EPIGENETIC SILENCING MECHANISMS IN THE HUMORAL IMMUNE RESPONSE AND LYMPHOMAGENESIS

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew R. Teater

August 2017

EPIGENETIC SILENCING MECHANISMS IN THE HUMORAL IMMUNE RESPONSE AND LYMPHOMAGENESIS

Matthew R. Teater, Ph.D.

Cornell University 2017

A hallmark of diffuse large B cell lymphoma (DLBCL) pathogenesis is the perturbation of epigenetic mechanisms. Germinal center (GC) B cells, from which DLBCL originates, are characterized by a specialized phenotype enabling rapid proliferation, sustained replicative potential, and tolerance to DNA damage. This GC phenotype facilitates affinity maturation by supporting clonal expansion with concurrent mutations and rearrangements of the B cell receptor gene. We found that cytosine methylation patterns in GC B cells involve relative loss of methylation and acquisition methylation heterogeneity. We found these effects to be largely mediated by AICDA, the enzyme responsible for somatic hypermutation of immunoglobulin loci. In DLBCL, expression of AICDA leads to increased epigenetic heterogeneity, a feature linked with poor clinical outcome. The GC phenotype is also mediated in part by histone modifications. We found that the histone methyltransferase EZH2 is required for GC formation and promotes the GC phenotype by silencing proliferation checkpoint and differentiation genes via repressive H3K27me3 modification of their promoters. Notably, we found that key regulatory loci implicated in GC exit are modified by EZH2 to establish GC-specific bivalent chromatin domains. We found that repression of these GC-specific bivalent domains requires cooperation of EZH2 with the BCL6 transcriptional repressor and a noncanonical PRC1 complex. Somatic mutations enhancing the activity

of EZH2 or BCL6 can “lock in” certain oncogenic features of GC B cells, resulting in malignant transformation.

BIOGRAPHICAL SKETCH

Matthew Teater was born and raised in Idaho. He attended New York University from 2000 to 2004 and graduated *cum laude* with a Bachelor of Arts in Mathematics. He returned to New York University in 2005 and obtained a Masters of Science in Biology in 2007. After these degrees, Matt was employed by the Department of Genetic Medicine at Weill Cornell Medical College, where he participated in collection and processing of bronchoscopy samples within the laboratory of Ronald Crystal to study the gene expression effects of cigarette smoking on lung tissues and the development of chronic obstructive pulmonary disease. He returned to graduate school in 2010, enrolling in the Department of Physiology, Biophysics, and Systems Biology at Weill Cornell Graduate School of Medical Sciences, working jointly in the laboratories of Drs. Olivier Elemento and Ari Melnick to study epigenetic regulation of gene expression in germinal center B cells and germinal center-derived lymphomas.

Dedicated to my beloved wife Kimberly,

a woman whose love and support

<and patience>

have made this work possible.

ACKNOWLEDGEMENTS

I am grateful to both of my advisors, Ari Melnick and Olivier Elemento, for the critical opportunities they have provided and for their invaluable advice, support, and motivation. I am also thankful to past and present members of the Elemento lab for all their troubleshooting suggestions and for introducing me to the fundamentals of computer coding. I am thankful to past and present members of the Melnick lab for their encouragement, assistance, and, most importantly, high quality data they have given me to play with. I would like to also thank my committee members Yariv Houvras, Jason Mezey, and Harel Weinstein for their support and ad. Finally, I would like to thank the Department of Physiology, Biophysics and Systems Biology, as well as all of my colleagues, collaborators, and friends for their cooperation and efforts during this endeavor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ABC-DLBCL	Activated B cell-like DLBCL
AICDA	Activation-induced cytidine deaminase
BAFF	B cell-activating factor
BCL2	B cell lymphoma 2
BCL6	B cell lymphoma 6
BCOR	BCL6 co-repressor
BCR	B cell receptor
CB	Centroblast
CBX	Chromobox
CC	Centrocyte
CDKN1B	Cyclin Dependent Kinase Inhibitor 1B
ChIP	Chromatin immunoprecipitation
CpG	Cytosine-guanine dinucleotide
CPM	Counts per million reads
CREBBP	cAMP-response element binding protein
CSR	Class switch recombination
CXCR4	C-X-C Motif Chemokine Receptor 4
DLBCL	Diffuse large B cell lymphoma
DMC	Differenetially methylated cytosine
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DZ	Dark zone
EED	Embryonic ectoderm development
EP300	E1A Binding Protein P300
ERRBS	Enhanced reduced representation bisulfite sequencing
ESC	Embryonic stem cell
EZH2	Enhancer of zeste homolog 2
FDC	Follicular dendritic cells
FDR	False discovery rate
FL	Follicular lymphoma
FPKM	Fragments per kilobase transcript per million reads
GCB	Germinal center B cell
GCB-DLBCL	Germinal center B cell-like DLBCL
GSEA	Gene set enrichment analysis
H3K4	Histone H3 lysine 4
H3K27	Histone H3 lysine 27
Ig	Immunoglobulin
IL	Interleukin
IP	Immunoprecipitation
IRF	Interferon regulatory factor
KDM2B	Lysine demethylase 2B

KDM2D	Lysine demethylase 2D
LZ	Light zone
MAPK	Mitogen-activated protein kinase
MC	Memory B cell
NB	Naïve B cell
NES	Normalized enrichment score
PC	Plasma cell
PCA	Principal component analysis
PCGF	Polycomb group ring finger
PNA	Peanut agglutinin
PRC	Polycomb repressive complex
PRDM1	PR/SET domain 1
R-CHOP	Cyclophosphamide, doxorubicin, vincristine, and prednisone plus rituximab
RING1	Ring finger protein 1
RNA	Ribonucleic acid
RNF2	Ring finger protein 2
SD	Standard deviation
SHM	Somatic hypermutation
SRBC	Sheep red blood cell
SUZ12	Suppressor of zeste 12
TES	Transcription end site
Tfh	Follicular helper T cells
TLR	Toll-like receptor
TSS	Transcription start site
VDJ	Variable-distal-joining
VH	Variable-heavy
WT	Wild-type
XBP1	X-box-binding protein 1
Y641F	EZH2 mutant, Tyr641Phe
Y641N	EZH2 mutant, Tyr641Asn

CHAPTER ONE:

INTRODUCTION

1.1 Germinal centers

During the adaptive humoral immune response, B cells facilitate detection and elimination of pathogens by generating long-lived memory B cells (MCs) and plasma cells (PCs) with the ability to produce high-affinity antibodies[1]. One developmental route leading to differentiation of antibody-secreting PCs and MCs involves germinal centers (GCs), transient secondary lymphoid organ structures that form upon exposure to antigen by infection or immunization[2,3]. Within GCs, B cells compete for affinity-dependent signals, ensuring that B cells with higher affinity B cell receptors (BCRs), complexes formed by surface immunoglobulin (Ig) and Igα and Igβ co-receptors, are able to progressively outcompete B cells with inferior BCRs or those with autoreactive specificities. These high-affinity B cell clones subsequently exit the GC and undergo terminal differentiation into PCs and MCs.

The formation of GCs begins with acquisition of antigen by mature, naive B cells (NBs), followed by migration to the interface between the B cell follicle and the T cell zone of a secondary lymphoid organ[4,5]. At this interface, NBs receive co-stimulatory signals from CD4 helper T cells[6,7], triggering a period of intense proliferation in which responding B cells relocate preferentially within the outer B cell follicle[8]. A fraction of these proliferating B cells coalesce into tight clusters in the center of the follicle, within close proximity to a network of

stromal cells known as follicular dendritic cells (FDCs), giving rise to the early GC.

The mature GC is dividend into two compartments: a “dark” zone (DZ), representing the GC pole closest to the T cell zone in lymph nodes and spleen, and a “light” zone, the pole distal to the T zone and closest to the capsule or the marginal zone in lymph nodes or spleen, respectively. The origin of these names refers to the histological appearance seen by conventional light microscopy using traditional DNA stains. The LZ historically had appeared “lighter” due to the sparser distribution of lymphocyte nuclei among a more abundant stromal network in this compartment[9].

The DZ consists primarily of a tight cluster of rapidly proliferating B cells, known as centroblasts (CB), which strongly express the chemokine receptor CXCR4[10-13]. Centroblasts have high expression of the enzyme activation-induced cytidine deaminase (AICDA) and the error-prone DNA polymerase eta, which introduces point mutations into DNA when repairing AICDA-induced lesions[12-14]. This mutagenesis is utilized during proliferation to diversify antigen-binding regions of Ig genes via a process called somatic hypermutation (SHM). During SHM, AICDA deaminates deoxycytosines (dCs) within the antibody variable V(D)J region, converting them into deoxyuracils (dUs) to yield dU:dG mismatches [15-19]. Error-prone repair of these dU:dG mismatches can lead to subsequent point mutations and changes in amino acid sequence, thereby providing the structural substrate for selection of BCR mutants with improved binding to the immunizing antigen [19,20].

DZ B cells then move to the LZ, which is less compact and more diverse than the DZ. The LZ contains GC B cells, follicular dendritic cells, a small population of follicular helper T (Tfh) cells, and a large proportion of infiltrating naïve follicular B cells[21]. In contrast to centroblasts, LZ B cells, or centrocytes, display an activated phenotype and exhibit higher expression of activation markers, such as CD86 and CD83, as well as genes associated with BCR pathways[12,13,22]. In the LZ, some B cells may undergo class switch recombination, wherein AICDA-mediated lesions can facilitate double strand breaks and recombination of Ig switch region DNA upstream of constant heavy-chain region exons[23]. Since such recombination does not change the Ig variable region, class switching will not affect antigen specificity, but can enable different signaling capabilities due to the extended cytoplasmic tails of switched Ig isotypes. Centrocytes engaging antigen with high-affinity BCRs exhibit greater antigen capture and will present a higher density of peptide-major histocompatibility complexes (MHC) on their surfaces. This allows high-affinity B cells to receive more signaling via CD40, B cell-activating factor, and Toll-like receptors [1], which transduce via multiple pathways (*e.g.* mitogen-activated protein kinase (MAPK) and phosphoinositide 3-kinase), to activate the NFκB transcription complex[24]. NFκB activation rescues the cells from apoptotic pathways by inducing anti-apoptotic genes[25]. The NFκB transcriptional signature is detectable in a subset of LZ B cells, but is absent in the DZ[26,27]. Given the presence of antigen, Tfh cells, and an activated phenotype, the LZ have been proposed to be the site of selection for high-affinity antibody variants.

Despite their compartmentalization, GCs are dynamic and GC B cells undergo rapid flux between the DZ and LZ[11,12,21]. Following positive selection, a subset of LZ B cells is instructed to recirculate to the DZ. In the DZ, these recirculated B cells undergo further proliferation and SHM, generating subclones that potentially feature BCR mutants with further improved affinity. The dynamics of selection during affinity maturation have indicated that the LZ is constantly repopulated by mass immigration from DZ B cells (at a rate of 50% of DZ cells transitioning to the LZ over a period of ~4 hours), while less than 10% of LZ B cells return to the DZ over a period of 6 hours[12]. Mathematical modeling of these rates estimate that 10-30% of B cells that migrate to the LZ are selected to re-enter the DZ[28]. The remaining B cells will either differentiate and exit the GC or will be eliminated by apoptotic mechanisms. Recirculation between the DZ and the LZ facilitates several iterative rounds of mutation and selection, thereby promoting the generation of high-affinity PCs and MCs (**Illustration 1.1**).

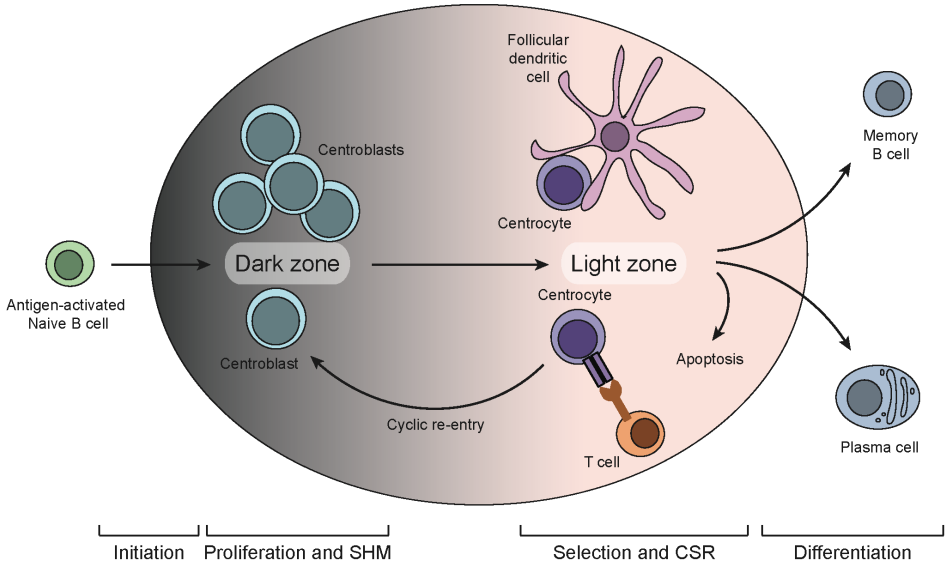

Illustration 1.1. Cellular composition and processes of the germinal center. B cells that have undergone V(D)J recombination and express functional BCRs migrate from the bone marrow to secondary lymphoid organs. Upon encounter with an antigen, naive B cells are activated by interaction with CD4+ T cells and aggregate into primary follicles to generate GCs. In the GC, B cells undergo somatic hypermutation (SHM) and class switch recombination (CSR) to remodel their immunoglobulin gene for selection of high affinity antibodies of different isotype classes. The outcome of the the GC reaction is the generation of both plasma cells and memory B cells.

The transcription factor B-cell lymphoma 6 (BCL6), originally identified from chromosomal translocations in B cell lymphomas[29], is a major regulator of the GC reaction[29-31]. *BCL6* gene expression can be initially detected in the interfollicular zone in a small subset of NBs that have successfully undergone T-cell-dependent antigen activation[32,33]. Although interferon regulatory factor 8 (IRF8), IRF4, and myocyte-specific enhancer factor 2B (MEF2B) contribute to the induction of *BCL6* expression[34-36], NBs do not produce detectable amounts of BCL6 protein[37,38]. After activation, however, BCL6 protein is abundantly expressed and B cells intrinsically require BCL6 for their

development into GC B cells[26,29,30,39,40]. BCL6 represses its target genes by binding specific DNA motifs and recruitment of histone deacetylase complexes, both directly and indirectly via interaction with multiple co-repressors[41-43]. Within GCs, *BCL6* expression is additionally regulated by an autoinhibitory mechanism in which the BCL6 protein binds its own promoter to repress further transcription[44]. BCL6 promotes the GC reaction by repressing genes implicated in cell cycle arrest, apoptosis, stress response following DNA damage, and B cell differentiation[45,46]. A subset of the promoters that are targeted by BCL6 in GCBs are also modulated by the repressive histone methyltransferase Enhancer of zeste homolog 2 (EZH2), suggesting that these two factors may cooperate during transcriptional repression of these genes[47].

EZH2, the enzymatic component of Polycomb repressive complex (PRC) 2, is critical for normal development and *Ezh2* null mice are embryonic lethal[48]. In embryonic and tissue-specific stem cells, EZH2 plays a major role in regulating gene expression patterning critical to the proper coordination of differentiation and proliferation[49]. Conditional knockout studies in early B cell differentiation has also shown EZH2 to be important for normal immunoglobulin VDJ recombination[50]. However, after the pre-B cell stage, EZH2 expression declines and is not detectable in mature B cells until the GC reaction[47,50,51]. EZH2 is highly upregulated following the GC transition, but is downregulated again when B cells differentiate and exit the GC[47,52]. Notably, EZH2 target genes in GC B cells only partially overlap with EZH2 targets in embryonic stem cells, suggesting that EZH2 has GC-specific functions[47].

In order to exit the GC and undergo terminal differentiation into PCs or MCs, GC B cells must repress BCL6 expression. BCR signaling, induced by high-affinity interactions with antigen, lead to MAPK-mediated BCL6 phosphorylation and degradation[53] and NFκB induces *IRF4*, which subsequently downregulates *BCL6* expression[54-56]. The release of BCL6-mediated silencing allows for IRF4-mediated upregulation of *PRDM1*, a master regulator of terminal B cell differentiation[57]. Expression of PRDM1 silences the 'GC B cell transcriptional program' by repressing several key transcriptional factors [58,59] and implements a PC-specific transcriptional program[60,61], which includes the acquisition of an X-box-binding protein 1 (XBP1)-dependent antibody-secreting phenotype[62].

GC B cells are unique in their ability to replicate at dramatically fast rates[11,63] while undergoing SHM and CSR. As such, the GC reaction is accompanied by suppression of genes involved in sensing or responding to DNA damage including *ATR*, *TP53,* CHEK1, and *CDKN1A*[64,65]. Additionally, GC B cells induce high expression of the enzyme telomerase, ensuring that vigorous clonal expansion does not result in telomere shortening and subsequence cellular senscence[66]. This concurrence of rapid proliferation with prolonged replicative potential and tolerance to DNA damage increases the risk of oncogenic mutations that can lead to the development of diffuse large B cell lymphoma (DLBCL) and follicular lymphoma (FL)[67,68]. Understanding how B cells impose and maintain the GC phenotype may provide important clues that could explain the pathogenesis of GC B cell-derived lymphomas.

1.2 Diffuse Large B Cell Lymphoma

B cell non-Hodgkin lymphomas comprise a range of genetically, pathologically, and clinically distinct malignancies. This heterogeneity partially reflects the diversity of the B cell system from which the lymphomas originate and the multiple pathways of transformation. The majority of B cell non-Hodgkin lymphomas, including Burkitt lymphoma, diffuse large B cell lymphoma (DLBCL), and follicular lymphoma (FL), are derived from GC B cells[69,70]. DLBCL is the most common form of B cell non-Hodgkin lymphoma, accounting for ~30-40% of newly diagnosed cases of non-Hodgkin lymphoma in adulthood[71,72]. Despite the common diagnosis, DLBCL itself is also a heterogeneous grouping of aggressive B cell lymphomas histologically characterized by diffuse proliferation of large neoplastic B-lymphoid cells with nuclear size equal to or exceeding normal histiocyte nuclei[72]. Gene expression profile analyses have classified two principle subtypes of DLBCL: GC B cell-like DLCBL (GCB-DLBCL), which is thought to arise from GC light zone B cells, and activated B cell-like DLBCL (ABC-DLBCL), which putatively originate from later stages of GC differentiation after B cells are committed to plasmablastic differentiation[13,67]. Patients with ABC-DLBCL subtype have a more aggressive clinical course and worse prognosis than those with GCB-DLBCL subtype[67,73].

