr virus nuclear antigens 3C and 3A maintain lymphoblastoid cell growth by repressing *p16INK4A* and *p14ARF* expression. *Proc. Natl Acad. Sci. USA* 108(5), 1919–1924 (2011).

98 Paschos K, Smith P, Anderton E, Middeldorp JM, White RE, Allday MJ. Epstein–Barr virus latency in B cells leads to epigenetic repression and CpG methylation of the tumour suppressor gene *Bim*. *PLoS Pathog.* 5(6), e1000492 (2009).

99 Caliskan M, Cusanovich DA, Ober C, Gilad Y. The effects of EBV transformation on gene expression levels and methylation profiles. *Hum. Mol. Genet.* 20(8), 1643–1652 (2011).

100 Pfeffer S, Zavolan M, Grasser FA *et al.* Identification of virus-encoded microRNAs. *Science* 304(5671), 734–736 (2004).

101 Barth S, Pfuhl T, Mamiani A *et al.* Epstein–Barr virus-encoded microRNA miR-BART2 downregulates the viral DNA polymerase BALF5. *Nucleic Acids. Res.* 36(2), 666–675 (2008).

102 Lo AK, To KF, Lo KW *et al.* Modulation of LMP1 protein expression by EBV-encoded microRNAs. *Proc. Natl Acad. Sci. USA* 104(41), 16164–16169 (2007).

103 Lung RW, Tong JH, Sung YM *et al.* Modulation of *LMP2A* expression by a newly identified Epstein–Barr virus-encoded microRNA miR-BART22. *Neoplasia* 11(11), 1174–1184 (2009).

104 Choy EY, Siu KL, Kok KH *et al.* An Epstein–Barr virus-encoded microRNA targets PUMA to promote host cell survival. *J. Exp. Med.* 205(11), 2551–2560 (2008).

105 Marquitz AR, Mathur A, Nam CS, Raab-Traub N. The Epstein–Barr virus BART microRNAs target the proapoptotic protein Bim. *Virology* 412(2), 392–400 (2011).

106 Xia T, O'hara A, Araujo I *et al.* EBV microRNAs in primary lymphomas and targeting of CXCL-11 by EBV-miR-BHRF1–3. *Cancer Res.* 68(5), 1436–1442 (2008).

107 Cameron JE, Fewell C, Yin Q *et al.* Epstein–Barr virus growth/latency III program alters cellular microRNA expression. *Virology* 382(2), 257–266 (2008).

108 Linnstaedt SD, Gottwein E, Skalsky RL, Luftig MA, Cullen BR. Virally induced cellular microRNA miR-155 plays a key role in B cell immortalization by Epstein–Barr virus. *J. Virol.* 84(22), 11670–11678 (2010).

109 Mrazek J, Kreutmayer SB, Grasser FA, Polacek N, Huttenhofer A. Subtractive hybridization identifies novel differentially expressed ncRNA species in EBV-infected human B cells. *Nucleic Acids Res.* 35(10), e73 (2007).

110 Cameron JE, Yin Q, Fewell C *et al.* Epstein–Barr virus latent membrane protein 1 induces cellular miRNA miR-146a, a modulator of lymphocyte signaling pathways. *J. Virol.* 82(4), 1946–1958 (2008).

111 Gatto G, Rossi A, Rossi D, Kroening S, Bonatti S, Mallardo M. Epstein–Barr virus latent membrane protein 1 trans-activates miR-155 transcription through the NF-κB pathway. *Nucleic Acids Res.* 36(20), 6608–6619 (2008).

112 Motsch N, Pfuhl T, Mrazek J, Barth S, Grasser FA. Epstein–Barr virus-encoded latent membrane protein 1 (LMP1) induces the expression of the cellular miRNA miR-146a. *RNA Biol.* 4(3), 131–137 (2007).

113 Yin Q, Wang X, Fewell C *et al.* MiRNA miR-155 inhibits bone morphogenetic protein (BMP) signaling and BMP-mediated Epstein–Barr virus reactivation. *J. Virol.* 84(13), 6318–6327 (2010).

114 Pomerantz RJ, Horn DL. Twenty years of therapy for HIV-1 infection. *Nat. Med.* 9(7), 867–873 (2003).

115 Coleman CM, Wu L. HIV interactions with monocytes and dendritic cells: viral latency and reservoirs. *Retrovirology* 6, 51 (2009).

116 Zhu T, Wang N, Carr A *et al.* Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. *J. Virol.* 70(5), 3098–3107 (1996).

117 Cheng-Mayer C, Weiss C, Seto D, Levy JA. Isolates of human immunodeficiency virus type 1 from the brain may constitute a special group of the AIDS virus. *Proc. Natl Acad. Sci. USA* 86(21), 8575–8579 (1989).

118 Alexaki A, Wigdahl B. HIV-1 infection of bone marrow hematopoietic progenitor cells and their role in trafficking and viral dissemination. *PLoS Pathog.* 4(12), e1000215 (2008).

119 Liszewski MK, Yu JJ, O'doherty U. Detecting HIV-1 integration by repetitive-sampling Alu-gag PCR. *Methods* 47(4), 254–260 (2009).

120 Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110(4), 521–529 (2002).

121 Kauder SE, Bosque A, Lindqvist A, Planelles V, Verdin E. Epigenetic regulation of HIV-1 latency by cytosine methylation. *PLoS Pathog* 5(6), e1000495 (2009).

122 El Kharroubi A, Piras G, Zensen R, Martin MA. Transcriptional activation of the integrated chromatin-associated human immunodeficiency virus type 1 promoter. *Mol. Cell. Biol.* 18(5), 2535–2544 (1998).

123 Blazkova J, Trejbalova K, Gondois-Rey F *et al.* CpG methylation controls reactivation of HIV from latency. *PLoS Pathog.* 5(8), e1000554 (2009).

124 Fang JY, Mikovits JA, Bagni R, Petrow-Sadowski CL, Ruscetti FW. Infection of lymphoid cells by integration-defective human immunodeficiency virus type 1 increases *de novo* methylation. *J. Virol.* 75(20), 9753–9761 (2001).

125 Kwon HS, Brent MM, Getachew R *et al.* Human immunodeficiency virus type 1 Tat protein inhibits the SIRT1 deacetylase and induces T cell hyperactivation. *Cell Host Microbe* 3(3), 158–167 (2008).

126 Triboulet R, Mari B, Lin YL *et al.* Suppression of microRNA-silencing pathway by HIV-1 during virus replication. *Science* 315(5818), 1579–1582 (2007).

▪▪ **Role of HIV-1 in promoting interactions between miRNA and histone modification is presented.**

127 Klase Z, Kale P, Winograd R *et al.* HIV-1 TAR element is processed by Dicer to yield a viral miRNA involved in chromatin remodeling of the viral LTR. *BMC Mol. Biol.* 8, 63 (2007).

128 Lu Q, Kaplan M, Ray D *et al.* Demethylation of ITGAL (CD11a) regulatory sequences in systemic lupus erythematosus. *Arthritis Rheum.* 46(5), 1282–1291 (2002).

129 Lu Q, Wu A, Tesmer L, Ray D, Yousif N, Richardson B. Demethylation of CD40LG on the inactive X in T cells from women with lupus. *J. Immunol.* 179(9), 6352–6358 (2007).

130 Lu Q, Wu A, Richardson BC. Demethylation of the same promoter sequence increases CD70 expression in lupus T cells and T cells treated with lupus-inducing drugs. *J. Immunol.* 174(10), 6212–6219 (2005).

131 Kaplan Mj, Lu Q, Wu A, Attwood J, Richardson B. Demethylation of promoter regulatory elements contributes to perforin overexpression in $CD4^+$ lupus T cells. *J. Immunol.* 172(6), 3652–3661 (2004).

132 Garaud S, Le Dantec C, Jousse-Joulin S *et al.* IL-6 modulates CD5 expression in B cells from patients with lupus by regulating DNA methylation. *J. Immunol.* 182(9), 5623–5632 (2009).

133 Nile CJ, Read RC, Akil M, Duff GW, Wilson AG. Methylation status of a single CpG site in the IL-6 promoter is related to IL-6 messenger RNA levels and rheumatoid arthritis. *Arthritis Rheum.* 58(9), 2686–2693 (2008).

134 Mastronardi FG, Noor A, Wood DD, Paton T, Moscarello MA. Peptidyl arginine-deiminase 2 CpG island in multiple sclerosis white matter is hypomethylated. *J. Neurosci. Res.* 85(9), 2006–2016 (2007).

135 Gonzalez S, Aguilera S, Alliende C *et al.* Alterations in type I hemidesmosome components suggestive of epigenetic control in the salivary glands of patients with Sjogren's syndrome. *Arthritis Rheum.* 63(4), 1106–1115 (2011).

136 Lei W, Luo Y, Lei W *et al.* Abnormal DNA methylation in $CD4^+$ T cells from patients with systemic lupus erythematosus, systemic sclerosis, and dermatomyositis. *Scand. J. Rheumatol.* 38(5), 369–374 (2009).

137 Wang Y, Fan Ps, Kahaleh B. Association between enhanced type I collagen expression and epigenetic repression of the *FLI1* gene in scleroderma fibroblasts. *Arthritis Rheum.* 54(7), 2271–2279 (2006).

138 Ozbalkan Z, Bagislar S, Kiraz S *et al.* Skewed X-chromosome inactivation in blood cells of women with scleroderma. *Arthritis Rheum.* 52(5), 1564–1570 (2005).

139 Chen M, Chen ZQ, Cui PG *et al.* The methylation pattern of *p16INK4a* gene promoter in psoriatic epidermis and its clinical significance. *Br. J. Dermatol.* 158(5), 987–993 (2008).

140 Zhang P, Su Y, Chen H, Zhao M, Lu Q. Abnormal DNA methylation in skin lesions and PBMCs of patients with psoriasis vulgaris. *J. Dermatol. Sci.* 60(1), 40–42 (2010).

141 Ruchusatsawat K, Wongpiyabovorn J, Shuangshoti S, Hirankarn N, Mutirangura A. SHP-1 promoter 2 methylation in normal epithelial tissues and demethylation in psoriasis. *J. Mol. Med.* 84(2), 175–182 (2006).

142 Zhang K, Zhang R, Li X, Yin G, Niu X. Promoter methylation status of *p15* and *p21* genes in HPP-CFCs of bone marrow of patients with psoriasis. *Eur. J. Dermatol.* 19(2), 141–146 (2009).

143 Zhang K, Zhang R, Li X, Yin G, Niu X, Hou R. The mRNA expression and promoter methylation status of the *p16* gene in colony-forming cells with high proliferative potential in patients with psoriasis. *Clin. Exp. Dermatol.* 32(6), 702–708 (2007).

144 Gloria L, Cravo M, Pinto A *et al.* DNA hypomethylation and proliferative activity are increased in the rectal mucosa of patients with long-standing ulcerative colitis. *Cancer* 78(11), 2300–2306 (1996).

145 Hsieh CJ, Klump B, Holzmann K, Borchard F, Gregor M, Porschen R. Hypermethylation of the p16INK4a promoter in colectomy specimens of patients with long-standing and extensive ulcerative colitis. *Cancer Res.* 58(17), 3942–3945 (1998).

146 Sato F, Harpaz N, Shibata D *et al.* Hypermethylation of the *p14(ARF)* gene in ulcerative colitis-associated colorectal carcinogenesis. *Cancer Res.* 62(4), 1148–1151 (2002).

147 Tahara T, Shibata T, Nakamura M *et al.* Effect of *MDR1* gene promoter methylation in patients with ulcerative colitis. *Int. J. Mol. Med.* 23(4), 521–527 (2009).