For decades, the standard treatment approach for DLBCL was cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) chemotherapy[74]. The introduction of the chimeric anti-CD20 monoclonal antibody rituximab, which specifically targets B cells, to the CHOP regimen has greatly improved outcomes for DLBCL patients[75], but approximately one-

third of DLBCL patients will develop either relapsed or refractory disease[76]. In addition, the median age at DLBCL presentation is in the mid-60s, an advanced age that predisposes DLBCL patients to comorbidities that may preclude them from tolerating high doses of chemotherapy or receiving stem cell transplants. This underscores the need for novel strategies to improve current first line therapies or to establish new and more effective therapeutic interventions.

Like most cancers, the mutational landscape of DLBCLs includes numerous genetic aberrations including amplifications, deletions, and nonsynonymous point mutations associated with gain- or loss-of-function. One additional genetic alteration that is uniquely associated with B cell non-Hodgkin lymphomas, particularly DLBCL, is aberrant SHM[77]. In normal GCBs, AICDA deamination is not restricted to Ig loci and it has been reported that ~25% of highly expressed genes can be targeted by SHM[78-81]. Such "off-target" activity of AICDA is thought to contribute to oncogenic mutation and AICDA has been shown to be critical for GC-derived lymphomagenesis[82]. Consistent with this, over half of DLBCL cases exhibit mutations targeting >10% of transcribed genes[77,83] (**Figure 1.1**). The mutations may occur with both untranslated and coding regions, altering regulation or function of the target genes.

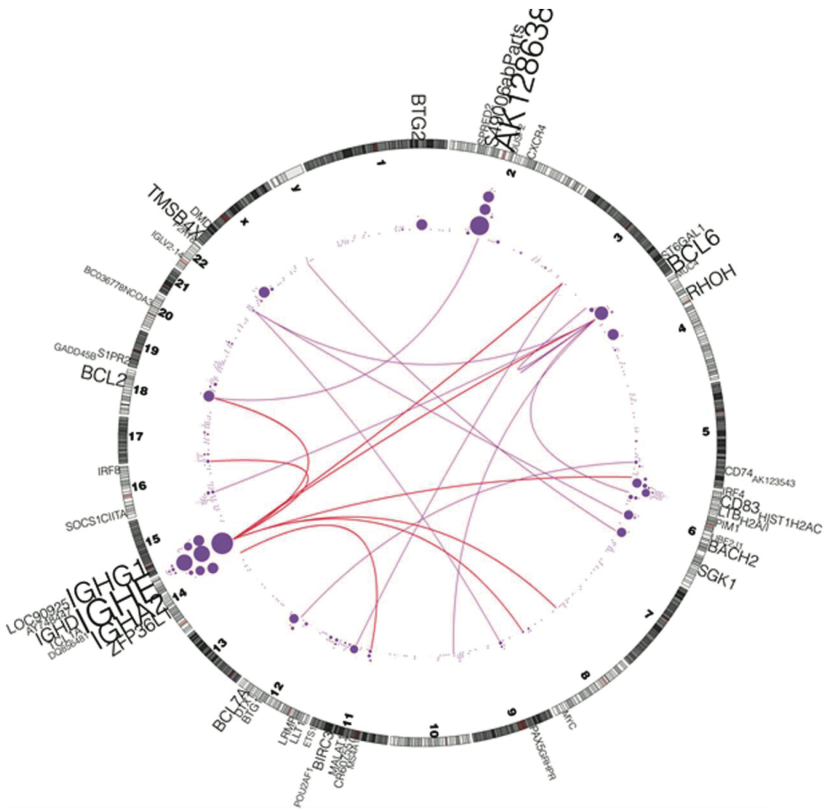

Figure 1.1 Distribution of somatic mutations in SHM-targets and correlation with genome rearrangements within DLBCL. A circos diagram showing the distribution of somatic mutations in recurrently mutated SHM-targets and genomic rearrangements such as translocations and inversions. The purple circles represent the count of SNVs in the corresponding SHM-targets, and the arcs represent the chromosomal translocation events. The red and purple arcs represent translocation involving IGH loci and non-IGH loci, respectively. The size of the circles and the gene labels are proportional to the number of mutations in the SHM-target. Reprinted from Khobabakhshi, *et al.*[83]

The cellular mechanisms regulating GC formation and maintenance are frequently implicated in the malignant transformation into lymphoma. The mutational landscape of GC-derived lymphomas suggests selection of epigenetic repression. Genetic inactivation of transcription co-activators acetyltransferse EP300 and/or CREBBP is observed in ~40% of DLBCL[84-87]. Additionally genetic aberrations in histone methyltransferase KMT2D, which catalyzes transcriptionally active histone modifications, occur in ~30% of

DLBCL cases[84-86]. An advantageous role of epigenetic silencing within malignant transformation is further supported in the GCB-DLBCL subtype, where heterozygous somatic point mutations targeting the EZH2 SET domain manifest in ~30% of cases[84]. Wild-type (WT) EZH2 displays the greatest catalytic activity for monomethylation of H3K27 and relatively weak efficiency for subsequent methylation (*e.g.* mono- to dimethylation or di- to trimethylation). Various point mutations at Y641 in EZH2, however, result in an enzyme with limited ability to catalyze the first methylation reaction (unmodified H3K27 to monomethylated H3K27), but with enhanced catalytic efficiency for the subsequent methylation reactions[88,89]. Heterozygous Y641 mutations may thus work in conjunction with WT EZH2 to increase levels of H3K27me3[88,90]. In accordance with pathogenic advantage to increased PRC2 repression, EZH2 has also been found to be highly expressed in GC-derived lymphomas[91] and inactivation (via small molecule inhibition or siRNA knockdown) of EZH2 in DLBLC cells causes acute cell cycle arrest at the G1/S transition[47,92-94]. Overall, the genetic aberrations targeting chromatin-modifying proteins suggest that lymphomagenesis may favor epigenome reprogramming to shift the balance from transcriptionally permissive states toward more transcriptionally repressive chromatin modifications (**Illustration 1.2**).

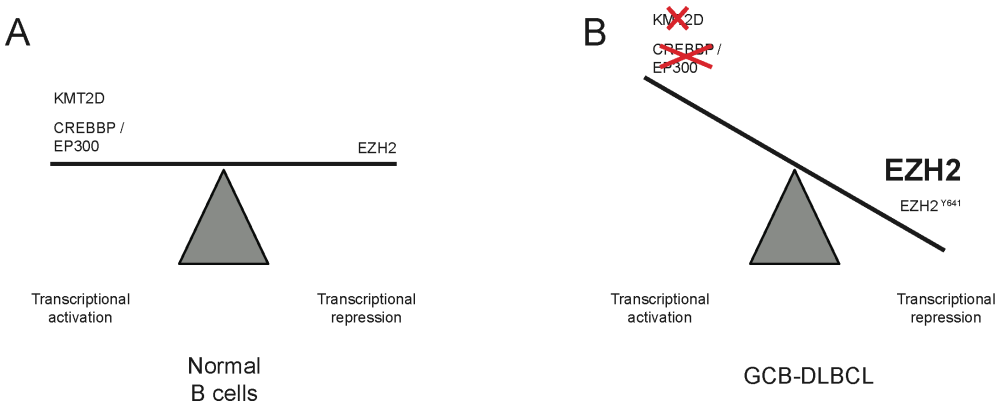

Illustration 1.2. Epigenetic alteration in diffuse large B cell lymphoma. (A) Epigenetic modifiers in normal B cells. Transcriptional co-activators EP300 and CREBBP catalyze acetylation of H3K27 and the histone methyltransferase KMT2D can add methyl groups to H3K4, both marks associated with transcriptional activity. In contrast, the histone methyltransferase EZH2 catalyzes the methylation of H3K27, which represents a marker of transcriptional repression. (B) Common genetic lesions in DLBCL involve genetic inactivation of EP300 and/or CREBBP and of KMT2D. In the GCB-DLBCL subtype, EZH2 is recurrently overexpressed or mutated to yield a gain-of-function phenotype.

1.3 Epigenetics: DNA methylation

DNA methylation describes the process by which methyl groups are covalently added to the fifth position of cytosine within a DNA molecule. Despite not altering the primary sequence of DNA, methylated cytosines play an significant role in gene silencing[95], tissue differentiation[96], mammalian development[97], and X-chromosome inactivation[98]. In mammals, the predominant form of DNA methylation occurs symmetrically on cytosine residues from both strands of cytosine-guanine (CpG) dinucleotides, although cytosine methylation is not limited to only CpG sequences[99-101]. In human somatic cells, methylated cytosines account for ~1% of total DNA bases and 70–80% of all cytosine-guanine (CpG) dinucleotides in the genome[102]. Notably, CpG dinucleotides

are unevenly distributed in the genome, being largely concentrated in small genomic regions termed “CpG islands[103].” CpG islands are found within the promoters of ~72% of human genes[104]. Although promoter-associated CpG islands tend to be unmethylated, a subset aquire methylation in specific tissues[105-107] or developmental stages [108], resulting in transcriptional repression of the adjacent genes.

DNA methylation patterns are established during development through a highly orchestrated process involving demethylation and *de novo* methylation and can be clonally inherited through the action of maintenance methyltransferase activity[108-110]. During pre-implantation development, both paternal and maternal genomes undergo substantial demethylation, erasing the majority of inherited methylation patterns. Shortly after implantation, the embryo undergoes a period of *de novo* methylation and establishes a genome-wide hypermethylation pattern[111-114]. In particular, pluripotency genes (*e.g. Oct4* and *Nanog*) and CpG island-associated germline-specific genes are tightly repressed by DNA methylation, preventing ectopic expression[115-123]. De *novo* methylation also occurs during gametogenesis, in both male and female germ cells, and is believed to play a critical role to establish genomic imprinting, an epigenetic process that yields differential methylation of paternal and maternal alleles and monoallelic expression of a small set of genes within the offspring[108,124,125]. Although d*e novo* methylation activity is present in pluripotent cells, it is largely suppressed in differentiated somatic cells and subsequent cell type-specific DNA methylation patterns are inherited via maintenance DNA methyltransferases[114,126-128].

Three active cytosine methyltransferases—DNA methyltransferase 1 (DNMT1), DNMT3A, and DNMT3B—have been identified in humans and mice[129-131]. DNMT1 is ubiquitously expressed in proliferating cells and localizes to DNA replication foci[132]. Loss of *Dnmt1* in murine ESCs induces extensive, nonspecific loss of cytosine methylation[127,133]. Furthermore, purified DNMT1 protein has strong preference for hemimethylated versus unmethylated DNA substrates and is unable to effectively initiate *de novo* methylation *in vivo*[134]. Overall, this suggests that DNMT1 functions primarily as a maintenance methyltransferase, responsible for copying the parental methylation patterns onto the daughter strand following DNA replication. In contrast, DNMT3A and DNMT3B are strongly expressed in ESCs, early embryos, and developing germ cells, but exhibit low expression in differentiated somatic cells[130,135]. Genetic studies have demonstrated that *Dnmt3a* and *Dnmt3b* are essential for *de novo* methylation in murine ESCs, as well as for *de novo* methylation of imprinted genes in the germ cells[136,137]. Although DNMT3A and DNMT3B function primarily to establish methylation patterns, they may also required for the maintenance of CpG methylation at some loci[136,138].

The relationship between cytosine methylation and gene expression is not entirely straightforward. Early observations of inverse correlations between gene expression and DNA methylation at CCGG sites in specific genes led to hypotheses that DNA methylation played an important role in the silencing gene expression[139-150]. Although later studies have substantiated a correlation between differentially methylated regions near promoters and gene expression changes[151-155], others have failed to support a major role for DNA methylation

in the regulation of gene expression[156,157]. Additionally, in cancers most aberrantly hypermethyled genes are repressed prior to hypermethylation, raising questions of a causal role for methylation in gene repression[158-161]. Furthermore, methylome profiling has revealed that, although expressed protein-coding genes generally exhibit low methylation around promoter regions, they also possess high methylation over their gene bodies[162]. The complexity of these patterns suggests that cytosine methylation is unlikely to have a clear, binary relationship with transcription.

DNA methylation is also implicated in genomic stability. Cytosine methylation outside of CpG islands is considered a major mechanism for silencing repetitive DNA elements, such as transposable DNA sequences or endoparasitic sequences, to prevent chromosomal instability, translocations and gene disruption[163,164]. Studies of murine ESCs deficient for DNMTs demonstrated a role for DNA methylation in maintaining telomere integrity[165]. DNMT3B deficiency results in expansion and rearrangements of pericentromeric repeats[136,166]. Furthermore, reduced DNMT1 activity has been implicated in increased microsatellite instability[167-171], frequency of chromosomal translocations[172], and sensitivity to genotoxic agents[171], which may promote the development of cancer.

During hematopoiesis, DNA methylation contributes to fate decisions in very early stages of hematopoietic stem cells as well as subsequent differentiating progenitors[173,174]. In hematopoietic stem cells from *Dnmt1* hypomorphic mice, lymphoid, but not myeloerythroid, genes were strikingly suppressed, skewing development toward myeloerythroid lineages with impaired B

lymphopoiesis[175]. Comprehensive methylome mapping further revealed that lymphoid lineage commitment relies more on DNA methylation for efficient suppression of myeloerthroid genes[173,174]. By contrast, Dnmt3a-disrupted hematopoietic stem cells strongly favor self-renewal over differentiation toward mature cells, a phenomenon exacerbated by combined deletion of Dnmt3b[176,177]. Among immune cells, DNA methylation profiling between GCB and NBs shows a marked shift in patterning, including differential methylation of genes affecting the NFκB and MAPK signaling pathways[179]. GCBs predominantly exhibit loss of methylation compared with NBs and AICDA-binding sites were highly over-represented among hypomethylated loci. Among DNMTs, only DNMT1 is upregulated during the GC reaction and *Dntm1* hypomorphic mice exhibit deficient GC formation and evidence of increased DNA damage[179].

DNA methylation patterning is also disrupted during lymphomagenesis[179,180]. DNMT1, DNMT3A, and DNMT3B are overexpressed in 48%, 13%, and 45% of DLBCLs and correlate with advanced clinical stages[181]. Furthermore, concomitant expression of DNMT1 and DNMT3B correlates with resistance to treatment, whereas DNMT3B overexpression correlates with shorter overall and progression-free survival. Cytosine methylation profiling studies indicate that gene promoter methylation patterning is perturbed in lymphomas compared to normal B-cells, and that promoter methylation is generally inversely correlated with gene expression[182-188]. Integrated DNA methylation and gene expression profiling can be used to classify ABC- and GCB-DLBCLs, indicating the subtypes to be epigenetically distinct entities[185]. DNA methylation profiling in normal B-cell populations, DLBCLs, and FLs also

revealed that lymphomas display striking and progressive intra-tumor heterogeneity and also inter-patient heterogeneity in their cytosine methylation patterns[189]. Epigenetic heterogeneity was initiated in normal germinal center B-cells, increased markedly with disease aggressiveness, and was associated with unfavorable clinical outcome[189].

In contrast to *de novo* DNA methylation, the identities of the enzymes in humans that catalyze DNA demethylation are largely unknown. Several studies in non-lymphoid tissues have demonstrated that AICDA can participate in loss of methylation. AICDA has been implicated in DNA demethylation during zebrafish development[190], reprogramming in heterokaryons[191] and pluripotent germ cells[192], and late reprogramming of induced pluripotent stem cells in mice[193]. Although no direct role in DNA demethylation for AICDA has been uncovered within *in vivo* GCBs, hypomethylated regions in GCBs are enriched for the putative AICDA binding site RGYW[179] and hypomethylation in GC-derived lymphomas correlates with AICDA expression[189]. As such, AICDA has been suggested to induce demethylation via base excision DNA repair of deaminated methylcytosines and replacement with unmethylated cytosines[68,69].

1.4 Epigenetics: histone modifications

In eukaryotic nuclei, genomic DNA is highly folded, constrained, and compacted by histone and nonhistone proteins into chromatin. The basic repeating unit of chromatin is the nucleosome, a complex consisting of approximately two superhelical turns of DNA wrapped around an octamer of

core histone proteins formed by four histone subunits: an H3-H4 tetramer and two H2A-H2B dimers[194]. Histones are small basic proteins that are comprised of a globular C-terminal domain and a flexible, charged NH_2-terminus, known as the "tail," that protrudes out from the nucleosome[195,196]. In conjunction with linker histone (H1), this chromatin fiber can subsequently twist and fold into increasingly more compacted filaments and lead to highr order structures (**Illustration 1.3**).