148 Brix TH, Knudsen GP, Kristiansen M, Kyvik KO, Orstavik KH, Hegedus L. High frequency of skewed X-chromosome inactivation in females with autoimmune thyroid disease: a possible explanation for the female predisposition to thyroid autoimmunity. *J. Clin. Endocrinol. Metab.* 90(11), 5949–5953 (2005).

Genome Biology

RESEARCH Open Access

The B cell transcription program mediates hypomethylation and overexpression of key genes in Epstein-Barr virus-associated proliferative conversion

Henar Hernando[1], Claire Shannon-Lowe[2], Abul B Islam[3], Fatima Al-Shahrour[4], Javier Rodríguez-Ubreva[1], Virginia C Rodríguez-Cortez[1], Biola M Javierre[1], Cristina Mangas[5], Agustín F Fernández[5], Maribel Parra[6], Henri-Jacques Delecluse[7], Manel Esteller[8], Eduardo López-Granados[9], Mario F Fraga[5,10], Nuria López-Bigas[3] and Esteban Ballestar[1*]

Abstract

Background: Epstein-Barr virus (EBV) infection is a well characterized etiopathogenic factor for a variety of immune-related conditions, including lymphomas, lymphoproliferative disorders and autoimmune diseases. EBV-mediated transformation of resting B cells to proliferating lymphoblastoid cells occurs in early stages of infection and is an excellent model for investigating the mechanisms associated with acquisition of unlimited growth.

Results: We investigated the effects of experimental EBV infection of B cells on DNA methylation profiles by using high-throughput analysis. Remarkably, we observed hypomethylation of around 250 genes, but no hypermethylation. Hypomethylation did not occur at repetitive sequences, consistent with the absence of genomic instability in lymphoproliferative cells. Changes in methylation only occurred after cell divisions started, without the participation of the active demethylation machinery, and were concomitant with acquisition by B cells of the ability to proliferate. Gene Ontology analysis, expression profiling, and high-throughput analysis of the presence of transcription factor binding motifs and occupancy revealed that most genes undergoing hypomethylation are active and display the presence of NF-κB p65 and other B cell-specific transcription factors. Promoter hypomethylation was associated with upregulation of genes relevant for the phenotype of proliferating lymphoblasts. Interestingly, pharmacologically induced demethylation increased the efficiency of transformation of resting B cells to lymphoblastoid cells, consistent with productive cooperation between hypomethylation and lymphocyte proliferation.

Conclusions: Our data provide novel clues on the role of the B cell transcription program leading to DNA methylation changes, which we find to be key to the EBV-associated conversion of resting B cells to proliferating lymphoblasts.

Background

Infection of B cells with Epstein-Barr virus (EBV), which is highly prevalent in humans, is an excellent model to investigate the molecular mechanisms associated with the acquisition of unlimited growth during disease. EBV-associated changes in B cells are relevant to the development and progression of lymphomas [1] and lymphoproliferative disorders in immune-suppressed individuals, and various autoimmune disorders like rheumatoid arthritis, systemic lupus erythematosus and multiple sclerosis [2]. In early primary human infection, EBV infects peripheral resting B cells and expresses six nuclear (EBNA1, 2, 3A, 3B, 3C and -LP) and two latent membrane proteins and small non-coding RNAs. This type of infection, in which these two groups of proteins

* Correspondence: eballestar@idibell.org
[1]Chromatin and Disease Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Avda. Gran Via 199-203, 08908 L'Hospitalet de Llobregat, Barcelona, Spain
Full list of author information is available at the end of the article

transform resting B lymphocytes into continuously proliferating lymphoblastoid cell lines, is referred to as type III latency [1,3]. This process mimics antigen-induced clonal expansion of resting B cells associated with MYC-mediated proliferation and upregulation of NF-κB, MAP kinases and antiapoptotic factors. Recent data have shown that EBNA2, which is essential to this process, enhances and exploits the B cell transcription program by binding to a variety of B cell transcription factor sites to achieve transformation [4]. *In vivo*, the vigorous cellular immune response directed against EBV-immortalized cells limits the proliferation and expansion of such latently infected cells at early stages of infection of a naïve host or in immunocompromised individuals. Studying type III latency lymphoblastoid cells is relevant because it not only allows the investigation of early steps in infection and the effects that the viral activity exerts on B cell function, but also is an excellent strategy for investigating changes related to the triggering of unlimited proliferation of B cells, before any additional secondary transforming genetic and epigenetic events occur.

The mechanisms by which B cell identity is altered in this process towards unlimited proliferation, triggered by EBV infection, involve the acquisition of epigenetic changes. In this context, DNA methylation might play a key role, since this epigenetic mark participates in regulating transcriptional activity [5] and is known to be highly aberrant in several types of EBV-associated lymphomas [6,7] and autoimmune diseases [8]. Despite its role in gene control, DNA methylation is not only a mechanism of transcriptional control but also guarantees genomic stability. The relationship between methylation and transcriptional activity has been best studied in promoter regions, particularly CpG island-associated promoters, where methylation is generally associated with transcriptional repression. In the context of the hematopoietic system, DNA methylation profiling has revealed overall higher methylation levels in the lymphoid branch relative to the myeloid one, and with respect to less differentiated progenitors [9].

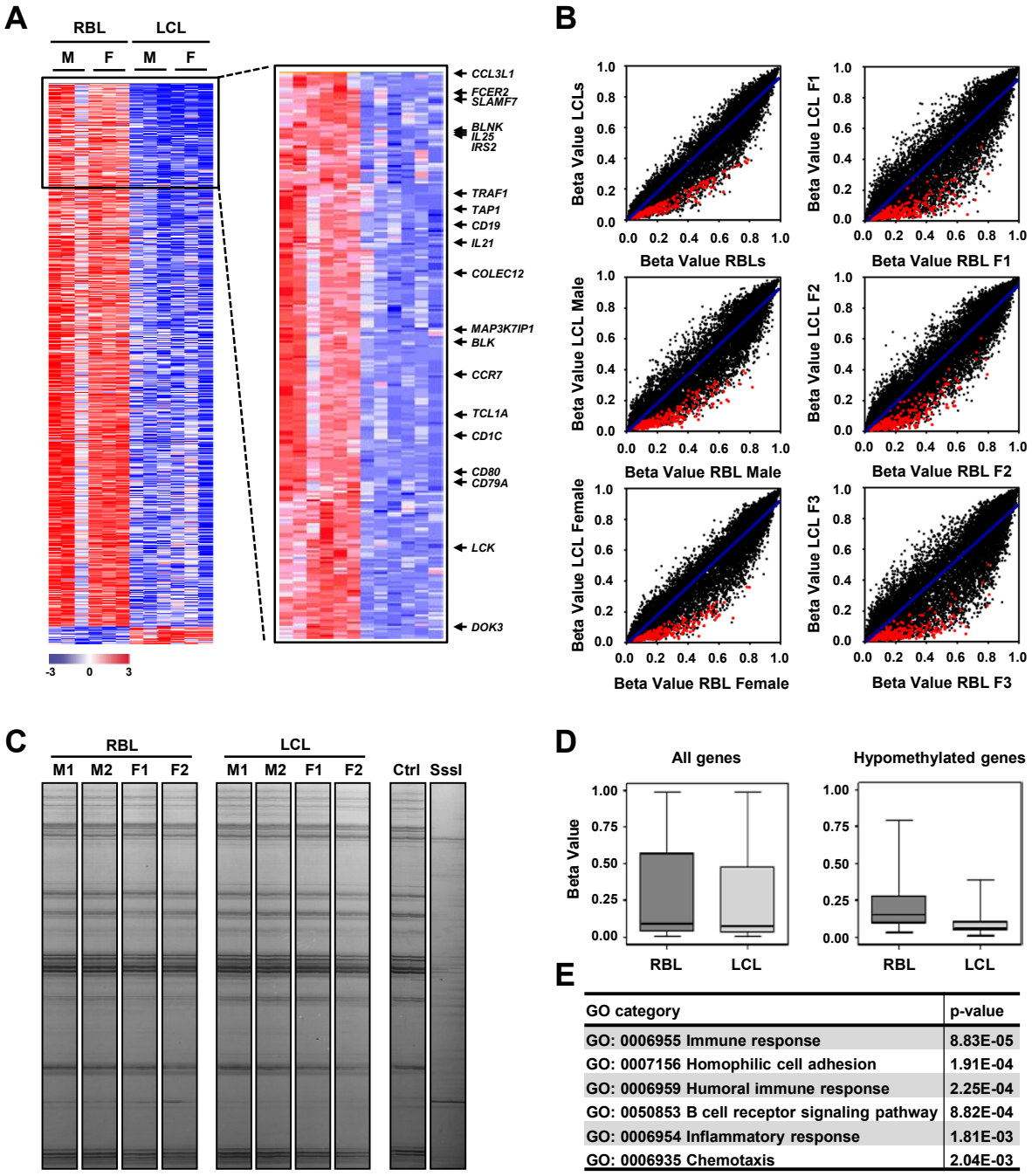
A number of studies have addressed the analysis of DNA methylation changes associated with EBV infection of B cells. Several of these have revealed that whereas the EBV genomic sequence is virtually unmethylated in free viral particles and lymphoblastoid cells, the genome is heavily methylated in both Burkitt and Hodgkin lymphomas [10]. Also, the DNA methylation status of EBV promoters has been widely studied in association with the activity of latency promoters [11-14]. By contrast, fewer studies have addressed the acquisition of DNA methylation changes by the host cell during EBV-mediated transformation between resting B cells and proliferating lymphoblasts. EBV influences changes in the DNA methylation status at specific sequences [15-17] and these are likely to influence or modify the B cell phenotype and function. It is therefore of inherent interest to investigate the extent and mechanisms of acquisition of changes in DNA methylation by B cells following EBV infection as well as their potential contribution to phenotypic changes during this process.

In this study we investigated the acquisition of DNA methylation changes during EBV-mediated transformation of resting B cells to lymphoblastoid cell lines by using methylation bead arrays. We exclusively observed significant hypomethylation of around 250 genes. No hypermethylation was found. Time course analysis indicated that hypomethylation occurs only when cell proliferation has started, suggesting the exclusive participation of replication-dependent mechanisms. Gene Ontology (GO) analysis, comparison with the expression patterns of different cell types and between resting B cells and proliferating lymphoblasts, high-throughput analysis of the presence of transcription factor binding motifs and occupancy revealed that most genes undergoing hypomethylation are active and display the presence of NF-κB p65 and other B cell-specific transcription factors. In addition, hypomethylation associates with upregulation of several genes that are essential in the transformation of resting B cells to continuously proliferating lymphoblasts. Pharmacologically induced DNA demethylation increases B cell transformation efficiency. Our data provide novel clues to the contribution of epigenetic mechanisms associated with EBV-associated conversion of resting B cells to proliferating lymphoblasts, the relevance of the cell type context and why DNA hypomethylation could be key in the efficiency of this process.

Results

DNA methylation profiling reveals that EBV-mediated B cell to lymphoblastoid transformation is associated with gene-specific hypomethylation

To investigate the acquisition of DNA methylation changes in association with EBV-associated transformation of resting B lymphocytes (RBLs), we first compared the DNA methylation profiles of six samples before and after EBV infection, once they had become lymphoblastoid cell lines (LCLs). To this end, we used methylation bead arrays that interrogate the DNA methylation status of over 27,000 informative CpG sites, including the region near the transcription start sites of more than 14,000 promoters. Statistical analysis of the combined data from the six pairs of samples revealed that 256 genes were hypomethylated (fold change (FC) > 2; false discovery rate (FDR) adjusted $P < 0.05$, Student's t-test) in B lymphoblastoid cells compared with resting B cells (Figure 1a; Additional file 1). By contrast, no hypermethylated genes were observed under these conditions. Scatterplots comparing the average DNA methylation patterns of the six RBLs



GO category	p-value
GO: 0006955 Immune response	8.83E-05
GO: 0007156 Homophilic cell adhesion	1.91E-04
GO: 0006959 Humoral immune response	2.25E-04
GO: 0050853 B cell receptor signaling pathway	8.82E-04
GO: 0006954 Inflammatory response	1.81E-03
GO: 0006935 Chemotaxis	2.04E-03

Figure 1 High-throughput methylation comparison of resting B lymphocytes and matching lymphoblastoid cells. **(a)** Heatmaps including the data for the six RBL/LCL pairs of samples showing significant differential methylation. The left panel shows all the genes showing a FDR < 0.05. Only those genes with a FC > 2 were selected (right panel). Data from the Illumina array were normalized as xi = xi - row.mean[i])/(0.333 × row.sd [i]). A scale is shown at the bottom, whereby positive (red) and negative (blue) values correspond, respectively, to higher and lower than average methylation status. M, male; F, female. **(b)** Scatterplots showing methylation profiles of matching RBL/LCL pairs. Genes with significant differences (FC > 2, FDR < 0.05) in averaged results from six samples are highlighted in red. Six panels are shown: top left, mean of six experiments/pairs of samples; middle left, male samples; bottom left, female samples; right panels, three individual RBL/LCL comparisons. **(c)** Band patterning corresponding to the analysis of unmethylated/methylated Alu (AUMA) repeats. Four RBL/LCL (M, male; F, female) pairs are shown. To illustrate the sensitivity toward DNA methylation changes, a B cell sample is compared with the same sample following limited treatment with SssI DNA methyltransferase. **(d)** Comparison of the methylation levels in RBLs and LCLs for all the CpGs represented in the 27k bead array and CpGs that undergo hypomethylation in this process. Box and whisker plots are presented, where the bottom and top of the box are the 25th and 75th percentile and the bar near the middle is the 50th percentile (the median). **(e)** Gene Ontology (GO) analysis of hypomethylated genes during EBV-mediated RBL to LCL conversion.

and matching LCLs had highly reproducible DNA methylation profiles among different samples (Figure 1b). Changes corresponding to the average six pairs of B cells/lymphoblastoid cells were almost identical to the pattern obtained for only male or only female RBL/LCL comparisons, or those changes obtained for each individual pair of samples (Figure 1b), highlighting the specificity of the differences observed.

Since disease-related hypomethylation alterations generally affect repetitive elements, we also investigated changes in CpG sites in this type of sequence. In humans, most of the methylated cytosines are found in CpG-rich sequences within the tandem and interspersed repeats that constitute up to 45% of the human genome, of which Alu repeats are the most common family. We used genome-wide amplification of unmethylated DNA Alu repeats (AUMA) [18] to perform high-throughput screening of DNA methylation at these sequences. This experiment revealed no differences between RBLs and LCLs, suggesting that loss of methylation does not occur in repetitive sequences (Figure 1c).

Our analysis revealed a mean 21.5% methylation among the 256 genes undergoing demethylation in RBLs (Figure 1d). This average level of methylation at or near promoters is compatible with active gene expression in resting B cells. In fact, GO analysis showed significant enrichment ($P < 0.05$) for the following categories assigned to biological processes: immune response (GO:0006955; P-value = 8.8×10^{-5}), humoral immune response (GO:0006959; P-value = 2.2×10^{-4}), B cell receptor signaling pathway (GO:0050853; P-value = 8.8×10^{-4}), inflammatory response (GO:0006954; P-value = 1.8×10^{-3}) and chemotaxis (GO:0006935; P-value = 2.0×10^{-3}) (Figure 1e). These categories include the presence of key markers of B cell function and identity, such as CD19, CD79a and BLNK. Among the list of hypomethylated genes, the presence of genes that are EBV-induced, such as *CCR7* (EBI1), *GPR183* (EBI2), *EBI3* (IL27 subunit beta) and *TRAF1* (EBI6), is also noteworthy.

To confirm that differences identified in DNA methylation between B cells and B lymphoblasts were robust, we carried out bisulfite genomic pyrosequencing of selected genes looking at CpG sites corresponding to the oligonucleotide probe represented in the methylation array, which is generally located around the transcription start site. We selected 20 genes on the basis of the magnitude of change in methylation, as revealed by the analysis of our methylation arrays, and their functional relevance in the context of B cell biology: *CD19*, *CD79a*, *BLK*, *FCER2*, *LCK*, *BLNK*, *IL21*, *CCL3L1*, *SLAMF7*, *IL25*, *IRS2*, *TAP1*, *COLEC12*, *MAP3K7IP1*, *TCL1A*, *CD1C*, *CD80* and *DOK3*, including two of the genes originally described as EBV-induced (*CCR7* and *TRAF1*). In all cases, bisulfite pyrosequencing of these genes confirmed an at least two-fold significant decrease for the aforementioned genes (Figure 2a; Additional file 2).

Hypomethylation associated with EBV-mediated transformation of RBLs to LCLs occurs in association with proliferation and does not involve active demethylation mechanisms

Our results indicated that the transformation of resting B cells to proliferating B lymphoblastoid cells is associated with gene promoter hypomethylation. Loss of methylation may occur as a result of the defective maintenance of DNA methylation as DNA replication and subsequent cell divisions start or, alternatively, as an active mechanism. To discriminate between these two possibilities, we first performed bisulfite genomic pyrosequencing of the genes previously analyzed in samples generated over time, including points at 1, 3 and 24 hours, before cell proliferation had been initiated [19], and LCLs after 1 and 2 weeks. To this end, we used a form of EBV that infects B cells very efficiently [20] and in which around 90% of B cells express EBNA2 (Figure 2b). The results showed that significant demethylation only occurred after cell divisions have started to take place (Figure 2c), although the coincidence in time with replication does not necessarily mean that hypomethylation takes place through a passive mechanism.

A variety of factors are known to be involved in active demethylation. Recent studies have drawn attention towards a family of enzymes, the Tet proteins, that convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and other modified forms of cytosine. 5hmC may represent intermediates in the process leading to active DNA demethylation [21,22]. In addition, activation-induced deaminase (AID) participates in active demethylation in a two-step process, whereby 5mC is first deaminated by AID to thymine, followed by G/T mismatch-specific thymine DNA glycosylase (TDG)-mediated excision repair [22,23]. This process could potentially take place on 5mC, although some studies have shown that hydroxymethylation may target the methylcytosine residues that are going to be demethylated through this process [22]. Analysis of the 5hmC levels at different times in several genes that become hypomethylated during RBL to LCL transformation showed neither a significant presence of 5hmC on the CpGs that undergo hypomethylation nor any changes during this process (Figure 2d). We did not see any significant change in the expression levels of Tet proteins (not shown) during this process. We also investigated the potential involvement of AID in hypomethylation. AID is overexpressed during EBV-mediated conversion of RBLs to LCLs. RT-PCR and western blot analysis of AID

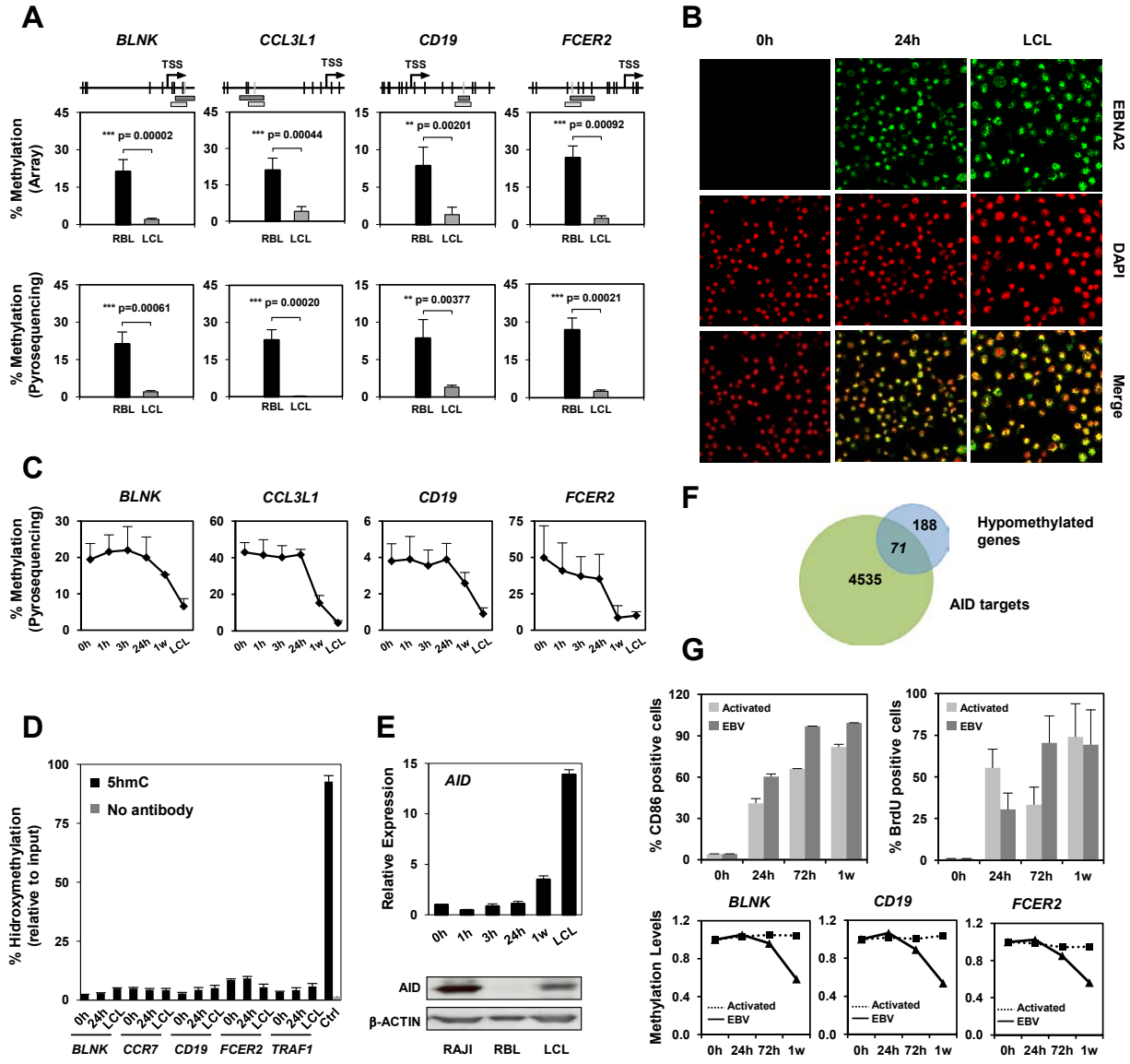


Figure 2 A comparison of the DNA methylation levels of selected genes between RBLs and resulting LCLs. **(a)** Bisulfite pyrosequencing was performed for the genes selected from experiments with methylation arrays. The distribution of CpG sites (vertical black line), the transcription start site (TSS; arrow) and the exact CpGs analyzed by the array (horizontal light grey bar) and pyrosequencing (horizontal dark grey bar) are indicated above each gene. Average data from six samples is represented. **(b)** EBNA2 and DAPI staining of B lymphocytes at 0 and 24 hours after infection with EBV derived from 293 cells carrying a recombinant B95.8 EBV genome, and LCLs, in which high levels of infection were achieved. **(c)** Pyrosequencing data showing the methylation changes at selected genes at different times. **(d)** Analysis of the 5-hydroxymethylcytosine content/changes at CpG sites undergoing hypomethylation in selected genes. As a positive control, *in vitro* hydroxymethylated DNA was used. **(e)** Changes in expression of activation-induced deaminase (AID) during RBL conversion to LCLs at the mRNA and protein levels. For protein analysis, the Burkitt's lymphoma cell line RAJI was used as a positive control for AID expression. **(f)** Overlap between AID target genes and hypomethylated genes. **(g)** Comparison between B cells activated with IL4 and CD40L and infected with EBV. The top panel shows the proportion of activated cells (left; determined by fluorescence-activated cell sorting (FACS) analysis with CD86) and proportion of dividing cells (right; determined with bromodeoxyuridine (BrdU)). Bottom, bisulfite pyrosequencing analysis of selected CpGs in association with EBV infection and with B cell activation by IL4 and CD40L. W, week.

revealed an increased expression of AID during this process (Figure 2e). Recent ChIP-seq data have also served to identify direct AID targets in B cells [24]. Comparison of these data with our own hypomethylation data showed that 71 out of 256 genes overlapped with AID binding sites (Figure 2f), although a chi-square test indicated that there is no significant enrichment. We also generated a retroviral inducible system for AID in a B cell line to test

whether forced expression of this factor could lead to hypomethylation in the aforementioned genes. Bisulfite pyrosequencing of the same CpG sites of the previously validated genes before and after expression of AID showed no differences in methylation (not shown).