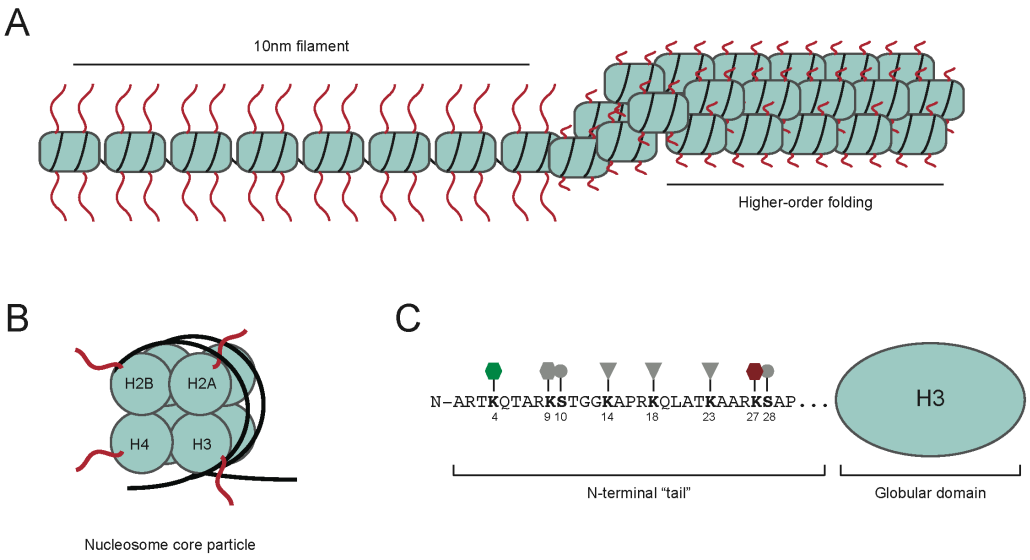

Illustration 1.3. Chromatin organization and the tail of histone H3. (A) Depiction of general chromatin organization. Histone "tails" represents a highly conserved domain that is likely to be exposed or extend outwards from the chromatin fiber. (B) A nucleosome core particle consists of approximately two superhelical turns of DNA wrapped around an octamer of histone subunits: an H3-H4 tetramer and two H2A-H2B dimer. (C) The N-terminus of human H3 is shown in single-letter amino-acid code, including a number of distinct post-translational modifications known to occur: acetylation (triangle), phosphorylation (circle), and methylation (hexagon). Methylation at H3K4 is associated with active transcription, while methylation at H3K27 is associated with repression.

Although histone proteins themselves come in both canonical or nonallelic variants with specific expression and localization[197], exquisite variation is provided by the diverse array of posttranslational modifications that can be

covalently added to the histone tail domains (**Table 1.1**). Depending on the chemical moiety and amino acid position on the tail, histone modifications can alter chromatin by modulating histone-DNA contact or provide modification-specific binding domains for effector proteins[198,199]. Several histone modifications have additionally been implicated as carriers of epigenetic information that can be transmitted through cell division, maintaining gene expression patterns within the daughter cells[200,201]. There are over 70 different residues on histones where modifications have been detected either by specific antibodies, mass spectrometry, or metabolic-labeling studies[202,203] (**Table 1.2**). However, this represents a huge underestimate of the total number of potential histone tail modifications and extra complexity is introduced by fact that methylation at lysines or arginines may be one of three different forms: mono-, di-, or trimethyl for lysines and mono- or di- (asymmetric or symmetric) for arginines. This vast array of modifications gives enormous potential for functional responses, but not all modifications occur on the same histone simultaneously. The specific regulation of a modification will be dependent on signaling conditions within the cell.

Table 1.1 Different Modifications Identified on Histones

Histone	Site	Modification	Proposed function
H1	Glu2	ADP ribosylation	Unknown
	Glu14	ADP ribosylation	Unknown
	Lys26	Methylation	Transcriptional silencing
	Ser27	Phosphorylation	Transcriptional activation
H2A	Ser1	Phosphorylation	Transcriptional repression
	Arg3	Methylation	Transcriptional activation, repression
	Lys4 (*S. cerevisiae*)	Acetylation	Transcriptional activation
	Lys5 (mammals)	Acetylation	Transcriptional activation
	Lys7 (*S. cerevisiae*)	Acetylation	Transcriptional activation
	Lys9	Biotinylation	Unknown
	Lys13	Biotinylation	Unknown
		ADP ribosylation	Unknown
	Thr119 (*D. melanogaster*)	Phosphorylation	Mitosis
	Lys119 (mammals)	Ubiquitylation	Spermatogenesis
	Thr120 (mammals)	Phosphorylation	Mitosis, transcriptional repression
	Ser122 (*S. cerevisiae*)	Phosphorylation	DNA repair
	Lys126 (*S. cerevisiae*)	Sumoylation	Transcriptional repression
	Ser129 (*S. cerevisiae*)	Phosphorylation	DNA repair
	Ser139 (mammalian H2A.X)	Phosphorylation	DNA repair
	Thr142 (mammalian H2A.X)	Phosphorylation	Apoptosis, DNA repair
H2B	Glu2	ADP ribosylation	Unknown
	Lys5	Acetylation	Transcriptional activation
	Lys6 (*S. cerevisiae*)	Sumoylation	Transcriptional repression
	Lys7	Sumoylation	Transcriptional repression
	Ser10 (*S. cerevisiae*)	Phosphorylation	Apoptosis
	Lys11 (*S. cerevisiae*)	Acetylation	Transcriptional activation
	Lys12 (mammals)	Acetylation	Transcriptional activation
	Ser14 (vertebrates)	Phosphorylation	Apoptosis
	Lys15 (mammals)	Acetylation	Transcriptional activation
	Lys16 (*S. cerevisiae*)	Acetylation	Transcriptional activation
	Lys20	Acetylation	Transcriptional activation
	Lys30	ADP ribosylation	Unknown
	Ser33 (*D. melanogaster*)	Phosphorylation	Transcriptional activation
	Ser36	Phosphorylation	Transcriptional activation
	Lys120 (mammals)	Ubiquitylation	Meiosis
	Lys123 (*S. cerevisiae*)	Ubiquitylation	Transcriptional activation

Table 1.1 (Continued)

H3	Arg2	Methylation	Transcriptional repression
	Thr3	Phosphorylation	Mitosis
	Lys4 (*S. cerevisiae*)	Acetylation	Transcriptional activation
		Methylation	Transcriptional activation
		Biotinylation	Gene expression
	Thr6	Phosphorylation	Unknown
	Arg8	Methylation	Transcriptional activation, repression
	Lys9	Acetylation	Transcriptional activation
		Methylation	Transcriptional repression
		Biotinylation	Gene expression
	Ser10	Phosphorylation	Transcriptional activation
	Thr11 (mammals)	Phosphorylation	Mitosis
	Lys14	Acetylation	Transcriptional activation
	Arg17	Methylation	Transcriptional activation
	Lys18	Acetylation	Transcriptional activation
		Biotinylation	Gene expression
	Lys23	Acetylation	Transcriptional activation
	Arg26	Methylation	Transcriptional activation
	Lys27	Acetylation	Transcriptional activation
		Methylation	Transcriptional repression
		ADP ribosylation	Unknown
	Ser28 (mammals)	Phosphorylation	Mitosis
	Lys36	Acetylation	Transcriptional activation
		Methylation	Transcriptional activation
	Lys37	ADP ribosylation	Unknown
	Tyr41	Phosphorylation	Transcriptional activation
	Arg42	Methylation	Transcriptional activation
	Tyr45	Phosphorylation	Apoptosis
	Lys56 (*S. cerevisiae*)	Acetylation	Transcriptional activation
	Lys79	Methylation	Transcriptional activation
H4	Ser1	Phosphorylation	DNA repair
	Arg3	Methylation	Transcriptional activation, repression
	Lys5	Acetylation	Transcriptional activation
	Lys8	Acetylation	Transcriptional activation
	Lys12	Acetylation	Transcriptional activation
		Biotinylation	DNA damage response
	Lys16	Acetylation	Transcriptional activation
	Lys20	Methylation	Transcriptional repression
	Lys59	Methylation	Transcriptional silencing
	Lys91 (*S. cerevisiae*)	Acetylation	Chromatin assembly

Large-scale mapping of histone modifications and chromatin-associated proteins using chromatin immunoprecipitation combined with high-throughput sequencing (ChIPseq) has enabled characterization of both the determinants and functional consequences of chromatin structure across the genome in diverse cell types. Two histone modifications, in particular, play crucial opposing roles in the epigenetic regulation of numerous developmental genes. Methylation of histone H3 lysine 4 (H3K4), catalyzed by the trithorax homolog myeloid-lymphoid leukemia (MLL)[206], is associated with transcriptional activation[207-210], whereas trimethylation of H3K27 (H3K27me3), catalyzed by the Polycomb-group protein EZH2, is associated with transcriptional repression (**Illustration 1.4A**). Polycomb-group proteins form the multisubunit Polycomb-repressive complexes (PRCs) 1 and 2[211,212]. PRC2 catalyzes H3K27 methylation, a pivotal mark in the establishment of repressive chromatin in both early development and adult organisms[213-215]. PRC1 encompasses a diverse range of complexes that all contain the RING1/RNF2 ubiquitin ligase and several additional subunits[212,216,217].

Polycomb target genes are often jointly occupied and repressed by PRC1 and PRC2, in part because H3K27me3 can bind a Polycomb-group Chromobox (CBX) family member protein and recruit PRC1. However, in ESCs, a proportion of PRC2 targets are not occupied by PRC1[218] and PRC1 has capacity to bind genes independently of PRC2[219]. Furthermore, even at jointly bound promoters, PRC1 and PRC2 act to repress genes independently and redundantly[220]. Notably, there is a cohort of developmental gene promoters in ESCs that exhibit a distinct histone modification signature consisting of

simultaneous active H3K4me3 and repressive H3K27me3 modifications[218,219,221-224] (**Illustration 1.4B**). This pattern of opposing histone marks on the same promoter, referred to as "bivalent" domains[225], are most prevalent in ESCs, but have been observed in cell types of restricted potency[226,227]. By exhibiting both active and repressive features, bivalently marked genes "poise" developmental genes, enabling rapid activation upon suitable developmental cues while maintaining minimal expression levels in the absence of differentiation signals[225].

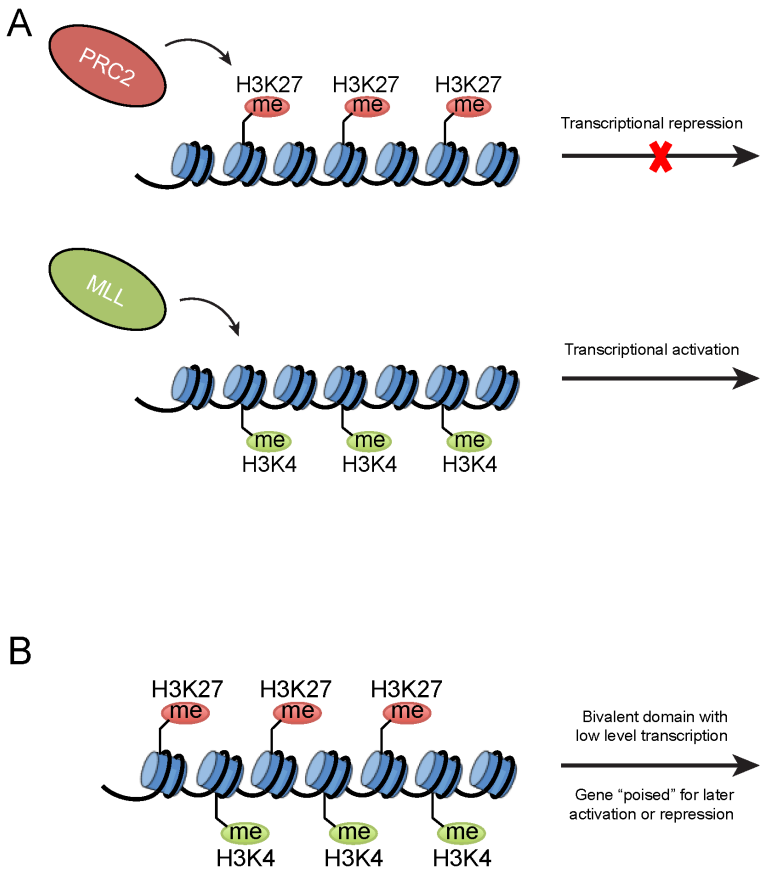

Illustration 1.4. Transcriptional "poising" at bivalent domains. (A) PRC2 catalyzes H3K27me associated with transcriptional repression, whereas myeloid-lymphoid leukemia catalyzes methylation of H3K4 associated with transcriptional activation. (B) In bivalent domains, both marks are present and associated wtih a low level of transcription; genes thus marked are thought to be poised for later transcriptional repression or activation.

During the GC reaction, PRC2 subunits, including the enzymatic subunit EZH2, are highly upregulated [52], allowing for the initiation of a repressive epigenetic program via addition of H3K27me3 to promoters[47]. This GC-specific repression program that includes silencing of cell cycle checkpoint genes and differentiation factors, facilitating rapid proliferation and maturation of the GCs[47]. In mammals, H3K27 methylation is catalyzed by the SET domain of EZH2 and requires the presence of two additional PRC2 subunits, embryonic ectoderm development (EED) and suppressor of zeste 12 (SUZ12)[228,229].

CHAPTER TWO:

DNA METHYLATION DYNAMICS DURING GC TRANSITION AND ROLE WITHIN GC-DERIVED LYMPHOMAGENESIS*

2.1 No correlation between cytosine methylation alterations and gene expression changes during GC transition

Although DNA methylation and gene expression changes are linked in pluripotent cells, it is not clear how correlated the relationship is in somatic cells. To investigate the effect of methylation levels on gene expression, sorted NBs (CD20+IgD+CD77) and GCBs (CD20+IgD CD77+) from reactive human tonsils and profiled their methylome using enhanced reduced representation sequencing (ERRBS), an efficient single-nucleotide resolution high-throughput technique that interrogates 2–4 million distinct CpGs[230]. Upon rigorous quality control of bisulfite conversion (>99.5% in all samples) and read mapping frequency (>70%), we quantified the levels of cytosine methylation at represented CpGs in both NB and GCBs. We found that transcriptionally inactive or lowly expressed genes exhibited more methylation

---

* Dominguez PM*, **Teater M***, Chambwe N, Kormaksson M, Redmond D, Ishii J, Vuong B, Chaudhuri J, Melnick A, Vasanthakumar A, Godley LA, Papavasiliou FN, Elemento O, Shaknovich R. (2015) DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID. *Cell Reports 12(12):2086-98*

* **Teater M***, Dominguez PM*, Redmond D, Chen Z, Ennishi D, Scott DW, Cimmino L, Ghione P, Chaudhuri J, Gascoyne RD, Aifantis I, Inghirami G, Elemento O, Melnick A, Shaknovich R. (2017) AICDA drives epigenetic heterogeneity and accelerates germinal center-derived lymphomagenesis. *In review*

in the region immediately surrounding the transcription start site (TSS), while genes with higher expression where depleted for cytosine methylation in these regions (**Figure 2.1A, B**). Notably, we also found at regions distal to the TSS in the gene body, highly expressed genes exhibited more cytosine methylation than their lowly expressed counterparts, consistent with previous studies[162]. Although we did not see an obvious difference in the relationship of cytosine methylation and transcription between the two cell types, we were interested in determine whether the gene expression changes observed during the GC transition were correlated with the methylation changes. To quantify gene expression changes during the GC transition, we performed RNA sequencing (RNAseq) on sorted human tonsillar NB and GCB populations. Comparing the change in mean methylation within the region ±250bp of TSS (GCB – NB) with the log2 ratio of gene expression of GCB compared to NB, we find no correlation (**Figure 2.1C**). Although we found a subset of genes with alteration in cytosine methylation levels in GCB, these genes did not vary in expression from GCB to NB. Similarly, we found that genes that changed expression during the GC reaction, where up- or downregulated, had generally constant methylation levels. Overall, we found no correlation between GC changes in methylation level at TSS and in expression of the associated genes (Pearson correlation coefficient = 0.0346).

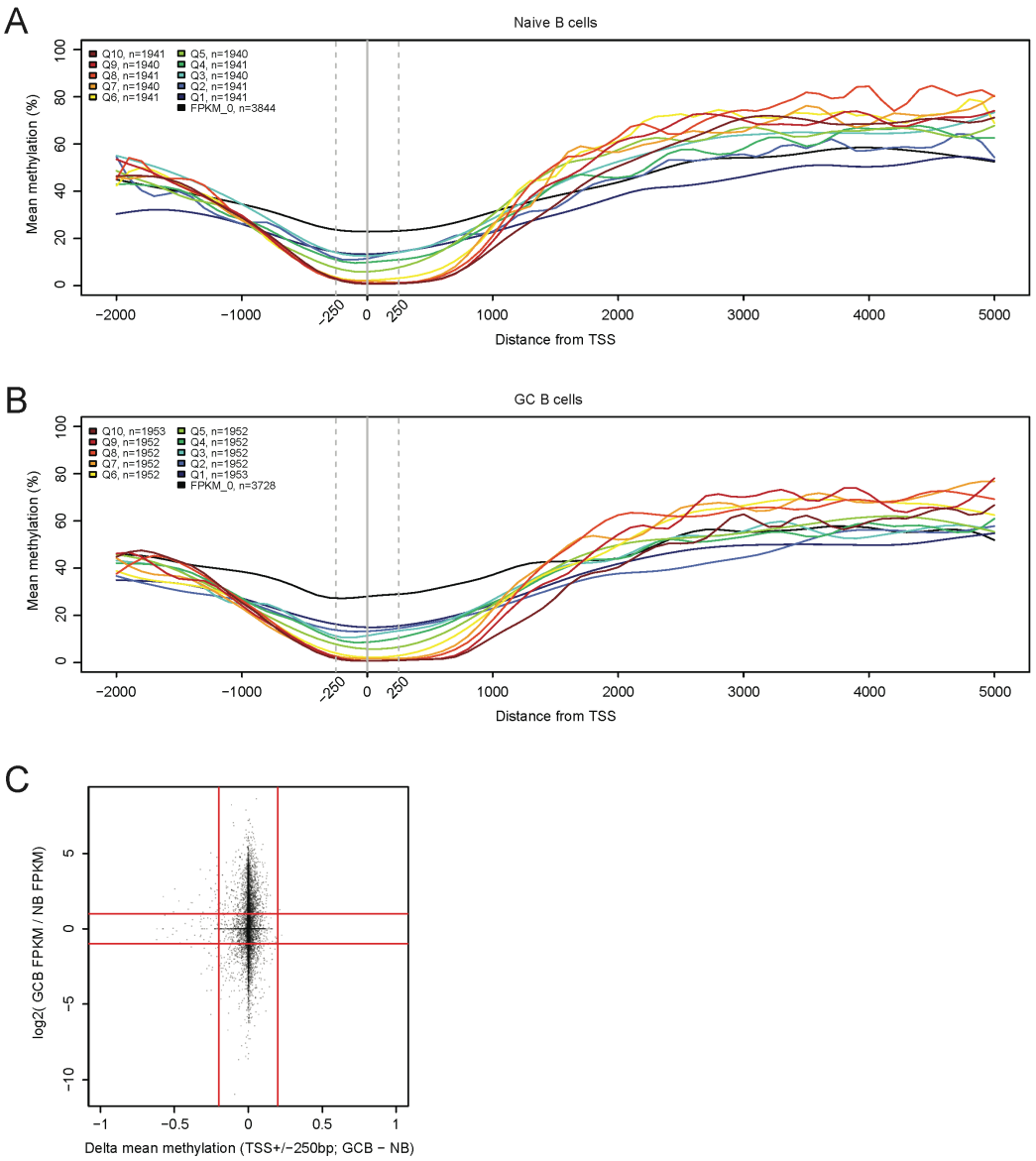

Figure 2.1. DNA methylation within gene promoter is associated with low expression, but shows no correlation with expression changes during GC transition. (A) Mean DNA methylation in proximity to gene promoters in human tonsillar naive B cells, separated into categories representing genes with no expression ("FPKM_0") and deciles of expressed genes according to FPKM (Q1-Q10). (B) Mean DNA methylation in proximity to gene promoters in human GC B cells. (C) Scatterplot showing relationship between change in mean DNA metylation at TSS versus change in gene expression during GC reaction. Horizontal red lines indicate fold change threshold of 2 and vertical red lines indicate methylation change threshold of 20%.

2.2 Loss of AICDA abrogates CpG methylation alteration during GC transition

Previous studies observed significant loss of DNA methylation in human GC B cells[179,231]. To investigate whether AICDA was responsible, at least in part, for such a decrease in DNA methylation, we induced T cell-dependent GC formation with 4-NP-chicken gamma globulin (NP-CGG) in wild-type (WT; n=7 replicates) and *Aicda*$^{-/-}$ (n=6 replicates) mice. Mice were sacrificed at day 10 post-injection, and splenic NBs (B220^{+}GL7^{-}CD95^{-}) and GC B cells (B220^{+}GL7^{+}CD95^{+}) were isolated. We then profiled genome-wide methylation changes between respective NB and GC B cells using ERRBS. Upon rigorous quality control of bisulfite conversion (>99.5% in all samples) and read mapping frequency (>70%), we called differentially methylated CpGs (DMCs) between NBs and GC B cells using a combination of statistical difference (false discovery rate [FDR] < 0.001 using Fisher's exact test) and methylation level difference greater than 20%. We observed that NB to GC B cell transition in WT mice was accompanied by significant changes in DNA methylation, including 8,308 hypomethylated DMCs (hypo-DMCs) and 3,390 hypermethylated DMCs (hyper- DMCs) (**Figure 2.2A and B**). This is consistent with previous results showing a genome-wide loss of methylation in primary human GC B cell samples compared to NBs[179]. In contrast, our profiling of *Aicda*$^{-/-}$ animals showed minimal changes in DNA methylation during the transition from NBs to GC B cells: only 703 of CpGs exhibited hypomethylation and 172 CpGs exhibited hypermethylation (**Figure 2.2B**). To query more comprehensive, global shifts in methylation, including at non-differentially methylated CpGs, we examined the distribution of methylation level changes for all represented CpGs and found that *Aicda*$^{-/-}$ mice

manifested reduced cytosine methylation differences during the NB to GC B cell transition, suggesting that loss of AICDA also resulted in less methylome plasticity (**Figure 2.2C**). This occurred despite comparable ERRBS coverage in WT and *Aicda*$^{-/-}$ cells and similar global methylation levels, as measured using liquid chromatography-mass spectrometry (LC-MS), in NBs from WT and *Aicda*$^{-/-}$ mice (**Figure 2.2D**). LC-MS analysis also revealed higher genome-wide levels of 5mC in *Aicda*$^{-/-}$ GC B cells compared to WT GC B cells (**Figure 2.2D**). Our results indicate that AICDA is responsible for the majority of the methylome changes that B cells undergo during their transit through the GC.

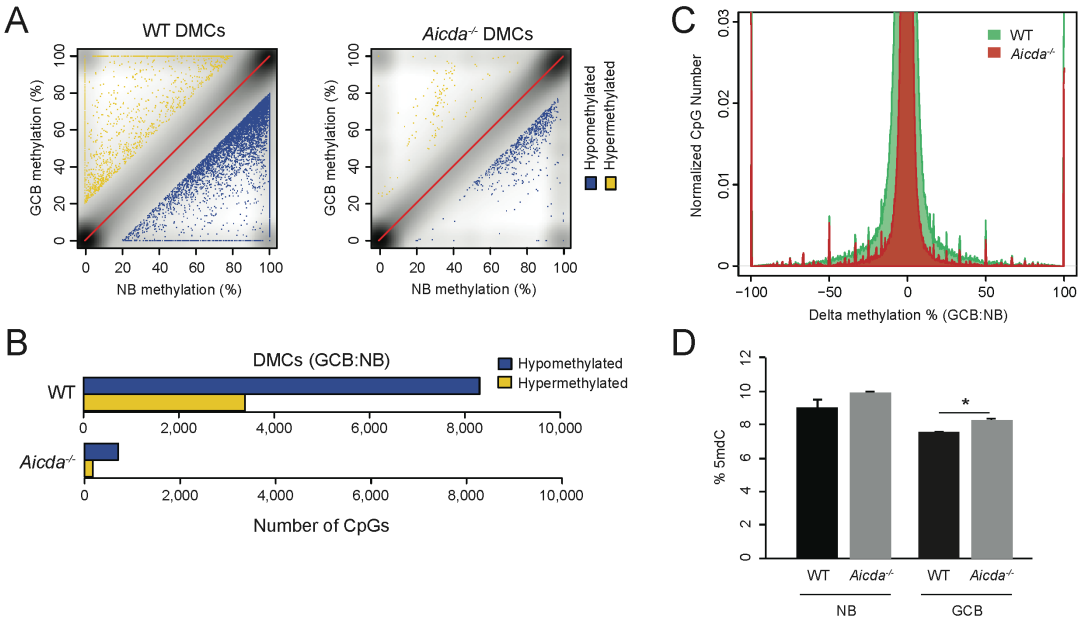

Figure 2.2. Loss of AICDA abrogates CpG methylation changes during GC transition. (A) Combined CpG methylation of NBs and GCBs from WT mice (seven repliates; left) and *Aicda*$^{-/-}$ mice (six replicates; right), determined by ERRBS using a 20% methylation difference and FDR < 0.001, Fisher's exact test. Hypo-DMCs are indicated in blue and hyper-DMCs are indicated in yellow. (B) Number of DMCs between GCBs and NBs from WT (top) and *Aicda*$^{-/-}$ (bottom) mice. (A) and (B) show that the DNA methylation changes characteristic of GC transition in WT mice are abrogated in *Aicda*$^{-/-}$ mice. (C) Density plot showing delta methylation values (GCB% - NB%) of all CpGs determined by ERRBS in WT and *Aicda*$^{-/-}$ mice. *Aicda*$^{-/-}$ delta methylation absolute values are decreased compared to WT, indicating less change in methylation during GC transition. (D) Liquid chromatography-mass spectrometry (LC-MS) was performed in n=2 replicates each for WT NB, *Aicda*$^{-/-}$ NB, WT GCB, and *Aicda*$^{-/-}$ GCB replicates, showing percentage of 5mdC in total cytosines in each condition (* $p<0.05$, t-test). Mass spectrometry was performed by M. Dominguez.

2.3 AICDA facilitates epigenetic diversity within GC B cells

We hypothesized that AICDA might also be responsible for the previously described increased in methylation diversity within GC B cells compared to NBs[189]. Taking advantage of the rectilinear properties of "Manhattan" distance, we devised a method by which we were able to compare the global similarity of any two ERRBS profiles by effectively assessing the number of methylation "events" that separated them. By applying this "pairwise methylation distance" metric to all possible pairs among our NB and GCB profiles from the WT and *Aicda*$^{-/-}$ mice, we were able to evaluate the degree of epigenetic diversity both among the respective groups and between cell types, without limiting the analysis to the smaller set of CpGs re commonly represented among all profiles.

We found that WT GC B cell replicates had greater pairwise methylation distances to each other than was observed amongst WT NB samples, corresponding to higher average methylation diversity (NB WT:GC B cell WT, Wilcoxon p = 4.50e-09; **Figure 2.3A and B**). This is consistent with epigenetic diversification of B cells during their passage through the GC. Importantly, *Aicda*$^{-/-}$ GCB replicates displayed significantly lower intra-group methylation distance than WT GCB replicates (GCB WT:GCB *Aicda*$^{-/-}$, Wilcoxon p = 3.59e-10; **Figure 2.3A and B**), indicating that loss of AICDA resulted in a more homogenous GCB methylome with diversity more akin to that of the *Aicda*$^{-/-}$ NB methylome (NB *Aicda*$^{-/-}$:GCB *Aicda*$^{-/-}$, Wilcoxon p = 0.000323; **Figure 2.3A** and **Figure 2.3B**). We also found lower pairwise methylation distance during

the transition from *Aicda*$^{-/-}$ NBs to GCBs than from WT NBs to GCBs (Wilcoxon p = 1.76e-20; **Figure 2.3A and B**). This decreased diversity is consistent with the abrogation of methylation changes during the NB to GCB transition observed in AICDA-deficient mice (**Figure 2.2**).

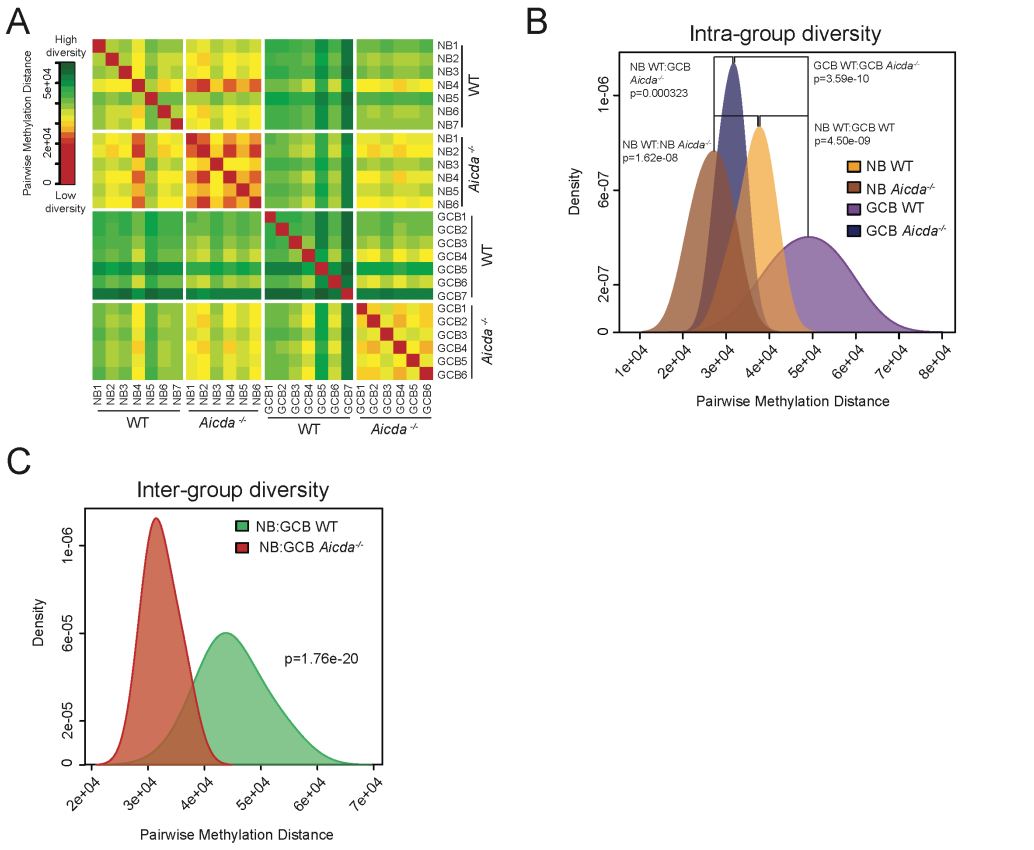

Figure 2.3. AICDA facilitates epigenetic diversity within GC B cells. (A) Heatmap showing pairwise methylation distance between ERRBS of NB and GCB replicates from WT and *Aicda*$^{-/-}$ mice. (B) Density plot showing the distribution of intra-group pairwise methylation distances between ERRBS profiles of NB and GCB replicates from WT and *Aicda*$^{-/-}$ mice. WT GCBs have greater pairwise distance than WT NBs, indicating increased diversity among methylation profiles. On the contrary, *Aicda*$^{-/-}$ GCB replicates have lower pairwise distance than their WT counterparts, closer to *Aicda*$^{-/-}$ NB samples. (C) Density plot showing distribution of pairwise methylation distances between NBs and GCBs from WT and *Aicda*$^{-/-}$ mice. NB to GCB transition in *Aicda*$^{-/-}$ mice is associated with lower pairwise distance than in WT mice, indicating less cytosine methylation changes between *Aicda*$^{-/-}$ NB and GCB profiles.

Previous experiments within an *ex vivo* system showed no cytosine methylation changes following depletion or overexpression of AICDA[232]. In order to reconcile our *in vivo* AICDA-dependent methylation changes with these earlier observations, we activated CD43 $^{-}$ splenic WT cells (NB) in the presence of lipopolysaccharide (LPS), interleukin-4 (IL-4), and anti-CD40 and infected them with either empty vector (EV) or a vector expressing the full-length *Aicda* cDNA (AICDA). We confirmed by qPCR and western blot that the AICDA-overexpressing cells expressed higher levels of AICDA than EV-infected cells (**Figure 2.4A**). Moreover, we detected a higher percentage of class-switched splenocytes when AICDA was overexpressed (**Figure 2.4B and C**). We performed ERRBS profiling on sorted GFP+IgG1+ cells (class-switched infected cells) and observed few differences in methylation between AICDA-overexpressing and EV-infected B cells, with less than 1,000 hyper- and hypo-DMCs (**Figure 2.4D**). We also calculated the pairwise methylation distance between all profiles and found a high degree of homogeneity among all sample ERRBS profiles, suggesting that methylation diversity between AICDA-overexpressing and EV-infected B cells was similar and comparable to the diversity observed in our NB controls (**Figure 2.4E**). These results support the previous findings obtained by Fritz *et al*[232]. The *ex vivo* stimulated B cells differ from GCBs in that they present significantly reduced levels of SHM—predominantly targeting the Sμ region—compared to GCBs induced *in vivo*[78,233,234]. Reduced SHM and the absence of methylome modification in *ex vivo* activated B cells support the hypothesis of convergence of the mechanisms for SHM repair and methylome editing.

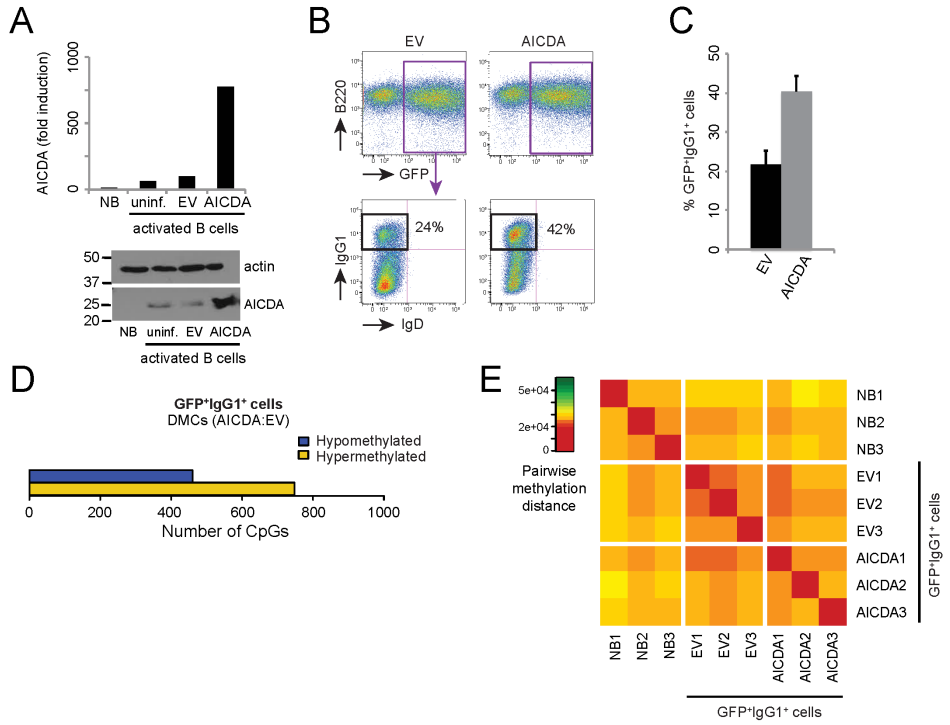

Figure 2.4. AICDA-overexpressing *ex vivo* stimulated splenic B cells do not show increased DNA methylation heterogeneity. (A) AICDA mRNA (top) and protein (bottom) expression in unstimulated CD43$^-$ splenic cells (NB) or stimulated splenic B cells (activated B cells). Activated B cells were stimulated with LPS, IL-4 and anti-CD40 for 96 h and left uninfected (uninf) or were simultaneously infected with either empty vector (EV) or AICDA-expressing pMIG vector (AICDA). (B) Flow cytometry analysis of activated B cells. Infected cells were GFP$^+$ and the class-switched cells among the GFP$^+$ fraction were identified by IgG1 expression. (C) Percentage of GFP$^+$ IgG1$^+$ splenic B cells between EV and AICDA-expressing vector conditions. (D) DMCs between AICDA-expressing vector and EV infected *ex vivo* splenic B cells were determined by ERRBS using a 20% methylation difference threshold and FDR < 0.001, Fisher's exact test. HypoDMCs are indicated in blue and hyperDMCs are indicated in yellow. (E) Heatmap showing pairwise methylation distance between ERRBS profiles from NB and GFP$^+$ IgG1$^+$ splenic B cells infected with AICDA-expressing vector or EV. *Ex vivo* culturing and flow cytometry experiments were performed by M. Dominguez.