Given that EBV-mediated transformation and B cell activation share common pathways [25], we also performed bisulfite pyrosequencing of these genes in B cells activated/stimulated with IL4 and CD40L to test the EBV-associated specificity of the observed DNA demethylation changes. Under our conditions we achieved similar levels of both cell activation, measured by cytometry analysis of CD86 surface marker, and proliferation, measured by bromodeoxyuridine (BrdU) incorporation (Figure 2g, top) in both types of stimulation of B cells. However, no changes in the DNA methylation levels of these genes were observed during CD40L/IL4-stimulation of B cells (see examples in Figure 2g, bottom) even after 1 week, suggesting that methylation changes associated with EBV-mediated transformation to LCLs are independent of B cell activation.

Hypomethylated genes are enriched for binding of NF-κB and other lymphocyte-specific transcription factors

Our analysis of the dynamics of DNA demethylation suggests that this process occurs in a replication-associated manner, since changes only occurred once cell divisions had started and did not involve changes in 5hmC or the action of AID. In this context, hypomethylation could potentially associate with genomic sites that are less efficient in maintaining methylation during DNA replication, perhaps in regions that are associated with active transcription in the context of B cell function.

Previous studies have shown that EBV transcription factor EBNA2 enhances and exploits the RBL transcription program by binding to a variety of B cell transcription factor sites [4]. We wondered whether hypomethylation is taking place at genomic sites bound by specific transcription factors, perhaps associated with active transcription in B cells. To address this, we investigated the possible enrichment of transcription factor motifs from the TRANSFAC database in a region of 500 bp around the CpG sites in which hypomethylation had been detected in our study. Significant enrichment of a small set of transcription factors was observed. Remarkably, significant enrichment of the binding motifs of two subunits of the NF-κB complex, specifically c-REL (9.3%, P-value = 2.0×10^{-5}) and p65/RELA (6.3%, P-value = 3.1×10^{-3}) (Figure 3a) was observed. NF-κB is known to be involved in the survival of LCLs and latency III-regulated cell gene expression [26]. We also found enrichment of additional hematopoietic transcription factor binding motifs in this 500-bp window around hypomethylated CpG sites, such as GATA3, STAT1 and the MEF2 family (Figure 3a).

To evaluate the extent to which hypomethylated genes correlate with NF-κB occupancy, we used our methylation data and ChIP-seq data for NF-κB p65. Enriched NF-κB p65 peaks from a ChIP-seq study in lymphoblastoid cells [27] were annotated to the nearest Ensembl gene build (version 54) [28] using the Bioconductor package ChIPpeakAnno [29]. All the targets from all four replicates were uniquely combined. Overlap of hypomethylated genes and NF-κB p65 targets were represented in a Venn diagram and the significance of overlap was determined by a standard chi-square test, written in the syntax of the R statistical program (Figure 3b).

We found that among the entire set of unique direct NF-κB p65 targets, 131 genes were shared with our list of 256 hypomethylated genes, that is, 51% of our hypomethylated genes were directly associated with the NF-κB p65 subunit (Figure 3b). Examples of the detailed binding of NF-κB p65 to these hypomethylated regions are shown in Figure 3c. Additional ChIP-seq data for other transcription factors were also used to investigate the overlap with hypomethylated genes. Specifically, we used ChIP-seq data from EBF1, IRF4, MEF2A, MEF2C, PAX5 and PU1 obtained from GM12878 cells (LCLs) from the ENCODE project (Figure 3D). Remarkably, significant enrichment was obtained for the binding of EBF1, IRF4 and MEF2C (P-value < 0.05). EBF1 is a transcription factor that is critical for both B lymphopoiesis and B cell function [30]. IRF4 is as a crucial transcription factor in the generation of functionally competent plasma B cells [31]. Finally, transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation [32].

chipseq

Our results suggested an association between genes that are regulated by B cell-specific transcription factors, particularly some associated with B cell activation and proliferation, and those that become hypomethylated. It is possible that during transformation of B cells to proliferative lymphoblasts, cells are less efficient in maintaining DNA methylation at actively transcribed sites, perhaps due to a lower presence of DNA methyltransferases (DNMTs), which are preferably bound to heterochromatic regions [33,34]. One could argue that in quiescent B cells there is a tendency for inactive genes to become methylated, and when the cells are activated to grow and proliferate, genes that become active lose methylation. This possibility was partially discarded by analyzing the methylation levels in CD40L/IL4-activated B cells (Figure 2g). Also, the analysis of bone marrow CD19+ cells, where a high proportion of B cells are proliferating [35], showed no methylation differences with respect to peripheral blood CD19+ cells, which are quiescent (Additional file 3). It could also be possible that the finding of lower methylation levels in proliferating B cells was due to the specific infection by EBV of a B cell subpopulation with lower levels of

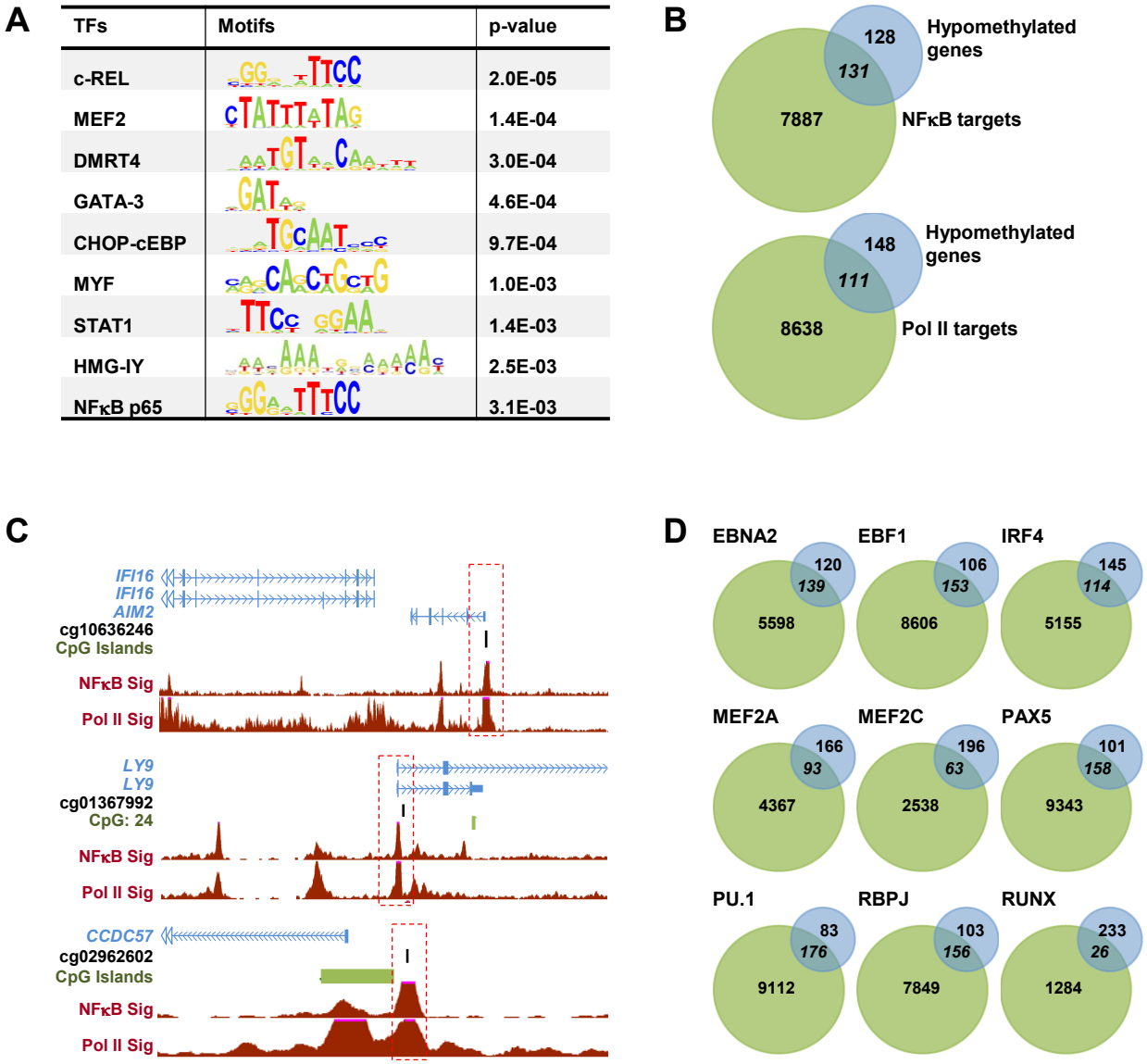


Figure 3 Association of factors with hypomethylation during RBL to LCL conversion. **(a)** Significant enrichment of predicted transcription factors (TFs; TRANSFAC motif) in hypomethylated regions. The 500-bp region around the center of significantly hypomethylated CpG sites was tested. Color intensity in the heatmap cells represents FDR *P*-values. Redder and yellower colors indicate greater and lower significance, respectively; gray indicates insignificance. **(b)** Venn diagram showing overlap of NF-κB targets from ChIPseq (considering targets that are also within the hypomethylated list) and hypomethylated genes (the 131 overlapping genes are listed in Additional file 4). **(c)** Three examples showing NF-κB p65 binding to the region neighboring the hypomethylated CpGs. The binding motif location is presented as a horizontal green bar. Below, NF-κB p65 and RNA polymerase II (Pol II) binding from GM12878 ChIP-seq data are shown. **(d)** Venn diagrams showing overlap between hypomethylated genes and transcription factors EBF1, IRF4, MEF2A, MEF2C, PAX5, PU.1, RBPJ, RUNX and EBNA2.

methylation. To address that, we pyrosequenced several genes in different B cell subpopulations present in peripheral blood, including naïve B cells, as well as unswitched and switched memory B cells. No differences in methylation were observed with respect to total CD19+ cells from peripheral blood or RBLs and they all displayed higher methylation than LCLs (Additional file 3).

Several key genes are overexpressed is association with hypomethylation during EBV-mediated transformation to LCLs

Our findings suggest that many of the genes undergoing hypomethylation are actively transcribed in B cells, and therefore loss of methylation may not directly affect their expression levels. However, it is important to determine

whether hypomethylation is also associated with overexpression of genes that may confer an advantage to proliferating B lymphoblasts.

To address this matter, we performed quantitative PCR of the selected genes in the set of six paired RBL and LCL samples (Figure 4a). Loss of methylation was in some cases associated with an increase in gene expression. This happens for instance for all *bona fide* EBV-induced genes (EBI), like *EBI1*, *EBI2* and *EBI3* (Figure 4a). Individual analysis also showed increase in mRNA levels for other genes, although in other cases this correlation with hypomethylation did not occur, and no changes or even a decrease in expression was observed, although the fold change was relatively low (see central and bottom panels in Figure 4a). This is not surprising, given that the DNA methylation levels of some of the genes undergoing hypomethylation in this process were already relatively low in resting B cells. As mentioned above, many of these hypomethylated genes are key factors in B cell function (*CD19*, *CD79a*, *BLNK*, and so on) and they are expected to be highly expressed in resting B cells. To test whether RNA polymerase II (Pol II) is associated with these genes we overlapped our data with the ChIP-seq data and found that at least 111 of them were Pol II-bound genes (Figure 3b). Examples of the detailed binding of Pol II to these hypomethylated regions are shown in Figure 3c.

We tested whether our 256 hypomethylated genes were generally highly expressed in B cells relative to other cell types. Thus, we compared normalized Affymetrix mRNA expression data of 73 normal human tissues [36]. Comparison of expression of the 256 hypomethylated genes indicated that around 28% of them were more highly expressed in B cells than the average expression level of the 73 other tissue types (Figure 4b), reinforcing our findings on the relevance of the B cell context in the genes that undergo hypomethylation.

For a general comparison in the context of transformation of RBLs to LCLs, we compared the expression profiles of the 256 genes hypomethylated in B and LCLs (available at the Gene Expression Omnibus (GEO) database under series accession number GSE26212) [16]. This revealed that around 41% of the genes displayed high levels of expression in RBLs (Figure 4c). However, around 59% of the remaining genes accomplished significant increase in gene expression (Figure 4b; Additional file 4), correlating with the over two-fold reduction in DNA methylation. Remarkably, the list of genes that become hypomethylated and are overexpressed includes several key genes in RBL to LCL conversion (Additional file 4). For instance, *EBI3* undergoes an approximately five-fold change in methylation and becomes overexpressed. This gene is a subunit of the heterodimeric cytokine IL27, known to be regulated through NF-κB activation [37] and to induce B cell proliferation, which is stronger in naïve than in memory B cells [38]. Another example is LTA, or lymphotoxin alpha, a member of the tumor necrosis factor family, and an autocrine growth factor induced upon binding of NF-κB [39]. Other examples in our list include genes like *SLAM1* and *SLAMF7*, two members of the 'signaling lymphocyte activation molecule' (SLAM) family, also implicated in B cell proliferation [40].

Pharmacologically induced DNA demethylation enhances proliferation during RBL to LCL transformation

Our analysis revealed that a high proportion of the genes that undergo hypomethylation during RBL to LCL transformation become overexpressed and that several of them are important for the transformation on continuous proliferating B lymphoblasts. To determine whether hypomethylation has any effect on the efficiency of transformation, we investigated whether pharmacologically induced hypomethylation influences RBL to LCL transformation. We tested the proliferation rate in RBLs infected with EBV and incubated in the presence of increasing concentrations with the demethylating agent 5-azadeoxycytidine (azadC) at concentrations between 50 pM and 50 μM. MTT assays showed that mock-treated cells started to proliferate around day 6 after infection with EBV (Figure 5a). Cell viability and proliferation rate decreased at very high doses of azadC, consistent with the toxicity properties of azadC at high doses, which we tested on cultured LCLs (Figure 5a). However, at low azadC concentrations (50 pM, 500 pM and 5 nM), where viability of cells is comparable to that of control cells (Figure 5a), a significant increase in cell proliferation was observed after day 8 (Figure 5b). This increase in proliferation in the presence of azadC was not observed when established LCLs were used as a control in proliferation experiments in the presence of azadC, suggesting that the effect of azadC is associated with the transformation of RBLs to proliferating B cells. Bisulfite genomic analysis of several of the target genes confirmed the loss of methylation at the studied CpG sites upon treatment with 5azadC at low concentrations (Figure 5c). In fact, we observed that azadC-treated cells had lower levels of methylation than control cells at day 10 (Figure 5c). Analysis of viral promoters showed that these were unmethylated in both free viral particles and in LCLs, ruling out the possibility that demethylation at EBV promoters and overexpression of EBV-encoded genes are responsible for more efficient transformation of RBLs to LCLs.

Discussion

Our results provide evidence that EBV-mediated transformation of RBLs to LCLs results in the demethylation

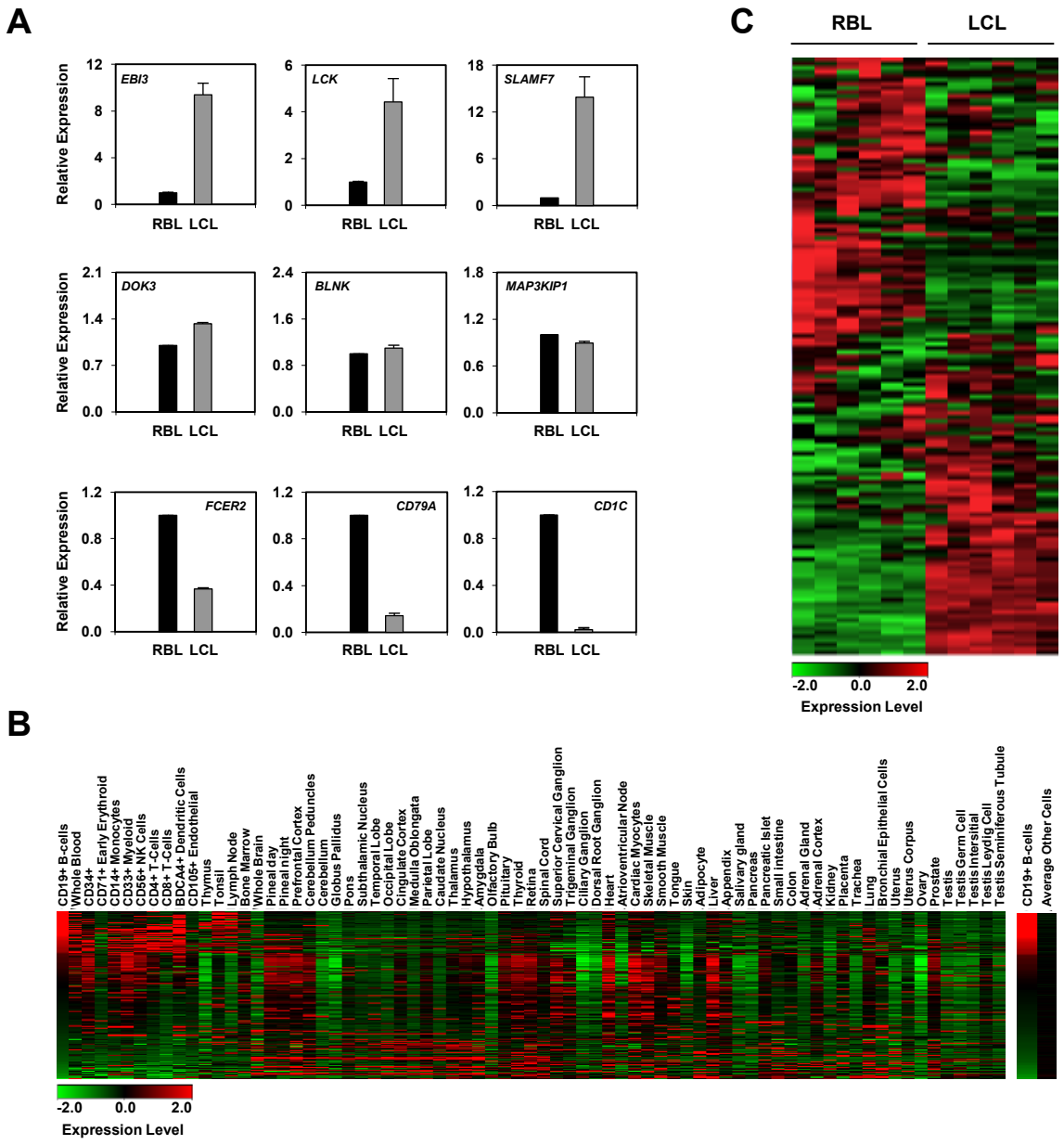


Figure 4 Expression analysis of genes undergoing hypomethylation during RBL to LCL conversion. **(a)** Expression analysis of selected genes in RBLs and matching LCLs. Error bars represent the standard deviation obtained from six independent measurements. **(b)** Heatmap showing the relative expression of the 256 genes hypomethylated in B cells with respect to other cell and tissue types. Expression data from Affymetrix mRNA expression analysis with 73 normal human tissues [36]. **(c)** Heatmap showing the relative expression of the 256 genes hypomethylated in B cells with respect to LCLs. Expression data obtained from GSE26212 [16].

of a cluster of around 250 genes. Conversely, no significant hypermethylation is observed. Hypomethylation has been described in various immune disease-related contexts, including ICF (immunodeficiency, centromere instability and facial anomalies) syndrome [41] and autoimmune diseases [8]. In most cases, disease-associated hypomethylation occurs at repetitive sequences, including Alu elements [42]. However, in RBL to LCL transformation, hypomethylation takes place near the transcription start sites of around 250 genes, and no apparent hypomethylation occurs at the major type of repetitive elements. Lack of hypomethylation in repetitive sequences is consistent with the minimal changes in genomic stability associated with EBV-associated lymphoproliferation.

Given that hypomethylation occurs in association with the acquisition of proliferation, it is likely to be related to

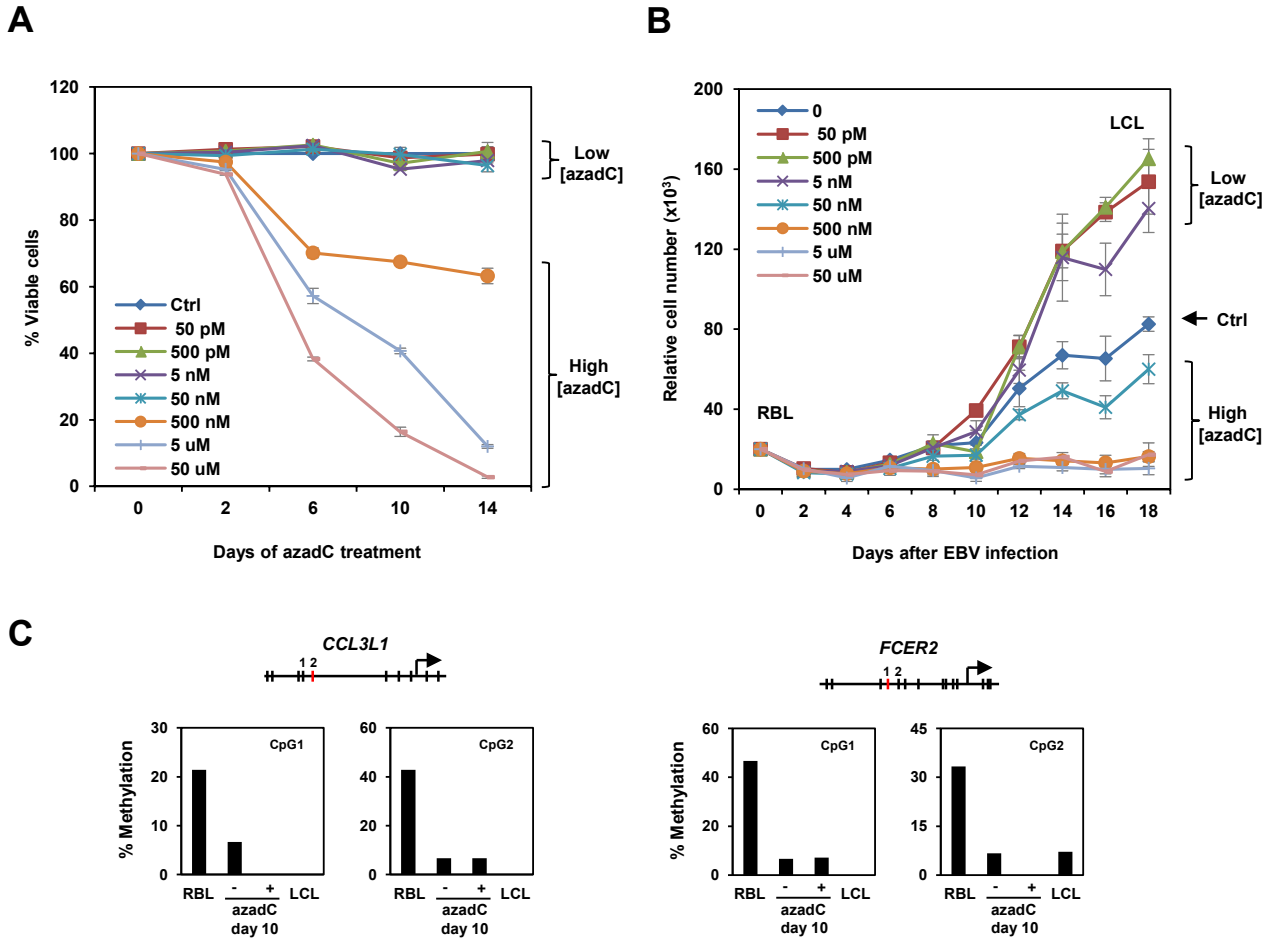


Figure 5 Effects of 5azadC in RBL to LCL conversion. **(a)** Influence of 5azadC treatments in viability of LCLs determined by flow cytometry based on propidium ioidide fluorescence. **(b)** Influence of 5azadC treatments in proliferation, as determined by MTT assays, during RBL to LCL conversion. **(c)** Bisulfite sequencing of two example genes, where methylation levels of two CpG sites is shown (the one represented in the array is highlighted in red). Methylation analysis was done in RBL cells at 10 days after EBV infection in the absence and presence of azadC, and in LCLs.