2.4 Methylation changes in *Aicda*$^{-/-}$ GCBs are not due to changes in the cellular composition or clonality within the GC

In order to rule out the possibility that abrogation of methylation changes in *Aicda*$^{-/-}$ GCBs arises due to changes in the cellular composition within the GC (content of CBs versus CCs) or clonal diversity, we carried out detailed analysis of GCBs from WT and *Aicda*$^{-/-}$ animals. Both WT and *Aicda*$^{-/-}$ animals

had the same proportion of CB (CXCR4^{high}) and CC (CXCR4^{low}) within the GC (**Figure 2.5A**). To confirm this result, we performed RNAseq on GCBs isolated from WT and *Aicda*$^{-/-}$ mice and compared their expression profiles for the genes that constitute the CB (DZ) and CC (LZ) signatures, identified by Victora *et al*.[12] (**Figure 2.5B**). The expression for these CB- and CC-specific genes was highly correlated between WT and *Aicda*$^{-/-}$ cells (Pearson correlation coefficient = 0.984 for DZ genes and 0.989 for LZ genes), indicating that both genotypes had comparable gene expression profiles. In addition, we investigated whether there were any differences in clonal complexity in the GC between WT and *Aicda*$^{-/-}$ mice. For that purpose, we amplified rearranged IgH, Igκ, and Igλ regions using primers capturing the most abundant families of Ig rearrangements[235-237] and performed high-throughput sequencing using the Illumina MiSeq (PE2x150). Statistical analysis of the Ig rearrangements revealed no significant difference in clonal complexity and or composition of VH regions between WT GCBs and *Aicda*$^{-/-}$ GCBs (Wilcoxon p=0.8571; **Figure 2.5C**). Furthermore, we observed no differences in the cell-cycle distribution based on bromodeoxyuridine (BrdU) incorporation and 7-AAD staining of GCBs from WT and *Aicda*$^{-/-}$ mice (**Figure 2.5D**). Altogether, these findings indicate that the composition and the clonality of WT and AICDA-deficient GCBs are equivalent.

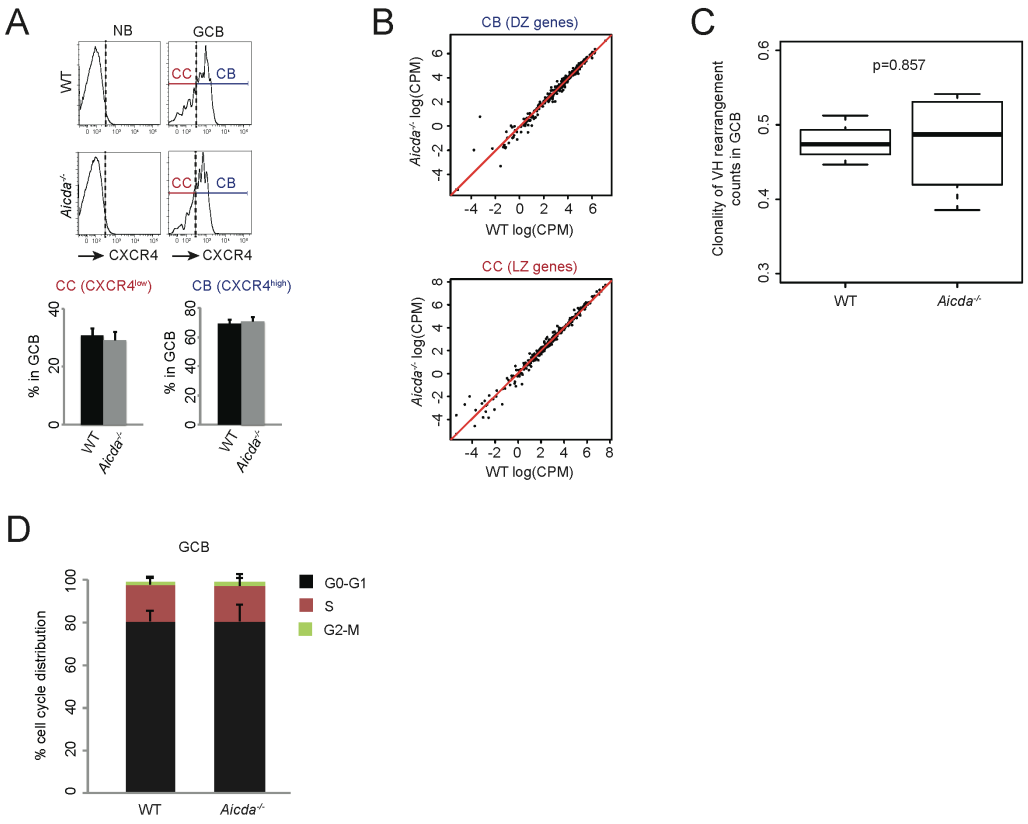

Figure 2.5. AICDA-deficient *in vivo* GC B cell populations do not show bias in proportion of centroblasts to centrocytes. (A) Percentage of CB and CC according to CXCR4 expression levels in GCB from WT (n=3) and *Aicda*$^{-/-}$ (n=3) mice. Expression of CXCR4 in NB is shown as negative control. (B) Scatterplot comparing the mean log-transformed expression of CB and CC gene signatures between GCB replicates of WT and *Aicda*$^{-/-}$ mice. (C) Boxplot showing similarity of clonal composition based on VH rearrangements in WT and *Aicda*$^{-/-}$ GCB. (D) Percentages of cells in G0/G1, S and G2/M phases, detected by BrdU incorporation and 7-AAD staining in GCB from WT (n=5) and *Aicda*$^{-/-}$ (n=5) mice. Flow cytometry and cell cycle experiments were performed by M. Dominguez. Clonality analysis was performed by D. Redmond.

2.5 AICDA-dependent hypomethylation in GC B cells is enriched to occur at SHM hotspot genes and dsDNA breaks

To investigate the genomic distribution of AICDA-dependent methylation changes, we defined AICDA-dependent hypo- and hyper-DMCs as CpGs that are hypomethylated (blue rectangle, **Figure 2.6A**) or hypermethylated (yellow

rectangle, **Figure 2.6A**) during NB to GCB transition in WT animals but show no respective differential methylation changes in *Aicda*$^{-/-}$ animals. We found that these AICDA-dependent hypo- and hyper-DMCs were significantly depleted in promoters of genes and enriched in introns and intergenic areas (Fisher's exact test, $p < 0.001$; **Figure 2.6B**). AICDA-dependent hypo- and hyper-DMCs were also depleted in CpG islands, shores, and shelves and enriched in inter-island regions (Fisher's exact test, $p < 0.001$; **Figure 2.6C**). Since AICDA-dependent DNA demethylation is thought to occur via deamination and subsequent DNA repair, similar to SHM, we investigated whether AICDA-dependent DMCs were enriched in genes reported to be targets of SHM in GCBs[78]. Notably, AICDA-dependent hypo-DMCs were enriched in occur within SHM hotspot genes (Fisher's exact test, $p < 0.001$; **Figure 2.6D**). As SHM occurs at highly expressed genes, we also tested if AICDA-dependent DMCs were enriched in genes highly expressed in WT GCBs (fragments per kilobase of transcript per million mapped reads (FPKM) > 20). We found no enrichment for hypo-DMCs (data not shown), suggesting that hypomethylation results from AICDA targeting specific genomic loci, not simply as a consequence of open chromatin structure or regions with high transcriptional activity. We also found enrichment of AICDA-dependent hypo-DMCs in AICDA-associated double-stranded DNA (dsDNA) breaks identified using high-throughput genomic translocation sequencing[238] (Fisher's exact test, $p < 0.001$; **Figure 2.6D**) and in loci associated with double-strand breaks defined through γ-H2AX occupancy[239] (Fisher's exact test, $p < 0.001$; **Figure 2.6D**), suggesting an association between AICDA-dependent hypomethylation and DNA breaks.

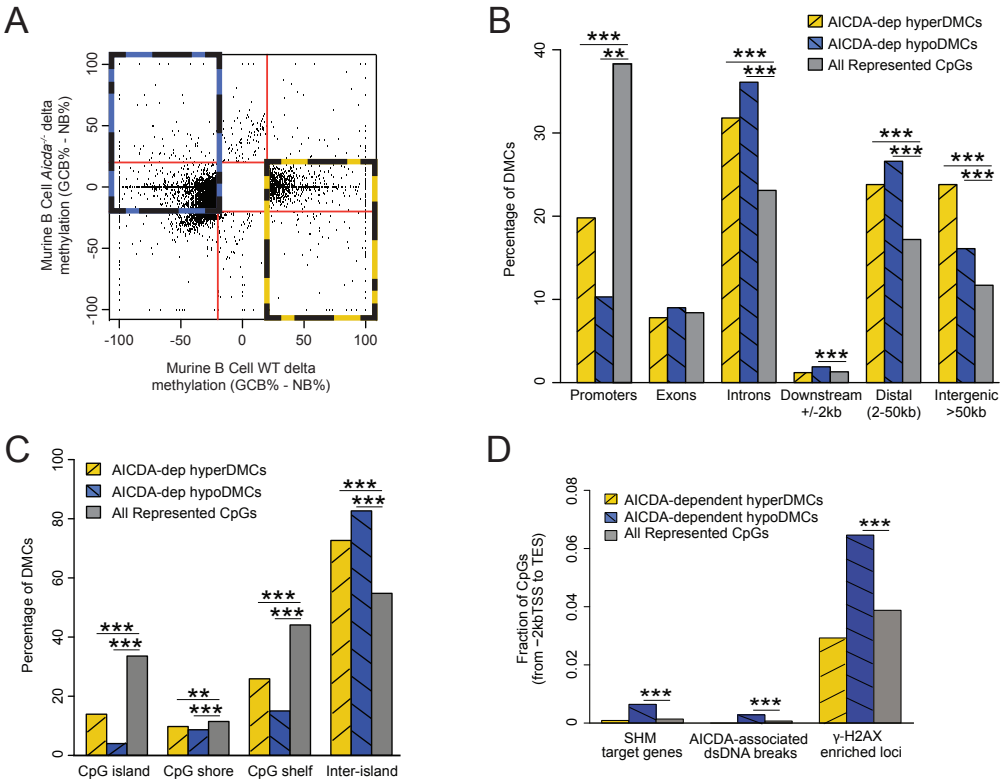

Figure 2.6. AICDA-dependent DNA methylation changes show characteristic distribution. (A) Scatterplot comparing delta methylation (GCB% - NB%) of WT versus *Aicda*$^{-/-}$ mice. AICDA-dependent hypo-DMCs are indicated by a yellow dashed box and hyper-DMCs by a blue dashed box. (B) Bar plot showing genomic distribution of AICDA-dependent hypo-DMCs and hyper-DMCs as well as all CpGs represented within ERRBS experiments. AICDA-dependent DMCs are depleted within promoters and over-represented in introns and distal and intergenic regions. (C) Bar plot showing the distribution of AICDA-dependent hypo-DMCs and hyper-DMCs within CpG islands, shores, shelves, and inter-island regions. AICDA-dependent DMCs are depleted in CpG islands, shores, and shelves and enriched in inter-island regions. (D) Fraction of AICDA-dependent DMCs enriched in SHM hotspot genes, AICDA-mediated dsDNA breaks, and loci with γH2AX occupancy (*p < 0.05, **p < 0.01, and ***p < 0.001).

The AICDA/APOBEC family of proteins has been known to contribute to intrinsic immunity against retrotransposition of endogenous and exogenous retroviruses[240]. Endogenous retroviruses are present in multiple copies in mammalian genomes and constitute up to a staggering 8% of human and mouse genomic DNA [241,242]. As AICDA-dependent DMCs are enriched in

introns and intergenic regions (**Figure 2.6B**), we investigated whether AICDA targets repetitive elements present in those regions of the genome. We annotated our AICDA-dependent DMCs according to RepeatMasker and identified six intergenic repetitive elements that were significantly enriched for the presence of AICDA-dependent DMCs, including L1 repeat element, IAPEY3_LTR, and MLT1J1 (**Figure 2.7A**). We also found two intragenic repetitive elements that were significantly enriched to contain AICDA-dependent DMCs (**Figure 2.7B**). In summary, although we do find cases of enrichment at specific repetitive elements, AICDA appears to acts mostly upon non-repetitive DNA sequences.

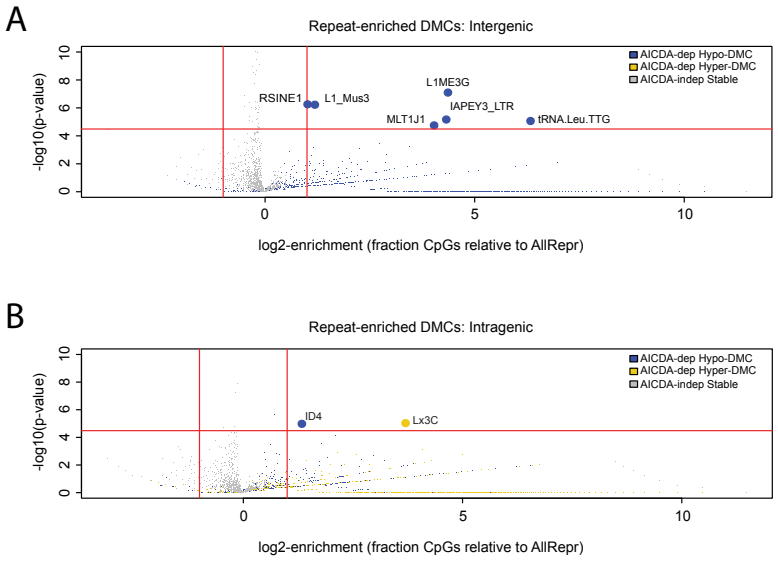

Figure 2.7. AICDA-dependent changes in DNA methylation targeting repetitive elements. (A) Volcano plot identifying intergenic repeats enriched for AICDA-dependent DMCs. Red lines indicate thresholds for two-fold enrichment versus all represented CpGs and Bonferroni-corrected p<0.05. (B) Volcano plot identifying intragenic repeats enriched for AICDA-dependent DMCs.

2.6 Epigenetic function of AICDA is conserved between human and murine GC B cells

We next investigated if, similar to SHM targets, there were hotspots of AICDA epigenetic activity in GCBs. In order to identify such ‘‘epigenetic hotspots,’’ we looked for differentially methylated regions (DMRs) based on the presence of at least five DMCs, a maximum distance of 250kb between DMCs, and at least 10% difference between average methylation within the region. We identified 119 DMRs between NBs and GCBs from WT mice, distributed throughout all chromosomes and consisting predominantly of hypo-DMRs (**Figure 2.8A**). We observed that these DMRs were mostly AICDA-dependent, and 88 hypo-DMRs and 16 hyper-DMRs were lost in *Aicda*$^{-/-}$ animals (**Figure 2.8A**), in agreement with our results from the analysis of DMCs (**Figure 2.2**).

To investigate whether the AICDA epigenetic program in mouse GC was conserved in human GC, we examined human tonsillar NB and GCB profiles. Using the same criteria applied to mouse data to call DMCs, we confirmed that human GCBs also underwent extensive hypomethylation compared to NBs and displayed greater epigenetic diversity than NBs (**Figure 2.8B and C**). Although we identified a greater number of DMCs in the human NB to GCB transition (69,277 hypo-DMCs and 5,991 hyper-DMCs), the genome-wide distribution of human DMCs was very similar to the distribution of murine DMCs, with enrichment in introns and depletion in promoters, CpG islands, and CpG shelves (**Figure 2.8D and E**). To address whether the methylation changes underlying the GC transition in human cells affected the same AICDA-epigenetic targets as in murine cells, we characterized DMRs between human NBs and GCBs, applying the same criteria used in mouse samples,

and assessed the significance of overlap between orthologs of murine genes associated with AICDA-dependent DMRs and human genes associated with DMRs. Remarkably, we observed a significant overlap between murine AICDA-dependent hypo-DMR-associated gene orthologs and human genes associated with hypo-DMRs (hypergeometric p = 5.73e-08; **Figure 2.8F**). The comparability of epigenetic reprogramming between mouse and human GC suggests that epigenetic changes associated with the NB to GCB transition are conserved between species, similar to conservation of the transcriptional programming.