the decreased efficiency in DNA methylation maintenance as cells start to divide. The lack of evidence for active demethylation mechanisms during this process also supports the above possibility. The transition of a differentiated resting cell to a proliferative status would require the participation of DNMTs to maintain the DNA methylation profiles throughout replication/division cycles. Various lines of evidence indicate that DNMTs are associated with repetitive sequences [33,34] and that heterochromatic regions act as a reservoir for DNMTs [33]. This would explain why active regions, which are less rich in DNMTs, lose methylation in the transformation from resting to proliferative B cells, whereas in ICF syndrome, which is characterized by mutations in DNMT3B, hypomethylation takes place in heterochromatic regions, which act as reservoirs of DNMTs.

In fact, analysis of the regions associated with hypomethylation during RBL to LCL transformation shows enrichment of highly expressed genes and target sites of transcription factors that are expressed in B cells or are specifically activated during EBV infection, such as factors in the NF-κB pathway [26,43]. The modulation of the NF-κB signaling by EBV is not only important for viral infection, but also contributes to the development of malignant neoplasia [43]. In addition, our results indicate that hypomethylated genes are regulated by B cell-specific factors like EBF1, IRF4 and MEF2C, the last two of which are implicated in B cell activation and proliferation processes. We also observed that many of the genes that undergo hypomethylation, like *CD19*, *CD79a*, *BLNK*, *BLK* and *LCK*, are highly expressed in B cells. This suggests that active regions are either more accessible to putative demethylating machinery or are less efficient at maintaining methylation as proliferation starts. Given the lack of demethylation events before proliferation begins, and that we were unable to detect links between any of the

proposed mechanisms associated with active demethylation, we favor the occurrence of a replication-coupled mechanism of demethylation, due to the inefficient maintenance of DNA methylation as replication is initiated. Similar to studies of the EBNA2 binding sites [4] in which EBV exploits intrinsic B cell transcription programs, hypomethylation takes place preferentially in genes that are expressed. Comparison of the lists of significantly hypomethylated genes and the presence of transcription factor binding motifs and occupancy reinforced this notion. Data from different sources suggest that transcriptionally active regions are devoid of DNMTs, making the maintenance of DNA methylation less efficient. On the other hand, since the main functional consequence of promoter hypomethylation would potentially be an increase in transcription, it is likely that demethylation contributes to overexpression of some of these genes. In this sense, we have identified several key genes that undergo both hypomethylation and overexpression during the conversion of RBLs to LCLs. These include genes like *EBI3* [37,38], a subunit of the cytokine IL27, *LTA* [39] and members of the *SLAM* family [40] among others that are known to contribute to lymphocyte proliferation (Additional file 4).

In addition, demethylation of genes that already have high expression levels in the first set of infected cells might also lock its transcriptional status, and therefore would reinforce the B cell phenotype, and might be key to enable infection of more B cells. EBV initially infects oropharynx epithelial cells, establishing a lytic replication that spreads to some B cells in nearby lymphoid tissues. Those expand as EBV establishes a type III latency growth-transforming infection of B cells that elicit a strong T-cell-specific response. Maintaining a B cell phenotype would favor recirculation and generalized expansion to B cell areas in secondary lymphoid organs where surviving B cells downregulate EBV antigen expression and establish long-term silent latency [44].

The functional relevance of demethylation in this process is highlighted by the observation that treatment of RBLs with azadC at low sub-toxic concentrations enhances RBL to LCL transformation. Since the EBV genome is virtually unmethylated in free viral particles, demethylation at EBV promoters and overexpression of EBV-encoded genes can be ruled out as responsible for more efficient transformation of RBLs to LCLs. Our findings on the enhanced cell proliferation concomitantly associated with the presence of low sub-toxic amounts of azadC suggests that initial DNA methylation changes may cooperate in the efficiency of this process.

Although type III latency lymphoblastoid cells do not necessarily correspond to initial steps of lymphomagenesis, they both share enhanced proliferation [45]. It is likely that hypomethylation occurs in early steps towards to lymphomagenesis, when enhanced proliferation starts. In other hematological neoplasias, like in acute myeloid leukemia, *DNMT3A* mutations are highly recurrent [46], and it has been proposed that these mutations in an enzyme responsible for the establishment of DNA methylation are an early event of clonal evolution [47]. In B-cell chronic lymphocytic leukemia, increased levels of *TCL1* result in inhibition of DNA methylation [48]. Inhibition of methylation is proposed to be a common oncogenic mechanism in leukemogenesis [48]. These findings highlight the relevance of our study as a potential early mechanism in this group of tumors.

Conclusions

Our study of DNA methylation changes in EBV-mediated transformation of RBLs to LCLs reveals that only promoter hypomethylation occurs during this process. Neither significant hypermethylation nor methylation changes at repetitive elements are observed. Hypomethylation takes place only when proliferation has started, and the analysis of putative elements of the active demethylation machinery does not indicate their implication during this process. Most genes undergoing hypomethylation are active and display the presence of NF-κB p65 and other B-cell-specific transcription factors. Since DNMTs tend to associate with heterochromatic regions, it is likely that maintenance of DNA methylation is less efficient in transcriptionally active regions as cells start to proliferate triggered by EBV; however, this process does not occur during CD40L/IL4-stimulation of B cells where no changes are observed. Our results show that hypomethylation is associated with further upregulation of gene expression of many of these genes. Also, pharmacologically induced demethylation increases B cell transformation efficiency and proliferation. Collectively, our data indicate that the B cell transcription machinery is associated with the subset of genes that undergo hypomethylation. The finding that relevant genes in EBV-mediated transformation of B cells are further upregulated indicates a key role of this mechanism. Hypomethylation has been proposed to play a role in early stages of hematological malignancies, including other B cell malignancies. Our data reinforce the notion of the role of hypomethylation in early lymphomagenesis and shed light on underlying mechanisms.

Materials and methods

Ethics statement

Human samples (blood) used in this study come from anonymous blood donors and were obtained from the Catalan blood donation center (Banc de Sang i Teixits). Since the samples are anonymous, no informed consent is therefore required. The protocol used to transform with EBV B cells obtained from these anonymous donors was approved by the Committee of Biosecurity of IDIBELL (CBS) on 5 May 2011 and the Ethics Committee of

the University Hospital of Bellvitge (CEIC) on 28 May 2011.

Subjects and sample preparation

Buffy-coats from anonymous blood donors were obtained from the Catalan blood donation center (Banc de Sang i Teixits). Viable peripheral blood mononuclear cells were isolated using LymphoprepTM density gradient centrifugation. Resting B cells were isolated by positive selection using CD19 MicroBeads (Miltenyi Biotec, Bergisch Gladbach, Germany), or by depletion using a B Cell Isolation Kit (Miltenyi Biotec). Isolated B cells were immortalized with the supernatant of the EBV producer cell line B95.8 for the methylation studies and with the 2089 EBV made from 293 cells carrying a recombinant B95.8 EBV genome [20] for expression analysis.

DNA methylation profiling using universal bead arrays

Infinium Methylation Assay (Illumina, Inc., San Diego, CA, USA) was used to analyze DNA methylation. The HumanMethylation27 panel allows researchers to interrogate 27,578 highly informative CpG sites per sample at single-nucleotide resolution. This panel targets CpG sites located within the proximal promoter regions of transcription start sites of 14,475 consensus coding sequencing (CCDS) in the NCBI Database (Genome Build 36). In addition, 254 assays cover 110 microRNA promoters. On average, two assays were selected per CCDS gene and from 3 to 20 CpG sites for > 200 cancer-related and imprinted genes. Bisulfite conversion of DNA samples was done using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA). After bisulfite treatment, the remaining assay steps were identical to those of the Infimium Methylation Assay, using reagents and conditions supplied and recommended by Illumina. Two technical replicates of each bisulfite-converted sample were run. The results were all in close agreement and were averaged for subsequent analysis. The array hybridization was conducted under a temperature gradient program, and arrays were imaged using a BeadArray Reader (Illumina Inc.). The image processing and intensity data extraction software and procedures were those described by Bibikova and colleagues [49]. Each methylation data point was represented as a combination of the Cy3 and Cy5 fluorescent intensities from the M (methylated) and U (unmethylated) alleles. Background intensity computed from a set of negative controls was subtracted from each data point.

Detection of differentially methylated genes from the methylation array and functional analysis

A *t*-test was carried out to identify probes differentially methylated between primary B cells relative to their counterparts obtained in the LCL. *P*-values were corrected for multiple testing using the method proposed by Benjamini and Hochberg [50] to control the FDR. Genes showing a FDR adjusted *P*-value < 0.05 and a minimum mean methylation FC of two were considered to be differentially methylated. Functional annotation of hypomethylated genes was based on GO (Consortium, 2000), as extracted from EnsEMBL [28] and the KEGG pathway database [51]. Accordingly, all genes were classified into three ontologies, based on their involvement in biological processes, molecular functions and cellular components. We took only the GO/pathway categories that had at least ten annotated genes. We used GiTools for enrichment analysis and heatmap generation [52]. Resulting *P*-values were adjusted for multiple testing using Benjamini and Hochberg's FDR method [50]. An FDR cutoff of 0.25 was used for selection of enriched terms. The relationships between the DNA methylation data (from standard bisulfite sequencing and/or quantitative bisulfite pyrosequencing) and the age and sex of the individuals were evaluated by the Pearson chi-square test.

Analysis of gene promoter methylation: bisulfite sequencing and pyrosequencing

CpG island DNA methylation status was determined by sequencing bisulfite-modified genomic DNA. Bisulfite modification of genomic DNA was carried out as described by Herman *et al.* [53]. To validate the DNA methylation data obtained by the Infinium methylation assay, bisulfite pyrosequencing was performed according to standard protocols and evaluated with the Pyro Q-CpG 1.0.9 program (Biotage, Uppsala, Sweden). Primer sequences, product lengths and annealing temperatures used in the bisulfite sequencing and bisulfite pyrosequencing PCR reactions are shown in Additional file 5. Raw data for bisulfite sequencing of all samples is presented in Additional file 6.