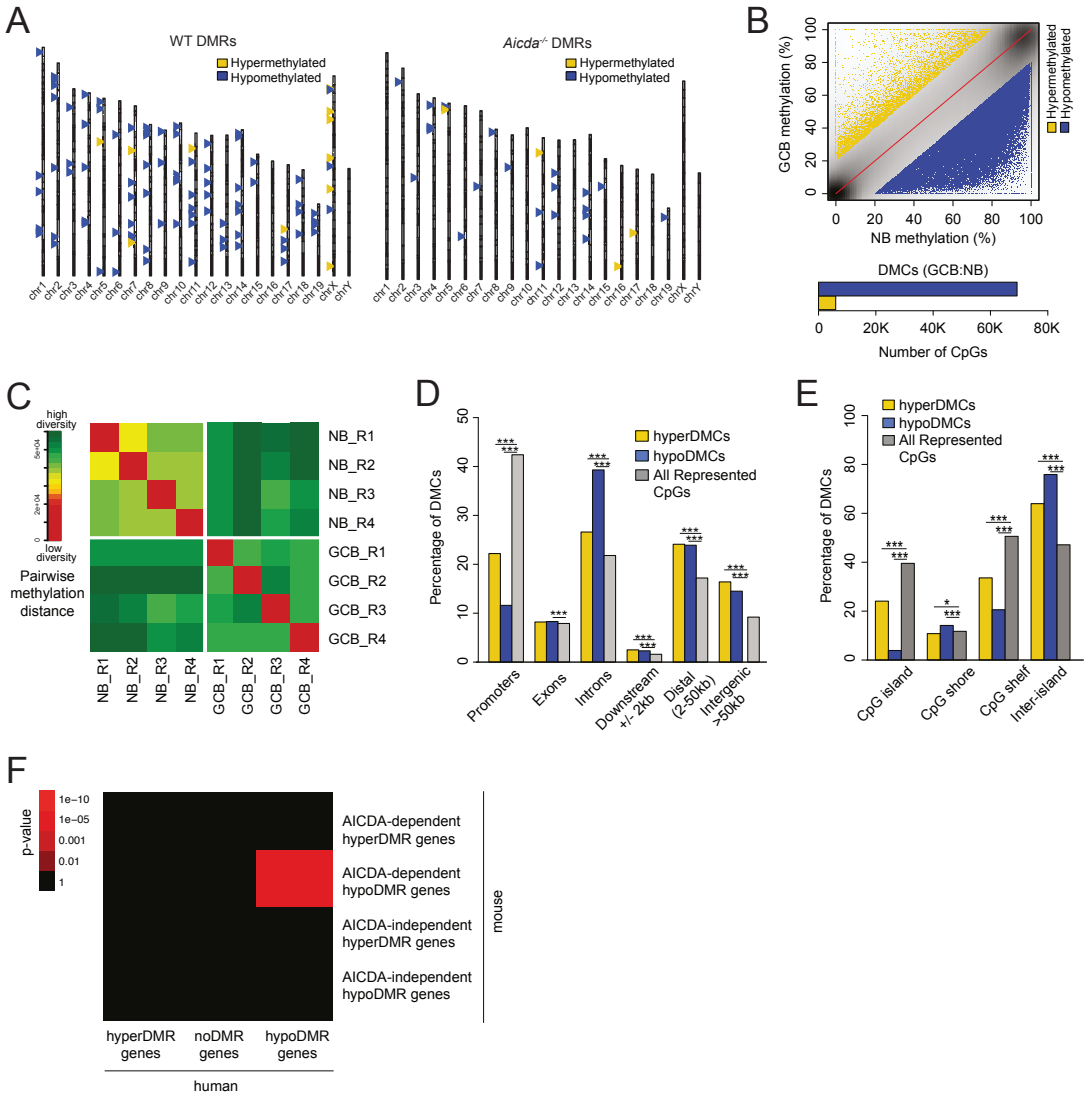

Figure 2.8. Conserved epigenetic function of AICDA between human and murine B cells. (A) Ideogram of DMRs in WT (left) and *Aicda*$^{-/-}$ (right) cells showing depletion of DMRs in *Aicda*$^{-/-}$ cells. Hypo-DMRs are indicated in blue, and hyper-DMRs are indicated in yellow. (B) Combined CpG methylation values (top) and DMCs (bottom) between NBs and GCBs from human tonsils (four replicates) determined by ERRBS using a 20% methylation difference threshold and FDR < 0.001 (Fisher's exact test). Hypomethylated CpGs are indicated in blue, and hypermethylated CpGs are indicated in yellow. (C) Heatmap showing pairwise methylation distance between ERRBS profiles of NB and GCB replicates from human tonsils with GCBs showing greater intra-group methylation distance (green versus yellow). (D) Bar plot showing genomic distribution of hypo-DMCs and hyper-DMCs as well as all CpGs represented within ERRBS experiments. DMCs show depletion in promoters and enrichment in introns and distal and intergenic regions. (E) Bar plot showing distribution of DMCs within CpG islands, shores, shelves, and inter-island regions. (F) Heatmap showing overrepresentation of murine AICDA-dependent hypo-DMR-associated orthologs with human hypo-DMR-associated genes (*p < 0.05, **p < 0.01, and ***p < 0.001).

2.7 Loss of AICDA in GCB reduces DNA methylation heterogeneity and causes relative gain in methylation

As we observed AICDA to be implicated two interrelated aspects of cytosine methylation patterning (cytosine hypomethylation and cytosine methylation diversity), we sought to identify AICDA cytosine methylation heterogeneity signatures in a comprehensive and unbiased manner. To this end, we performed a principal component analysis (PCA) on all CpGs according to these two dimensions of AICDA-associated changes: change in mean methylation level and the change in "spread" of methylation values, measured as interquartile range (IQR), between respective *Aicda*$^{-/-}$ GCB and WT GCB (**Figure 2.9A**). This analysis found that, consistent with the results from **Figure 2.2**, the most pronounced alteration to the GCB *Aicda*$^{-/-}$ methylome involved a relative gain of methylation and decreased heterogeneity between individuals (**Figure 2.9B**). Using the component loading factor from this approach, we identified an AICDA cytosine methylation signature containing 64,323 CpGs (GCB AICDA-perturbed CpGs; **Figure 2.9B**).

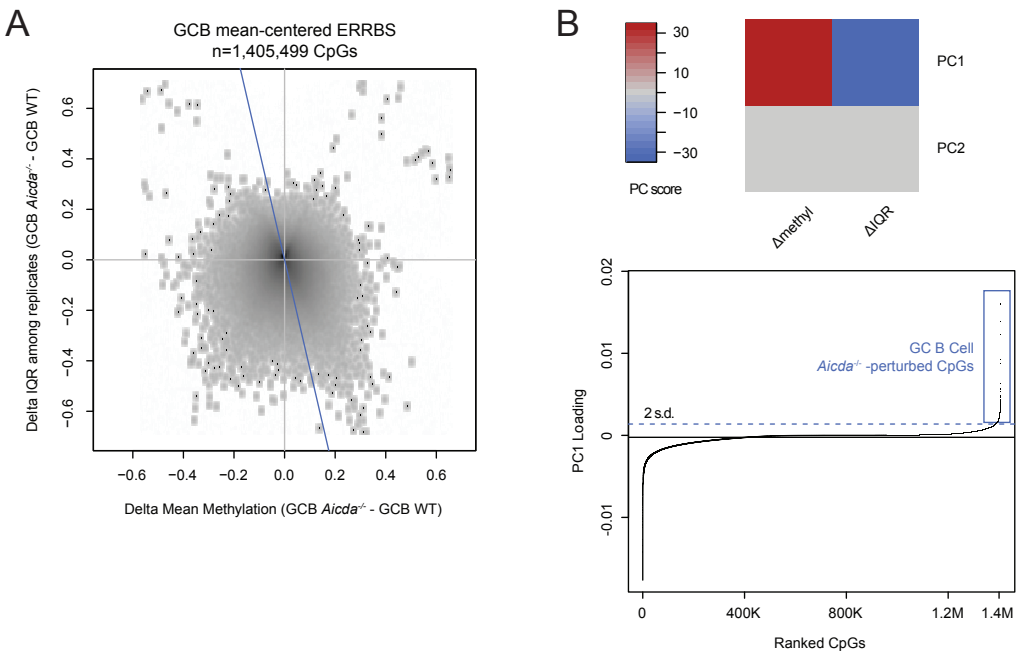

Figure 2.9. Multidimensional integrated analysis of methylation. (A) Density scatterplot showing centered data representing change in mean methylation level versus change in IQR between Aicda$^{-/-}$ and wild-type GC B cells. The blue line indicates the principal component axis. (B) Principal component decomposition of centered data from (A). Top, principal component (PC) scores indicating methylation pattern manifesting increase in methylation and decrease in inter-individual heterogeneity within Aicda$^{-/-}$ GC B cells relative to wild-type controls. Bottom, ranked CpG component loading factors. Horizontal black line indicates mean loading factor and blue rectangle indicates AICDA perturbation CpG signature.

We found the GCB AICDA-perturbed signature to be comprised of CpGs that gained methylation, primarily from a highly methylated state within WT controls (**Figure 2.10A**). The signature CpGs also exhibited a loss of inter-individual heterogeneity among *Aicda*$^{-/-}$ GCBs (median WT GCB IQR = 17.35%; median *Aicda*$^{-/-}$ GCB IQR = 4.89%; **Figure 2.10B**). Notably, the gain of methylation in this signature resulted in a reduction of intra-individual heterogeneity compared to WT (paired Wilcoxon test p<1e-300; **Figure 2.10C**), as many CpGs exhibited a uniformly methylated state in *Aicda*$^{-/-}$ GCBs in contrast to a more mosaic highly methylated state in WT GCBs. Consistent with the

genomic distribution of AICDA-depednt hypoDMCs, we found these signature CpGs to be significantly depleted from gene promoters and highly enriched within introns and intergenic regions (Fisher's exact test p<0.001; **Figure 2.10D**). We also found this signature to be largely depleted from CpG islands, but were slightly enriched with CpG "shelves" (Fisher's exact test p<0.001; **Figure 2.10E**).

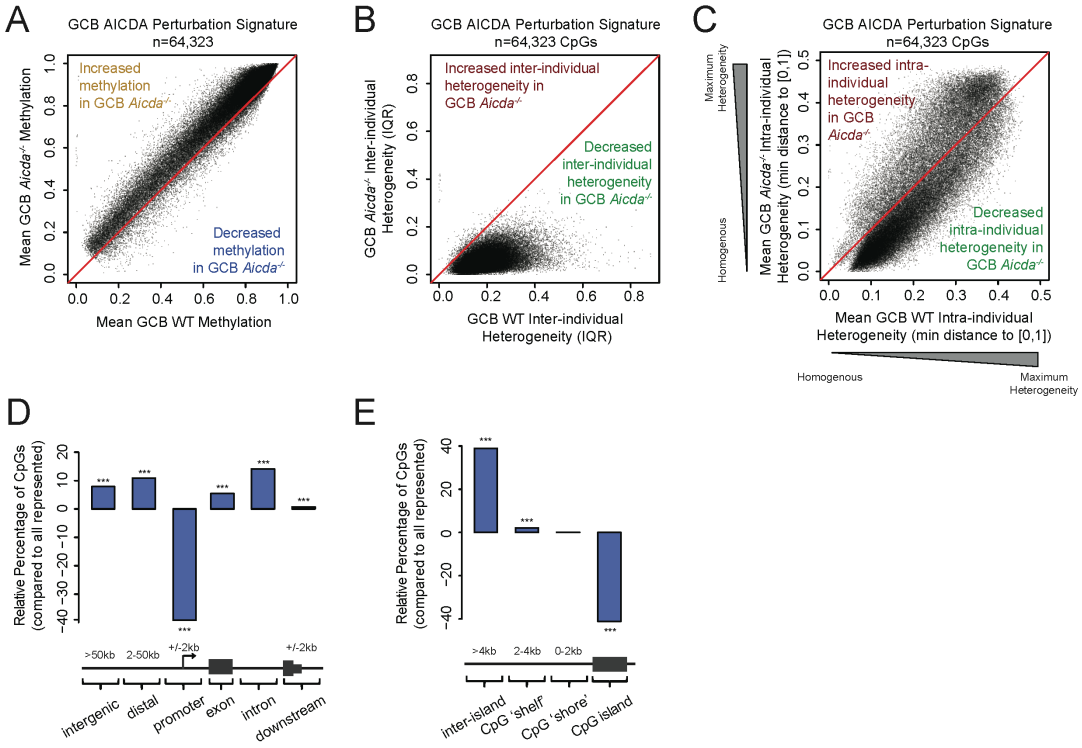

Figure 2.10. Loss of AICDA in GC B cells reduces DNA methylation heterogeneity and causes relative gain in methylation. (A-C) Scatterplots showing shift in mean methylation values (A), inter-individual methylation heterogeneity (B), and intra-individual methylation heterogeneity (C) of GCB *Aicda*$^{-/-}$ perturbation signature between *Aicda*$^{-/-}$ and wild-type GC B cells. (D) Bar plot showing the distribution of GCB *Aicda*$^{-/-}$ perturbed CpGs relative to the distribution of all represented CpGs; Fisher's exact test. (E) Bar plot showing the relative distribution GCB *Aicda*$^{-/-}$ perturbed CpGs within proximity to CpG islands; Fisher's exact test. (* p<0.05, ** p<0.01, *** p<0.001)

2.8 Overexpression of AICDA in GC-derived lymphomas results in more aggressive disease

To characterize the actions of AICDA in GC-derived lymphomagenesis, we transplanted bone marrow cells from VavP-*Bcl2* transgenic mice transfected with AICDA expressing retrovirus (VavP-*Bcl2*+*Aicda*) or empty vector control (VavP-*Bcl2*) into lethally irradiated recipients. All mice were sacrificed after eight months, the timepoint at which VavP-*Bcl2*+*Aicda* animals presented signs of morbidity. Histopathological examination revealed more aggressive disease in VavP-*Bcl2*+*Aicda* (n=8) than in VavP-*Bcl2* (n=7) mice, with greater disruption of splenic architecture and neoplastic B cell expansion in organs such as lung, liver and kidney (**Figure 2.11A**). Reminiscent of human DLBCLs, the spleens of VavP-*Bcl2*+*Aicda* but not VavP-*Bcl2* control mice exhibited white pulp expansion with replacement by sheets of neoplastic B-cells (**Figure 2.11A**). The VavP-*Bcl2*+*Aicda* neoplastic cells were larger, exhibited greater pleomorphic morphology and had higher Ki-67 positivity (**Figure 2.11B**), indicating increased number of proliferating cells, a feature that is correlated with more aggressive DLBCL[243]. Necropsy of diseased animals revealed greater burden of disease in spleen, lung, kidney, and livers of VavP-*Bcl2*+*Aicda* animals, as quantified by the degree of neoplastic expansion and infiltration (**Figure 2.11C**). A separate cohort of mice was followed longitudinally to assess impact of AICDA expression on survival. VavP-*Bcl2*+*Aicda* mice (n=10) manifested significantly shortened lifespan (Log-rank test P=0.0289), with a median survival of 214 days as opposed to 293 days for VavP-*Bcl2* (n=9) animals (**Figure 2.11D**). Hence, AICDA overexpression was associated with a more aggressive disease phenotype and decreased survival.

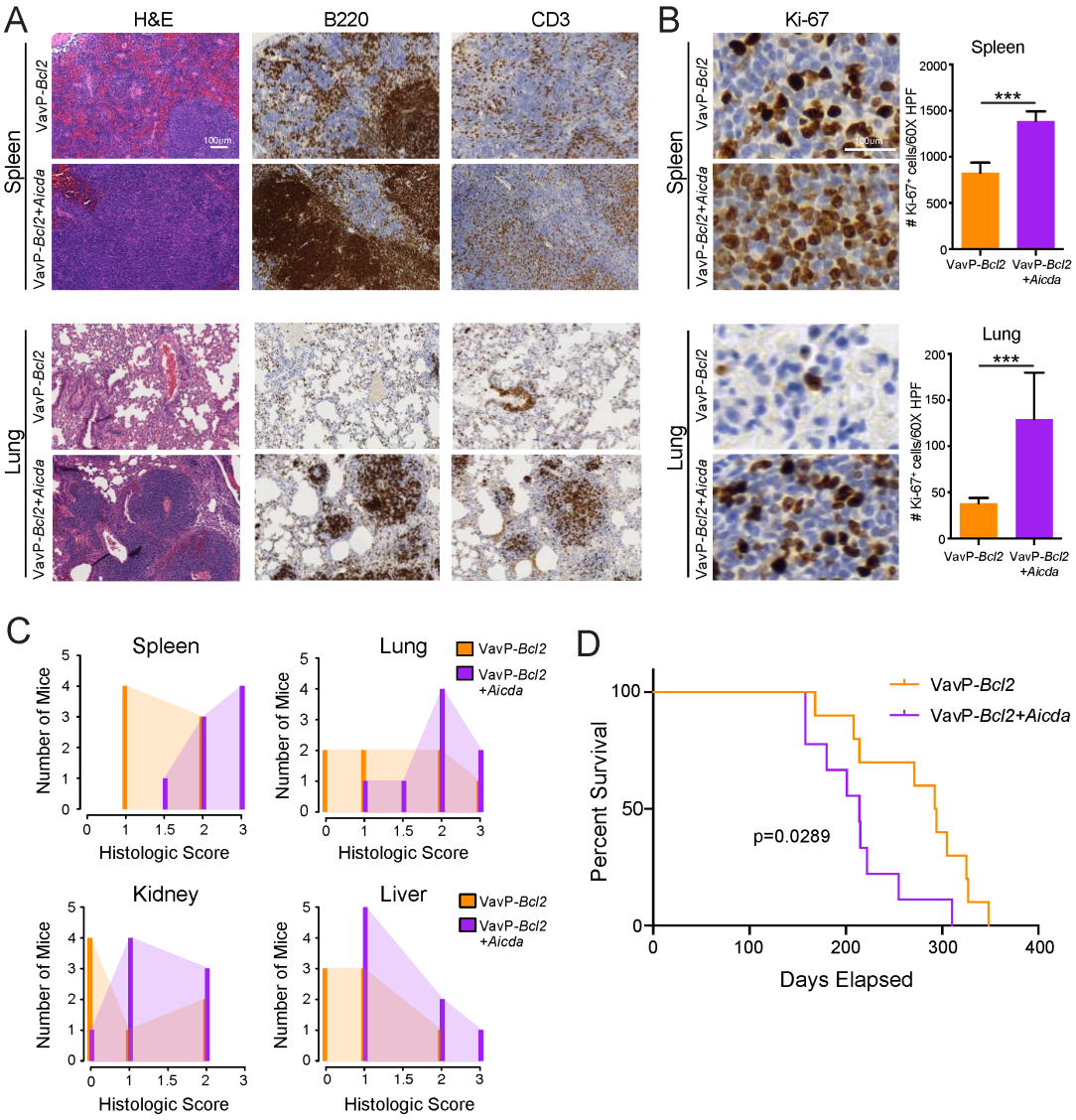

Figure 2.11. AICDA overexpression results in more aggressive BCL2-driven lymphomas. (A) Representative histologic sections of formalin-fixed, paraffin-embedded spleen and lung tissues from VavP-*Bcl2* and VavP-*Bcl2*+ *Aicda* mice. Sections were stained with H&E and antibodies specific for B220 and CD3. (B) Representative histologic sections stained with anti-Ki-67 (left) and quantification of Ki67+ cells (right) in the spleen and lung tissues from VavP-*Bcl2* and VavP-*Bcl2*+*Aicda* mice. Bars represent mean of number of Ki67+ cells in 10 fields of spleen and lung sections and error bars indicate standard deviation; two-tailed t-test ***p < 0.001. (C) Histologic score, measuring relative organ infiltration,of spleen, lung, kidney and liver from VavP-*Bcl2* (n=7) and VavP-*Bcl2*+*Aicda* (n=8) mice. Scores correspond to no (0), mild (1), moderate (2) and marked (3) infiltration by neoplastic lymphocytes. (D) Kaplan–Meier survival curve of VavP-*Bcl2* (n=10) and VavP-*Bcl2*+*Aicda* (n=9) mice. Significant differences in survival were evaluated by log-rank (Mantel–Cox) test. Histology and survival data prepared by M. Dominguez and analyzed by R. Shaknovich.