Amplification of unmethylated Alus (AUMA)

DNA digestion with SmaI enzyme and ligation to the linker were performed as described elsewhere [18]. The product was purified using the GFX Kit (Amersham Biosciences, Uppsala Sweden) and eluted in 250 μl of sterile water. A chimeric primer comprising the complementary linker sequence (ATTCGCAAAGCTCTGA), the cut SmaI site (GGG) and three additional nucleotides homologous to the Alu consensus sequence were used to enrich for Alu sequences: AUMA-TTC (ATTCGCAAAGCTC TGAGGGTTC). Single primers were used for each PCR reaction. Products were resolved on denaturing sequencing gels. Bands were visualized by silver staining the gels. Faint bands with inconsistent display due to small variations in gel electrophoresis resolution were not considered. Band reproducibility was assessed with the analysis of

PCR duplicates. AUMA fingerprints were visually checked for methylation differences between bands in different samples. Based on these premises, a given band was scored according to three possible behaviors: hypomethylation (increased intensity), hypermethylation (decreased intensity) and no change (no substantial difference between samples). Only those bands showing clear changes in their fingerprint intensities were considered to represent methylation changes.

Quantitative RT-PCR expression analyses

We reverse-transcribed total RNA extracted with TRIzol (Invitrogen), using a Transcriptor First Strand cDNA Synthesis Kit from Roche Diagnostics (Indianapolis, Indiana, USA). Quantitative real-time PCR analysis was performed in a PCR Real Time LightCycler 480 (Roche) with Sybr green. Primer sequences are shown in Additional file 5.

Analysis of transcription factor binding

Possible enrichment of transcription factor motif in the 500- to 1,000-bp region around the center of the hypomethylated probes and all other probes were predicted with the STORM algorithm [54], assuming *P*-value cutoffs of 0.00002 and 0.00001, respectively, using position frequency matrices (PFMs) from the TRANSFAC database (Professional version, release 2009.4) [55]. Enrichment analysis of predicted transcription factors in the probes of significant hypomethylated probes (n = 421) were conducted using GiTools [52]. We calculated a two-tailed *P*-value, and a finally adjusted FDR *P*-value (with 0.25 cutoff) was considered in establishing statistical significance.

We downloaded EBF1, IRF4, MEF2A, MEF2C, PAX5 and PU1 binding ChIPseq data for the GM12878 cell line from the ENCODE project [56]. Original hg19 genomic co-ordinates were converted to hg18 using the USCS 'liftover' tool. RUNX ChIPseq target data were obtained from the study of Hollenhorst and colleagues [57] (Jurkat cell line, GEO Database ID: GSE17954). RBPJ and EBNA2 taken from a 2011 study of Zhao *et al.* [4] (GEO database ID: GSE29498). In this case, mapped data were analyzed using the MACS (version 1.3.7.1) [58] pipeline to call peaks. In all cases, peaks were annotated to the nearest EnsEMBL [28] gene (version 54) using the Bioconductor package ChIPpeakAnno [29].

Expression analysis of hypomethylated genes in various tissue types

Expression data on RBL and LCL from the study of Caliskan and colleagues [16] (GEO accession GSE26210) were used to determine the relative expression of hypomethylated genes in these two cell types. Probes were annotated to Refseq genes and when more than one probe was present for the same gene, they were averaged. Also, all replicates on same sample were averaged. Expression data were normalized by median-centering the expression value of each gene across all the samples and dividing the value by the standard deviation. These normalized values were delineated in a color-coded heatmap using GiTools [52].

Normalized mRNA expression of data of 73 normal human tissues was downloaded from the BioGPS database [36] and analyzed similarly.

Cell proliferation and viability assays

In cell proliferation analysis, different dilutions of cells were plated and cultured at 37°C in 5% CO_2/95% O_2 for 20 days. AzadC (Sigma, St. Louis, MO, USA) was used in serial dilutions between 50 pM and 50 μM and refreshed on day 4 of treatment. Every 2 days, cells were fixed and stained with MTT and incubated for 4 hours at 37°C. The reaction was stopped with 50% N,N-dimethylformamide, 30% SDS, 2.5% glacial acetic acid and 2.5% acid chloride 1 N, and incubated overnight at 37°C in 5% CO_2/95% O_2. Cell quantities were determined by measuring the optical density at 560 nm. All assays were performed in triplicate. Cell viability was determined by the incorporation of propidium iodide in dead cells measured by flow cytometry.

CD40L activation of B cells

B cells were cultured at 1.5 × 106 cells/ml and activated with 50 ng/ml of M.CD40L (Enzo Life Sciences, Lausen, Switzerland) and 50 ng/ml of IL4 (Gentaur, Kampenhout, Belgium). The percentage of activated RBLs was determined by CD86 expression measured by flow cytometry and proliferating B cells were detected by measuring BrdU incorporation.

Use of a B-cell-based inducible system to test AID activity

Jiyoye B cells with inducible expression of AID were generated using the Retro-X TM Tet-ON® Advanced Inducible Expression System (Clontech, Saint-Germain-en-Laye, France). This system works through the sequential infection of the RetroX-Tet-ON advanced vector and the pRetroX-Tight-Pur vector. Carboxy-terminal hemagglutinin (HA)-tagged human AID was cloned in the pRetroX-Tight-Pur vector. The stable doubly infected cell line was selected with Geneticin (1 mg/ml) and Puromycin (0.3 μg/ml). AID expression was induced by the addition of doxycycline (500 ng/ml) for 24 hours. Nuclear export was inhibited by the addition of leptomycin B (10 ng/ml) for 2 hours.

Data access

Methylation array data for this publication have been deposited in NCBI's GEO and is accessible through GEO series accession number GSE41957 [59].

Additional material

Additional file 1: Hypomethylated genes in RBL to LCL transformation for a FDR adjusted *P*-value < 0.05 and fold change (FC) ≥ 2 from bead array analysis.

Additional file 2: Comparison of the methylation levels (in percentage) for selected genes from the bead arrays and calculated after bisulfite pyrosequencing.

Additional file 3: A comparison of the DNA methylation levels of selected genes in different B cell types. Bisulfite pyrosequencing was performed for the genes selected from experiments with methylation arrays. The analysis includes bone marrow (BM) CD19+ cells, naïve B cells, unswitched (US) memory B cells and switched (S) memory cells. Also peripheral blood resting B cells (RBLs) and corresponding lymphoblastoid B cells (LCLs) are included.

Additional file 4: Expression changes between those undergoing hypomethylation during the conversion between resting B cells and lymphoblastoid cells.

Additional file 5: Primer sequences.

Additional file 6: Individual raw data corresponding to bisulfite pyrosequencing presented in Figure 2.

Abbreviations

5hmC: 5-hydroxymethylcytosine; 5mC: 5-methylcytosine; AID: activation-induced deaminase; AUMA: amplification of unmethylated Alu repeats; azadC: 5-azadeoxycytidine; BrdU: bromodeoxyuridine; ChIP: chromatin immunoprecipitation; DNMT: DNA methyltransferase; EBV: Epstein-Barr virus; FC: fold change; FDR: false discovery rate; GEO: Gene Expression Omnibus; GO: Gene Ontology; LCL: lymphoblastoid cell line; NF: nuclear factor; RBL: resting B lymphocyte; Pol II: RNA polymerase II.

Authors' contributions

HH, CS-L, JR-U, VR-C, BJ, CM and MP performed experiments, AI, FA-S, AF, MF performed analysis, H-JD contributed with key tools, ME, EL-G, NL-B contributed with key analytic tools and interpreted data, and EB designed the project and wrote the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by grants PI081346 (FIS) and SAF2011-29635 from the Spanish Ministry of Science and Innovation (MICINN) and grant 2009SGR184 from AGAUR (Catalan Government). AI was supported by fellowship from AGAUR, Government of Catalonia, Spain. MP is supported by Ramon y Cajal Programme.

Author details

[1]Chromatin and Disease Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Avda. Gran Via 199-203, 08908 L'Hospitalet de Llobregat, Barcelona, Spain. [2]CR-UK Institute for Cancer Studies, University of Birmingham, Vincent Drive, Birmingham B15 2TT, UK. [3]Department of Experimental and Health Sciences, Barcelona Biomedical Research Park, Universitat Pompeu Fabra (UPF), c/Dr. Aiguader, 88, 08003 Barcelona, Spain. [4]Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA and Hematology Division, Brigham and Women's Hospital, Harvard Medical School, One Blackfan Circle, Brookline, MA 02115, USA. [5]Cancer Epigenetics Laboratory, Instituto Universitario de Oncología del Principado de Asturias (IUOPA), HUCA, Universidad de Oviedo, C/Dr. Emilio Rodríguez Vigil, s/n, 33006 Oviedo, Spain. [6]Cellular Differentiation Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Avda. Gran Via 199-203, 08908 L'Hospitalet de Llobregat, Barcelona, Spain. [7]Department of Virus Associated Tumours, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [8]Cancer Epigenetics Group, Cancer Epigenetics and Biology Programme (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Avda. Gran Via 199-203, 08908 L'Hospitalet de Llobregat, Barcelona, Spain. [9]Clinical Immunology Department, University Hospital La Paz, P° de la Castellana, 261, 28046 Madrid, Spain. [10]Department of Immunology and Oncology, Centro Nacional de Biotecnología/CNB-CSIC, Cantoblanco, Darwin 3, 28049 Madrid, Spain.

Received: 31 July 2012 Revised: 19 December 2012
Accepted: 15 January 2013 Published: 15 January 2013