2.9 AICDA overexpression induces DNA methylation heterogeneity and hypomethylation, but no increase in somatic mutation burden

To explore whether AICDA-overexpressing lymphomas manifested cytosine methylation pattern differences, we performed ERRBS on VavP-*Bcl2*+*Aicda* (n=7) and VavP-*Bcl2* (n=6) lymphoma B cell populations. Initial examination of global methylation levels indicated that VavP-*Bcl2*+*Aicda* tumors showed fewer highly methylated CpGs, consistent with the role of AICDA in demethylation (Fisher's exact test p=2.27e-14; **Figure 2.12A**). To interrogate whether VavP-*Bcl2*+*Aicda* tumors also exhibited more diverse cytosine methylation patterning, we examined pairwise methylation distance between cytosine methylation profiles and found that VavP-*Bcl2*+*Aicda* lymphomas indeed manifested greater global inter-tumor heterogeneity than their VavP-*Bcl2* counterparts (Wilcoxon p=0.00112; **Figure 2.12B**). Since AICDA induces DNA hypomethylation in GC B-cells, we postulated that the AICDA-mediated DNA methylation heterogeneity would be more prominently observed among hypomethylated cytosines in VavP-*Bcl2*+*Aicda* lymphoma cells. We therefore evaluated two-dimensional AICDA-associated changes (mean methylation level and IQR) between respective tumors, at each CpG showing >20% methylation difference. Among these selected CpGs, inter-tumor heterogeneity was significantly associated with hypomethylation in VavP-*Bcl2*+*Aicda* mice compared to control mice (Fisher exact test p=3.68e-40; **Figure 2.12C**).

Using PCA to identify an AICDA cytosine methylation signature, we found that, inverse and reciprocal to the results from AICDA depletion, the most pronounced alteration caused by AICDA overexpression within the VavP-*Bcl2*

methylome involved methylation loss and increased inter-tumor heterogeneity. We identified a VavP-*Bcl2*+*Aicda* cytosine methylation signature containing 49,750 CpGs (VavP-*Bcl2* AICDA-perturbed CpGs). This signature was comprised of CpGs that lost methylation, primarily from a highly methylated state within VavP-*Bcl2* controls (**Figure 2.12D**). The signature CpGs also exhibited increased inter-tumor heterogeneity among the VavP-*Bcl2*+*Aicda* lymphomas, compared to relatively stable methylation states within VavP-*Bcl2* controls (median VavP-*Bcl2* IQR = 5.56%; median VavP-*Bcl2*+*Aicda* IQR = 19.2%; **Figure 2.12E**). The loss of methylation of this VavP-*Bcl2*+*Aicda* cytosine methylation signature also resulted in a relative increase in intra-tumor heterogeneity compared to the VavP-*Bcl2* tumors (paired Wilcoxon test p<1e-300; **Figure 2.12F**). VavP-*Bcl2*+*Aicda*-perturbed CpGs were significantly depleted from gene promoters, but were highly enriched within introns and intergenic regions (Fisher's Exact Test p<0.001; **Figure 2.12G**). VavP-*Bcl2*+*Aicda*-perturbed CpGs were also largely depleted within CpG islands, but were slightly enriched within CpG "shores" and "shelves" (Fisher's Exact Test p<0.001; **Figure 2.12H**). These AICDA-mediated epigenetic changes are unlikely to stem from differences in the proliferation rate between AICDA-overexpressing and control lymphomas since DNA methylation heterogeneity in lymphoma cells is independent of the mitotic ratio[189]. Given the similar genomic distribution and reciprocal nature of epigenetic effects between *Aicda* loss and overexpression, respectively, in normal and malignant B-cells, we investigated whether AICDA preferentially affected the same genes within the two systems. Indeed, there was statistically significant overlap between genes enriched for VavP-*Bcl2* tumor AICDA signature CpGs

within gene bodies and genes over-representing GC *Aicda*$^{-/-}$ signature CpGs within gene bodies (hypergeometric test p=1.98e-121, **Figure 2.12l**).

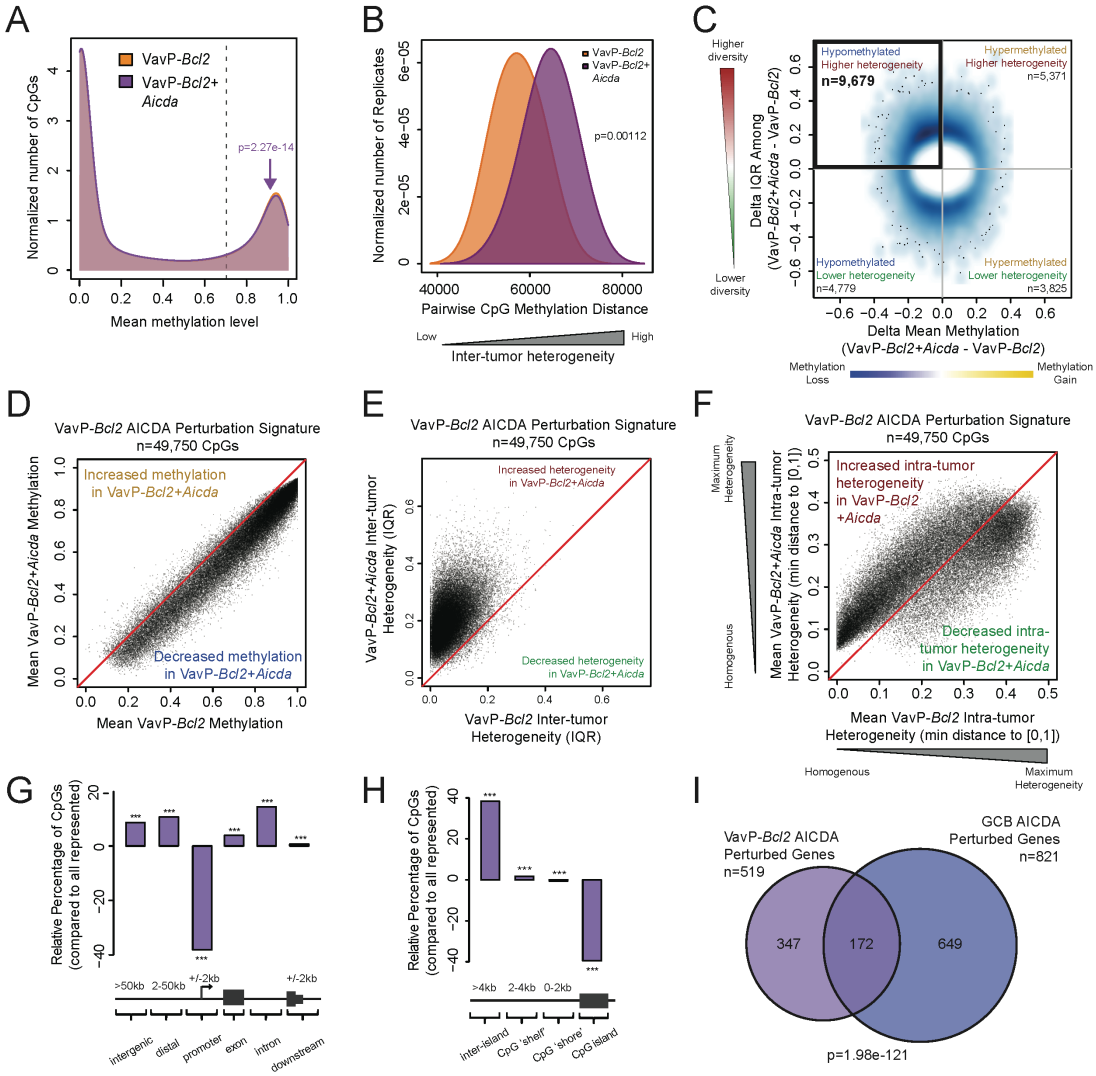

Figure 2.12. AICDA overexpression induces DNA methylation heterogeneity and hypomethylation in VavP-*Bcl2+Aicda* tumors. (A) Density plot showing the mean methylation distribution of all CpGs respresented in the VavP-*Bcl2* (n=6) and VavP-*Bcl2+Aicda* (n=7) ERRBS profiles. Overexpression of AICDA is associated with fewer highly methylated CpGs, Fisher's Exact Test. (B) Density plot showing the inter-tumor pairwise methylation distance between ERRBS profiles of VavP-*Bcl2* and VavP-*Bcl2+Aicda* tumors. VavP-*Bcl2+Aicda* tumors have greater pairwise distance, indicating increased inter-tumor heterogeneity among methylation profiles; two-sided Wilcoxon signed-rank test (C) Density scatterplot showing the relationship between methylation change (x-axis) and change in inter-tumor diversity (y-axis) for all CpGs manifesting >20% combined methylation level difference and/or IQR difference. (D-F) Scatterplots showing shift in mean methylation values (D), inter-tumor heterogeneity (E), and intra-tumor heterogeneity (F) of AICDA perturbation signature between VavP-*Bcl2* and VavP-*Bcl2+Aicda*. (G) Bar plot showing the distribution of AICDA-perturbed CpGs relative to the distribution of all represented CpGs; Fisher's exact test. (H) Bar plot showing the relative distribution of AICDA-perturbed CpGs within proximity to CpG islands; Fisher's exact test. (I) Venn diagram showing the overlap between genes significantly over-representing GCB Aicda$^{-/-}$ signature CpGs and genes over-representing VavP-Bcl2+Aicda signature CpGs; hypergeometric test. (* p<0.05, ** p<0.01, *** p<0.001)

2.10 AICDA overexpression does not increase somatic mutation burden

Overexpression of AICDA could also contribute to lymphomagenesis by introducing somatic mutations at non-Ig loci, such as proto-oncogenes. To evaluate the contribution of AICDA mutational activity to the aggressive phenotype observed in VavP-*Bcl2*+*Aicda* mice, we first assessd the mutational burden of known Ig (J_H4 and *S*μ) and non-Ig loci (*Bcl6*, *Cd83* and *Pax5*) using targeted resequencing. We observed no significant differences in burden of non-synonymous mutations (**Figure 2.13A**) or indels (**Figure 2.13B**) within Ig loci, *Cd83,* or *Pax5*, although indels were enriched at the *Bcl6* locus in VavP-*Bcl2*+*Aicda* mice. To further explore genomic effects of AICDA overexpression, we performed whole exome sequencing in VavP-*Bcl2*+*Aicda* (n=4) and VavP-*Bcl2* lymphomas (n=4) and found no significant difference in non-synonymous or indel mutation burdens (**Figure 2.13C and D**). These data suggest that the more aggressive VavP-*Bcl2*+*Aicda* phenotype is not significantly associated with increased mutational burden. Hence, overexpression of AICDA in B-cell lymphomas induces increased burden of epigenetic changes, without significant change in the burden of somatic mutation compared to lymphomas with endogenous AICDA levels, recapitulating earlier evidence that the level of AICDA expression did not increase aberrant SHM in DLBCL[244].

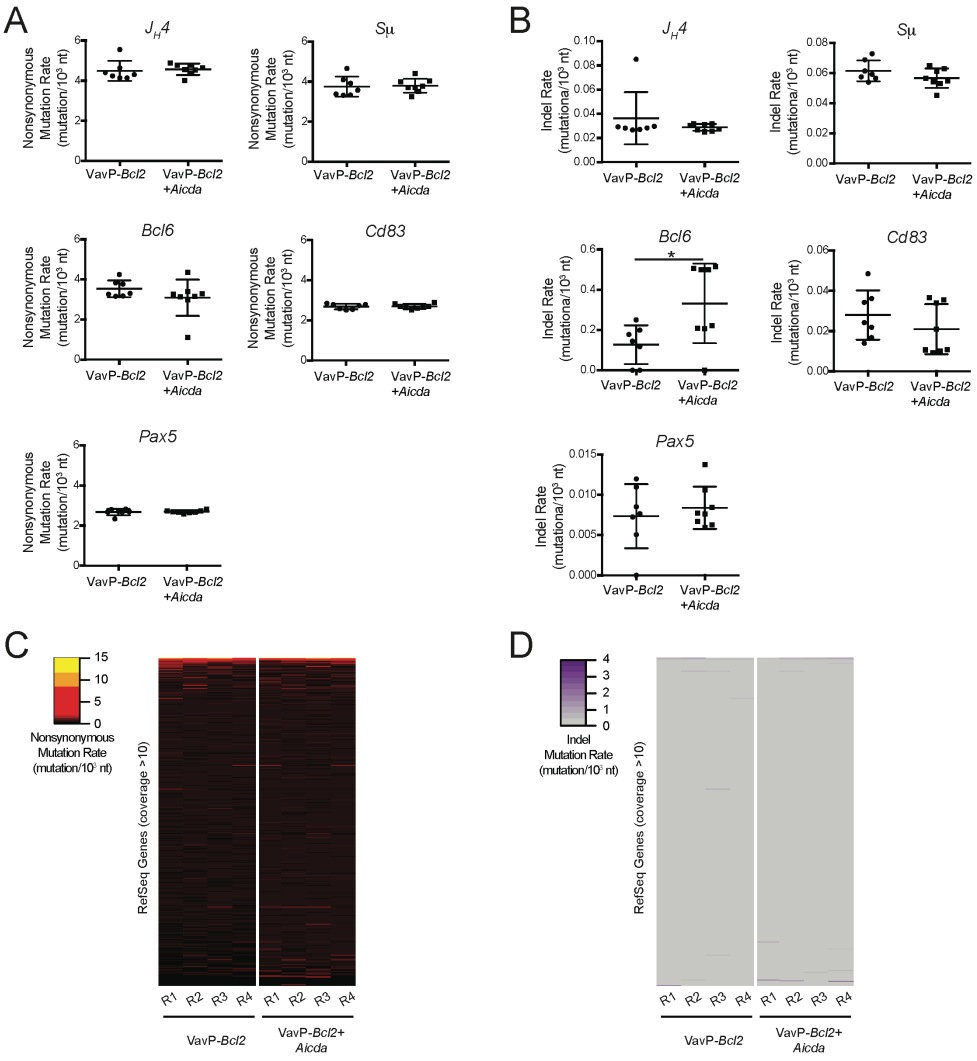

Figure 2.13. AICDA overexpression does not increase somatic mutation burden in VavP-*Bcl2* tumors. (A) Analysis of the nonsynonymous mutation rate (mutation/10^3 nt) and (B) indel mutation rate (mutation/10^3 nt) in AICDA target genes (JH4, Sμ, Bcl6, Cd83, and Pax5) in B220+ cells isolated from the spleens of VavP-*Bcl2* (n=7) and VavP-*Bcl2*+*Aicda* (n=8) mice. Dot plots represent mean of mutation rates and error bars indicate standard deviation. (C) Heatmap showing the nonsynonymous mutation rates and (D) indel mutation rates in VavP-*Bcl2* and VavP-*Bcl2*+*Aicda* lymphomas. No significant differences in mutation burden are observed between the two conditions; two-sided Wilcoxon signed-rank test.

2.11 High expression of AICDA in DLBCL is associated with higher DNA methylation heterogeneity and hypomethylation

To investigate whether AICDA cytosine methylation signatures can be detected in primary human lymphomas, we performed ERRBS and RNAseq in a cohort of 63 primary DLBCL patients. We compared cases with the highest AICDA expression (>50 CPM; AICDA-high DLBCL; n=10) against the cases with lowest or no AICDA expression (<2 CPM; AICDA-low DLBCL; n=19, **Figure 2.14A**). Similar to mouse VavP-*Bcl2*+*Aicda* lymphomas, we found that AICDA-high DLBCL patients manifested greater global inter-tumor heterogeneity than AICDA-low cases (Wilcoxon p=4.17e-06; **Figure 2.14B**). Given the concordance of these results with our previous murine findings, we performed a PCA to determine if higher levels of AICDA would have similar consequences upon the DLBCL methylome. Indeed, AICDA-high DLBCLs exhibited a pattern of increased inter-tumor heterogeneity and hypomethylation compared to DLBCL profiles with low expression of AICDA. From this PCA, we identified a DLBCL AICDA-high methylation signature (n=37,557 DLBCL AICDA-perturbed CpG) that exhibited similar methylation dynamics to VavP-*Bcl2*+*Aicda* (and opposite to *Aicda*$^{-/-}$ GC B-cells), including loss of CpG methylation from intermediate and highly methylated states (**Figure 2.14C**) alongside gain of inter-tumor methylation heterogeneity (**Figure 2.14D**). We also found that the AICDA-high DLBCLs had higher intra-tumor methylation heterogeneity compared to their lower AICDA-expressing controls (paired Wilcoxon p=5.22e-229; **Figure 2.14E**). This CpG signature was depleted in gene promoters and enriched in exons, introns, and distal regions (Fisher's exact test p<0.001; **Figure 2.14F**). AICDA-high DLBCL perturbed CpGs were also under-represented in CpG islands and enriched in

CpG “shelves” and “shores” (Fisher’s exact test p<0.001; **Figure 2.14G**). Notably, we found significant overlap of gene orthologs affected by AICDA-perturbed CpGs in human AICDA high DLBCLs and murine VavP-*Bcl2*+AICDA lymphomas (hypergeometric p=2.21e-23; **Figure 2.14H**) and with the gene orthologs affected by AICDA in normal GC B-cells (hypergeometric p=8.48e-33; **Figure 2.14H**). To assess whether this high AICDA-perturbed cytosine methylation signature was associated with DLBCL subtype, we categorized our AICDA-high and AICDA-low DLBCL cases according to their gene expression[67]. Despite presumably reflecting late or post-GC B-cells, we found that AICDA-high DLBCL were more likely to be classified as ABC-subtype than AICDA-low DLBCL (Fisher’s exact test p=0.021; **Figure 2.14I**). This is consistent with previous reports that AICDA is more highly expressed in ABC-DLBCL[244,245] and that epigenetic heterogeneity is more pronounced in the ABC-subset of DLBCL[189].