References

1. Saha A, Robertson ES: **Epstein-Barr virus-associated B-cell lymphomas: pathogenesis and clinical outcomes.** *Clin Cancer Res* 2011, **17**:3056-3063.
2. Pender MP: **Infection of autoreactive B lymphocytes with EBV, causing chronic autoimmune diseases.** *Trends Immunol* 2003, **24**:584-588.
3. Cohen JI, Wang F, Mannick J, Kieff E: **Epstein-Barr virus nuclear protein 2 is a key determinant of lymphocyte transformation.** *Proc Natl Acad Sci USA* 1989, **86**:9558-9562.
4. Zhao B, Zou J, Wang H, Johannsen E, Peng CW, Quackenbush J, Mar JC, Morton CC, Freedman ML, Blacklow SC, Aster JC, Bernstein BE, Kieff E: **Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth.** *Proc Natl Acad Sci USA* 2011, **108**:14902-14907.
5. Deaton AM, Bird A: **CpG islands and the regulation of transcription.** *Genes Dev* 2011, **25**:1010-1022.
6. Esteller M, Gaidano G, Goodman SN, Zagonel V, Capello D, Botto B, Rossi D, Gloghini A, Vitolo U, Carbone A, Baylin SB, Herman JG: **Hypermethylation of the DNA repair gene O(6)-methylguanine DNA methyltransferase and survival of patients with diffuse large B-cell lymphoma.** *J Natl Cancer Inst* 2002, **94**:26-32.
7. Ammerpohl O, Haake A, Pellissery S, Giefing M, Richter J, Balint B, Kulis M, Le J, Bibikova M, Drexler HG, Seifert M, Shaknovic R, Korn B, Küppers R, Martín-Subero JI, Siebert R: **Array-based DNA methylation analysis in classical Hodgkin lymphoma reveals new insights into the mechanisms underlying silencing of B cell-specific genes.** *Leukemia* 2011, **26**:185-188.
8. Javierre BM, Fernandez AF, Richter J, Al-Shahrour F, Martin-Subero JI, Rodriguez-Ubreva J, Berdasco M, Fraga MF, O'Hanlon TP, Rider LG, Jacinto FV, Lopez-Longo FJ, Dopazo J, Forn M, Peinado MA, Carreño L, Sawalha AH, Harley JB, Siebert R, Esteller M, Miller FW, Ballestar E: **Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.** *Genome Res* 2010, **20**:170-179.
9. Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P, Lee H, Aryee MJ, Irizarry RA, Kim K, Rossi DJ, Inlay MA, Serwold T, Karsunky H, Ho L, Daley GQ, Weissman IL, Feinberg AP: **Comprehensive methylome map of lineage commitment from haematopoietic progenitors.** *Nature* 2010, **467**:338-342.
10. Fernandez AF, Rosales C, Lopez-Nieva P, Graña O, Ballestar E, Ropero S, Espada J, Melo SA, Lujambio A, Fraga MF, Pino I, Javierre B, Carmona FJ, Acquadro F, Steenbergen RD, Snijders PJ, Meijer CJ, Pineau P, Dejean A, Lloveras B, Capella G, Quer J, Buti M, Esteban JI, Allende H, Rodriguez-Frias F, Castellsague X, Minarovits J, Ponce J, Capello D, *et al*: **The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer.** *Genome Res* 2009, **19**:438-451.
11. Schaefer BC, Strominger JL, Speck SH: **Host-cell-determined methylation of specific Epstein-Barr virus promoters regulates the choice between distinct viral latency programs.** *Mol Cell Biol* 1997, **17**:364-377.
12. Bakos A, Banati F, Koroknai A, Takacs M, Salamon D, Minarovits-Kormuta S, Schwarzmann F, Wolf H, Niller HH, Minarovits J: **High-resolution analysis of CpG methylation and in vivo protein-DNA interactions at the alternative Epstein-Barr virus latency promoters Qp and Cp in the nasopharyngeal carcinoma cell line C666-1.** *Virus Genes* 2007, **35**:195-202.
13. Fejer G, Koroknai A, Banati F, Gyory I, Salamon D, Wolf H, Niller HH, Minarovits J: **Latency type-specific distribution of epigenetic marks at the alternative promoters Cp and Qp of Epstein-Barr virus.** *J Gen Virol* 2008, **89**:1364-1370.
14. Hutchings IA, Tierney RJ, Kelly GL, Stylianou J, Rickinson AB, Bell AI: **Methylation status of the Epstein-Barr virus (EBV) BamHI W latent cycle promoter and promoter activity: analysis with novel EBV-positive Burkitt and lymphoblastoid cell lines.** *J Virol* 2006, **80**:10700-10711.
15. Paschos K, Smith P, Anderton E, Middeldorp JM, White RE, Allday MJ: **Epstein-barr virus latency in B cells leads to epigenetic repression and CpG methylation of the tumour suppressor gene Bim.** *PLoS Pathog* 2009, **5**: e1000492.

16. Caliskan M, Cusanovich DA, Ober C, Gilad Y: **The effects of EBV transformation on gene expression levels and methylation profiles.** *Hum Mol Genet* 2011, **20**:1643-1652.
17. Leonard S, Wei W, Anderton J, Vockerodt M, Rowe M, Murray PG, Woodman CB: **Epigenetic and transcriptional changes which follow Epstein-Barr virus infection of germinal center B cells and their relevance to the pathogenesis of Hodgkin's lymphoma.** *J Virol* 2011, **85**:9568-9577.
18. Rodriguez J, Vives L, Jorda M, Morales C, Munoz M, Vendrell E, Peinado MA: **Genome-wide tracking of unmethylated DNA Alu repeats in normal and cancer cells.** *Nucleic Acids Res* 2008, **36**:770-784.
19. Shannon-Lowe C, Baldwin G, Feederle R, Bell A, Rickinson A, Delecluse HJ: **Epstein-Barr virus-induced B-cell transformation: quantitating events from virus binding to cell outgrowth.** *J Gen Virol* 2005, **86**:3009-3019.
20. Delecluse HJ, Hilsendegen T, Pich D, Zeidler R, Hammerschmidt W: **Propagation and recovery of intact, infectious Epstein-Barr virus from prokaryotic to human cells.** *Proc Natl Acad Sci USA* 1998, **95**:8245-8250.
21. Wu H, Zhang Y: **Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation.** *Genes Dev* 2011, **25**:2436-2452.
22. Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A: **Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair.** *Cell* 2011, **146**:67-79.
23. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR: **DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45.** *Cell* 2008, **135**:1201-1212.
24. Yamane A, Resch W, Kuo N, Kuchen S, Li Z, Sun HW, Robbiani DF, McBride K, Nussenzweig MC, Casellas R: **Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes.** *Nat Immunol* 2011, **12**:62-69.
25. Thorley-Lawson DA: **Epstein-Barr virus: exploiting the immune system.** *Nat Rev Immunol* 2001, **1**:75-82.
26. Cahir-McFarland ED, Carter K, Rosenwald A, Giltnane JM, Henrickson SE, Staudt LM, Kieff E: **Role of NF-kappa B in cell survival and transcription of latent membrane protein 1-expressing or Epstein-Barr virus latency III-infected cells.** *J Virol* 2004, **78**:4108-4119.
27. Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, Hong MY, Karczewski KJ, Huber W, Weissman SM, Gerstein MB, Korbel JO, Snyder M: **Variation in transcription factor binding among humans.** *Science* 2010, **328**:232-235.
28. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, Fitzgerald S, Fernandez-Banet J, Graf S, Haider S, Hammond M, Herrero J, Holland R, Howe K, Howe K, Johnson N, Kahari A, Keefe D, Kokocinski F, Kulesha E, Lawson D, Longden I, Melsopp C, Megy K, *et al*: **Ensembl 2007.** *Nucleic Acids Res* 2007, **35**:D610-617.
29. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, Green MR: **ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data.** *BMC Bioinformatics* 2010, **11**:237.
30. Hagman J, Ramirez J, Lukin K: **B lymphocyte lineage specification, commitment and epigenetic control of transcription by early B cell factor 1.** *Curr Top Microbiol Immunol* 2012, **356**:17-38.
31. Klein U, Casola S, Cattoretti G, Shen Q, Lia M, Mo T, Ludwig T, Rajewsky K, Dalla-Favera R: **Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination.** *Nat Immunol* 2006, **7**:773-782.
32. Wilker PR, Kohyama M, Sandau MM, Albring JC, Nakagawa O, Schwarz JJ, Murphy KM: **Transcription factor Mef2c is required for B cell proliferation and survival after antigen receptor stimulation.** *Nat Immunol* 2008, **9**:603-612.
33. Kashiwagi K, Nimura K, Ura K, Kaneda Y: **DNA methyltransferase 3b preferentially associates with condensed chromatin.** *Nucleic Acids Res* 2011, **39**:874-888.
34. Choi SH, Heo K, Byun HM, An W, Lu W, Yang AS: **Identification of preferential target sites for human DNA methyltransferases.** *Nucleic Acids Res* 2011, **39**:104-118.
35. McGinnes K, Letarte M, Paige CJ: **B-lineage colonies from normal, human bone marrow are initiated by B cells and their progenitors.** *Blood* 1991, **77**:961-970.
36. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB: **A gene atlas of the mouse and human protein-encoding transcriptomes.** *Proc Natl Acad Sci USA* 2004, **101**:6062-6067.
37. Wirtz S, Becker C, Fantini MC, Nieuwenhuis EE, Tubbe I, Galle PR, Schild HJ, Birkenbach M, Blumberg RS, Neurath MF: **EBV-induced gene 3 transcription is induced by TLR signaling in primary dendritic cells via NF-kappa B activation.** *J Immunol* 2005, **174**:2814-2824.
38. Larousserie F, Charlot P, Bardel E, Froger J, Kastelein RA, Devergne O: **Differential effects of IL-27 on human B cell subsets.** *J Immunol* 2006, **176**:5890-5897.
39. Thompson MP, Aggarwal BB, Shishodia S, Estrov Z, Kurzrock R: **Autocrine lymphotoxin production in Epstein-Barr virus-immortalized B cells: induction via NF-kappaB activation mediated by EBV-derived latent membrane protein 1.** *Leukemia* 2003, **17**:2196-2201.
40. Sayos J, Wu C, Morra M, Wang N, Zhang X, Allen D, van Schaik S, Notarangelo L, Geha R, Roncarolo MG, Oettgen H, De Vries JE, Aversa G, Terhorst C: **The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM.** *Nature* 1998, **395**:462-469.
41. Kondo T, Bobek MP, Kuick R, Lamb B, Zhu X, Narayan A, Bourc'his D, Viegas-Pequignot E, Ehrlich M, Hanash SM: **Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2.** *Hum Mol Genet* 2000, **9**:597-604.
42. Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, Ehrlich M, Laird PW: **Analysis of repetitive element DNA methylation by MethyLight.** *Nucleic Acids Res* 2005, **33**:6823-6836.
43. de Oliveira DE, Ballon G, Cesarman E: **NF-kappaB signaling modulation by EBV and KSHV.** *Trends Microbiol* 2010, **18**:248-257.
44. Young LS, Rickinson AB: **Epstein-Barr virus: 40 years on.** *Nat Rev Cancer* 2004, **4**:757-768.
45. Kuppers R, Klein U, Hansmann ML, Rajewsky K: **Cellular origin of human B-cell lymphomas.** *N Engl J Med* 1999, **341**:1520-1529.
46. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'Laughlin M, McMichael JF, Delehaunty KD, McGrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, *et al*: **DNMT3A mutations in acute myeloid leukemia.** *N Engl J Med* 2010, **363**:2424-2433.
47. Shih AH, Abdel-Wahab O, Patel JP, Levine RL: **The role of mutations in epigenetic regulators in myeloid malignancies.** *Nat Rev Cancer* 2012, **12**:599-612.
48. Palamarchuk A, Yan PS, Zanesi N, Wang L, Rodrigues B, Murphy M, Balatti V, Bottoni A, Nazaryan N, Alder H, Rassenti L, Kipps TJ, Freitas M, Croce CM, Pekarsky Y: **Tcl1 protein functions as an inhibitor of de novo DNA methylation in B-cell chronic lymphocytic leukemia (CLL).** *Proc Natl Acad Sci USA* 2012, **109**:2555-2560.
49. Bibikova M, Chudin E, Wu B, Zhou L, Garcia EW, Liu Y, Shin S, Plaia TW, Auerbach JM, Arking DE, Gonzalez R, Crook J, Davidson B, Schulz TC, Robins A, Khanna A, Sartipy P, Hyllner J, Vanguri P, Savant-Bhonsale S, Smith AK, Chakravarti A, Maitra A, Rao M, Barker DL, Loring JF, Fan JB: **Human embryonic stem cells have a unique epigenetic signature.** *Genome Res* 2006, **16**:1075-1083.
50. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc B* 1995, **57**:289-300.
51. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
52. Perez-Llamas C, Lopez-Bigas N: **Gitools: analysis and visualisation of genomic data using interactive heat-maps.** *PLoS One* 2011, **6**:e19541.
53. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB: **Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.** *Proc Natl Acad Sci USA* 1996, **93**:9821-9826.
54. Schones DE, Smith AD, Zhang MQ: **Statistical significance of cis-regulatory modules.** *BMC Bioinformatics* 2007, **8**:19.
55. Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E: **TRANSFAC: transcriptional regulation, from patterns to profiles.** *Nucleic Acids Res* 2003, **31**:374-378.

56. **Transcription Factor Binding Sites by ChIP-seq from ENCODE/HAIB..** [http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeHaibTfbs].
57. Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, Graves BJ: **DNA specificity determinants associate with distinct transcription factor functions.** *PLoS Genet* 2009, **5**:e1000778.
58. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome Biol* 2008, **9**:R137.
59. **GEO Series GSE41957.** [http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41957].

doi:10.1186/gb-2013-14-1-r3
Cite this article as: Hernando *et al.*: The B cell transcription program mediates hypomethylation and overexpression of key genes in Epstein-Barr virus-associated proliferative conversion. *Genome Biology* 2013 **14**:R3.