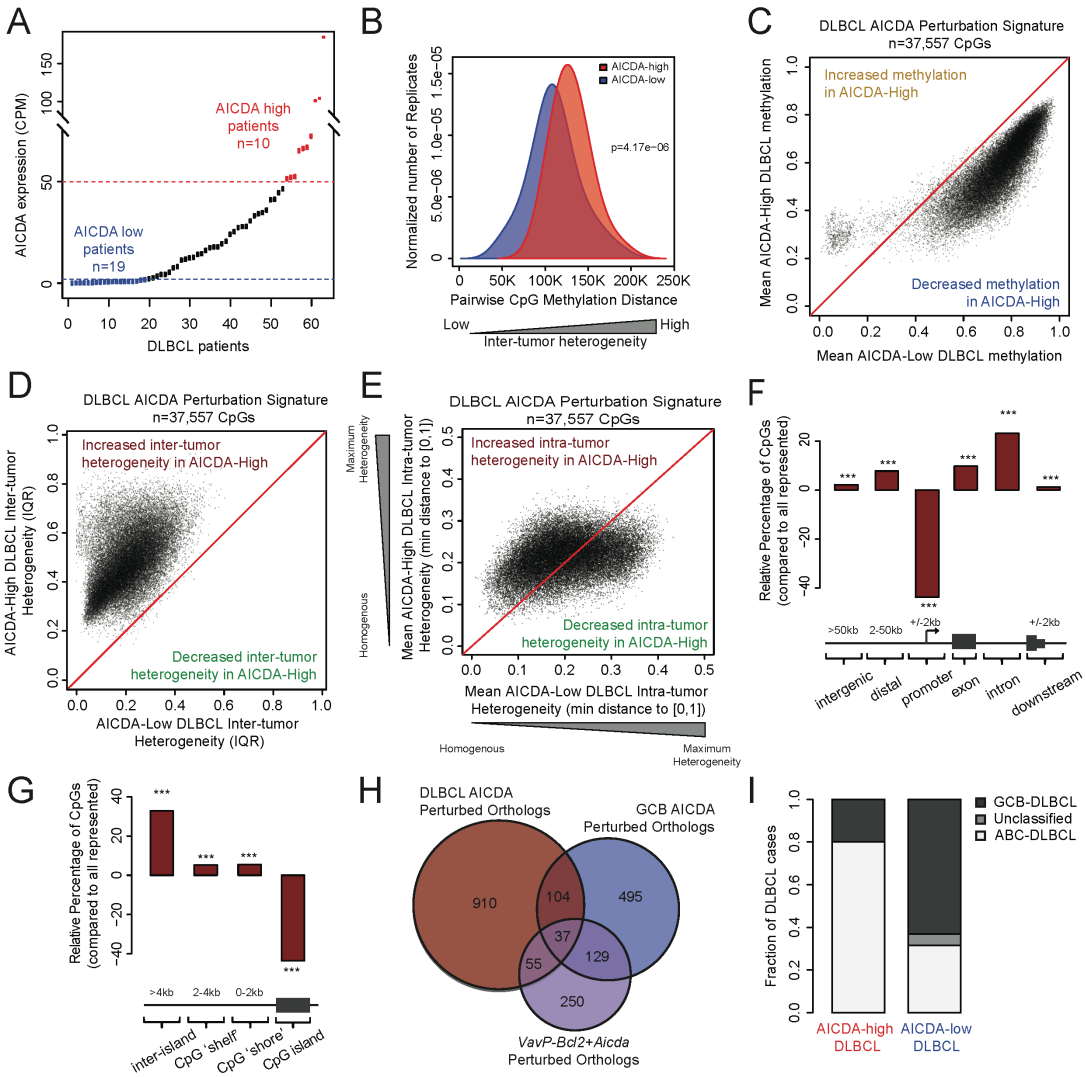

Figure 2.14. High expression of AICDA in DLBCL is associated with higher DNA methylation heterogeneity and hypomethylation. (A) Expression of AICDA within DLBCL cohort. Cases with AICDA expression <2 CPM were characterized as low expressors (AICDA-low) and cases with expression >50 CPM were characterized as high expressors (AICDA-high). (Density plot showing the inter-tumor pairwise methylation distance between ERRBS profiles of AICDA-low and AICDA-high DLBCLs. AICDA-high DLBCLs have greater pairwise distance, indicating increased inter-tumor heterogeneity; two-sided Wilcoxon signed-rank test. (B-D) Scatterplots showing shift in mean methylation (B), inter-tumor diversity (C), and intra-tumor heterogeneity (D) of DLBCL AICDA perturbation signature between AICDA-low and AICDA-high DLBCL. (E) Bar plot showing the distribution of DLBCL AICDA-perturbed CpGs relative to the distribution of all represented CpGs; Fisher's exact test. (F) Bar plot showing the relative distribution of DLBCL AICDA-perturbed CpGs within proximity to CpG islands; Fisher's exact test. (g) Venn diagram showing the overlap between genes significantly over-representing DLBCL AICDA signature CpGs and murine orthologs over-representing VavP-*Bcl2+Aicda* and GCB AICDA signature CpGs . (H) Bar plot showing the relative fraction of DLBCL subtypes within AICDA-high and AICDA-low DLBCL cases. (*p<0.05, **p<0.01, ***p<0.001)

2.12 Notes on analysis

Although quantification of inter-tumor heterogeneity at individual CpGs via IQR or standard deviation is relatively straightforward, quantification of inter-tumor heterogeneity between global ERRBS profiles is substantially more difficult. Many ERRBS profiles are disproportionately weighted with uniformly methylated or unmethylated CpGs. This makes comparison of CpG IQR distributions between groups problematic as low heterogeneity CpGs overwhelm the distribution and may lead to false negatives (**Figure 2.15A**). This problem can be overcome using statistical procedures that reduce dimensionality, *e.g.* principal component analysis, as low variance CpGs will not contribute meaningfully to the first few principal components (PCs). Subsequent comparison of samples using Euclidean distance within the space defined by PC vectors with greatest variance permits evaluation of the inter-tumor heterogeneity of ERRBS profiles (**Figure 2.15B**). Such PC space distance approaches, however, still have a number of limitations. The interpretation of distance metrics within low-dimensional orthogonally transformed space can be difficult. Also, given their inability to tolerate missing data, only CpGs represented in all ERRBS profiles can be evaluated, potentially leading to omission of relevant data. Additionally, this approach is fragile to outliers within the data. Gross errors in methylation values for even a few CpGs can have a disproportionate weight on variance within the data and designation of the first PC (**Figure 2.15C**). As each successive PC vector is independent and orthogonal to preceding PC vectors, such outliers will drive the PC space and may lead to artificial increase or decrease within heterogeneity measurements. Our “pairwise methylation distance” approach addresses these issues by evaluating Manhattan distance on the full-

dimensionality of the data. The contribution of low heterogeneity or uniform CpGs is minimized, as they will not contribution greatly to the overall distance. Furthermore, given the rectilinear properties of Manhattan distance, interpretation is intuitive and the resulting distance value can be construed as the number of methylation "events" that separate the respective sample pair. Also, we can evaluate the similarity of methylation values at all CpGs common just to the two ERRBS profiles under comparison, regardless of whether they are represented in every other profile. By normalizing the resulting Manhattan distance to 10^6 CpGs, we are able to evaluate the distribution of pairwise methylation distances between relevant groups (*e.g.* high AICDA expression versus low AICDA expression) to assess differences in inter-tumor heterogeneity (**Figure 2.12B, 2.14B**).

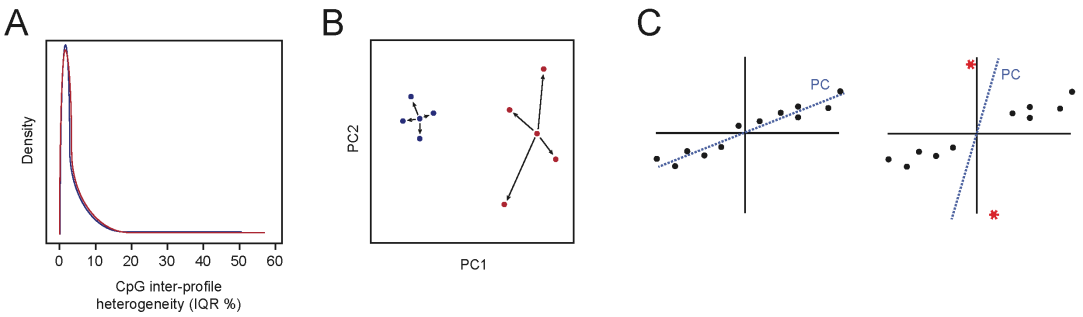

Figure 2.15. Notes on Analysis. (A) ERRBS profiles typically over-represent CpGs with uniform methylation values or low inter-profile heterogeneity. This can obscure differences in groups with higher inter-profile heterogeneity (red) from groups with lower inter-profile heterogeneity (blue). (B) Diversity among groups of ERRBS profiles can be measured using euclidean distance within space defined by first two principal components (PCs). In this example, the blue group exhibits shorter distance between points and less heterogeneity than the red group. (C) PC analysis is fragile to outliers and gross errors of even a few points (as shown on left) can result in deviation of the PC vector.

Many contemporary analytical approaches exist that assess differential cytosine methylation changes levels or changes in cytosine methylation heterogeneity, either inter-sample (*e.g.* variance or IQR) or intra-sample (typically via entropy-based methods on four-CpG segments on bisulfite sequencing reads). These approaches, however, treat changes in methylation levels as separate and independent from changes in methylation heterogeneity. This assumption is erroneous, as methylation heterogeneity is generated by non-uniform gain or loss of methylation at CpGs. Recognizing the interconnected nature of cytosine hypomethylation and AICDA-mediated increase in inter-sample heterogeneity, we devised an approach that accounts for both changes in mean methylation level and change in inter-tumor heterogeneity. Using an unconventional application of principal component analysis, we were able to determine the linear combination of alterations in methylation level and “spread” associated with the overexpression of AICDA.

Additionally, by using the decomposed PC loadings as a measure of the weight each CpG contributes to the overall pattern of hypomethylation and inter-tumor methylation heterogeneity gain, we were able to identify a signature of CpGs that best drive the cytosine methylation alterations we observe from AICDA overexpression.

2.13 Discussion

We have demonstrated using a genome-wide approach that B cell transit through the GC is accompanied by locus-specific hypomethylation and minor gains of methylation, along with a substantial increase in DNA methylation diversity. More importantly, our results indicate that such changes are largely mediated by AICDA. In the last decade, we have gained in-depth knowledge regarding the function of epigenetic alterations in normal development and cancer biology. DNA methylation is understood to play a key role in gene imprinting, X chromosome inactivation, and regulation of gene expression specific to tissue identity, developmental stage, and cell lineage[246,247]. Changes in the DNA methylome mark specific stages of B cell ontogeny and play an important role in B cell lymphomagenesis[185,189,231,248-251]. The GC stage of B cell development is associated with a proliferative burst, affinity maturation of B cells with associated SHM and CSR, all of which contribute to adaptive immune response and determine antibody diversity[68], but the contribution of the methylome to these processes is not clearly defined. It has been demonstrated that changes in methylation are required for the successful formation of the GC and that such modifications are dependent on DNMT1, a methyltransferase highly expressed in GCBs[179]. While the mechanism of DNA

methylation gain is well understood, the mechanism of demethylation, the factors responsible for the loss of methylation in GCBs, and its biological significance remain almost completely unknown. Since AICDA is highly expressed in GCBs and has been implicated in DNA demethylation during embryonic development and epigenetic reprogramming[190-193], we hypothesized that AICDA is involved in the active demethylation of B cells during GC transition.

To prove this hypothesis, we isolated NBs and GCBs from *in vivo* WT and *Aicda*$^{-/-}$ mice and profiled their methylome using ERRBS, a genome-wide approach capable of interrogating three million CpGs. We observed that over 90% of methylome alterations characterizing the transition from NBs to GCBs were lost in *Aicda*$^{-/-}$ animals, confirming the role of AICDA in the DNA demethylation of the GCB genome. We also found that AICDA depletion caused loss of hypermethylation in GCBs. We suspect this to be a result of reduced recruitment of DNMT1 to double-strand breaks[252], putatively generated as a consequence of the AICDA deamination activity[253]. Several prior attempts to link AICDA to demethylation in GCBs were made before. Fritz *et al.* addressed this same question using an *ex vivo* system, activating primary splenocytes in the presence of anti-CD40, LPS, and IL-4[232]. The authors could not detect AICDA-induced changes in the B cell methylome, consistent with our results with *ex vivo* stimulated B cells. This suggests that AICDA-dependent demethylation requires additional factors and is perhaps coupled to the rate of SHM, which is much lower in the *ex vivo* system than in GCBs[233,234]. In this regard, it has recently been demonstrated that *ex vivo* stimulated B cells are defective in SHM because the initiating form of RNA

polymerase II is not retained in the variable regions of the Ig genes, hampering the recruitment of the cofactor Spt5 and AICDA[254]. Another attempt to delineate the demethylation function of AICDA in GCBs was made by Hogenbirk *et al.* using MethylCap sequencing (MethylCap-seq) and failed to find any AICDA-dependent changes[255]. MethylCap-seq is an affinity-purification-based technique, which is likely not to be sufficiently sensitive to detect variable methylation changes in CpGs scattered throughout genome. Here, we have used ERRBS, a genome-wide technique with higher coverage compared to MethylCap-seq and single-nucleotide-level resolution[256]. We think that the above differences are due to the experimental system and the techniques used in earlier studies.

Importantly, we have demonstrated that the epigenetic diversification of the B cell methylome during GC transition is dependent on AICDA activity. It is tempting to speculate that this methylation diversification may contribute, along with SHM, to clonal evolution among normal GCBs. We show here that the genomic distribution of hypo-DMRs in GCBs is similar to the distribution of AICDA binding sites revealed by Liu *et al*[78]. We also provide circumstantial evidence that the demethylase function of AICDA may arise from its deaminase activity, showing that AICDA-dependent hypo-DMCs are enriched within known AICDA target genes for SHM. Despite expectation that AICDA-dependent differential methylation would be concentrated around transcription start sites (TSSs) of genes, similar to SHM hotspots, our data reveal that DMCs are enriched in gene introns and intergenic regions. This is consistent with the location of AICDA-dependent demethylation observed in other systems[192,193]. We also found enrichment of AICDA-dependent hypo-DMCs at

loci associated with dsDNA breaks. It is possible that, despite the intense focus on SHM target genes, AICDA may bind genome-wide, with the majority of binding similarly distributed outside of TSS and gene bodies. This would suggest AICDA deamination activity to have more far-reaching consequences than we have yet appreciated. It has been proposed that AICDA-dependent regions of demethylation may extend beyond the deamination sites as a result of the activity of processive DNA repair pathways (mismatch repair or long-patch base excision repair). These pathways can result in the replacement of long stretches of DNA (up to 2 kb) with concomitant possible repair of all somatic mutations[257]. Such broad extension of hypomethylation could, in turn, have various consequences, including instability of transposable elements, chromosomal translocations, and gene deregulation, as suggested by Guadet *et al*[258].

Although AICDA loss abrogates the majority of the methylation changes experienced by GCB, we observe residual hypomethylation in *Aicda*$^{-/-}$ GCBs, suggesting that other demethylation mechanisms are likely to exist in these cells. The most plausible scenario is TET-dependent oxidative demethylation, which has been proposed as an alternative to AICDA deamination-dependent demethylation[259]. Another source of demethylation may be passive loss of methylation in highly proliferative GCBs. The fact that *Aicda*$^{-/-}$ GCBs are highly proliferative, and nevertheless have minimal loss of methylation, argues against this theory. Moreover, passive stochastic loss of methylation would likely be randomly distributed throughout the genome, and our results show preferential genomic distribution of the methylation changes.

Our work additionally suggests that AICDA is a critical source of epigenetic plasticity and heterogeneity in DLBCL. We find that AICDA-linked epigenetic heterogeneity is predominantly associated with relative loss of cytosine methylation, consistent with the known mechanism of action of AICDA in cytosine deamination. Although AICDA targeting seems to show preference for certain chromatin features, we anticipate that the effect on specific cytosines will likely be stochastic, resulting in gradual divergence of methylation landscapes between cells. Consistent with the known association of cytosine methylation heterogeneity with inferior clinical outcome in human DLBCL patients[189,260,261], we found that AICDA overexpression in mice was associated with increased inter-tumor methylation heterogeneity (**Figure 2.12**) and was linked to more aggressive disease in murine B-cell tumors (**Figure 2.11**). We suggest that AICDA-induced epigenetic heterogeneity increases plasticity, thereby allowing cancer cells a greater degree of population diversity and enhancing the adaptive capacity of the overall tumor. Additionally, we found further evidence that ABC-DLBCLs, which are more aggressive than their GCB-DLBCL counterparts with inferior clinical outcome[73], are associated with high AICDA expression along with greater epigenetic heterogeneity. This suggests that AICDA may be among the contributing factors to epigenetic heterogeneity in ABC-DLBCL.

Previously, the role of AICDA in lymphomagenesis has been linked to its mutagenic potential. Our data suggest additional dimensions to the deleterious effect of this protein in DLBCL, via enhanced epigenetic plasticity. The epigenetic role of AICDA in neoplastic transformation and cancer progression may also reach beyond GC-derived lymphomas. AICDA can be

transcriptionally upregulated in epithelial cells via NFκB signaling following cytokine stimulation or exposure to pathogenic factors, such as *Helicobacter Pylori*[19]. AICDA has also been implicated in non-lymphoid cancers, including melanoma and pancreatic carcinomas[262,263]. Such association between inflammation, infection, and expression of AICDA may prove to be a missing link between chronic inflammation and increased risk of various malignancies.

Aberrant cytosine methylation patterning, along with other epigenetic modifications, plays a critical role in the pathogenesis and progression of lymphoid malignancies[264] and cancer epigenomes have been long appreciated to differ from their normal tissue counterparts[265,266]. As opposed to the genomic DNA sequence, cytosine methylation distribution has great plasticity and can be dynamically redistributed in response to environmental changes or the influence of transcription regulatory factors[97]. One consequence of such plasticity is the epigenetic divergence of individual tumor cells from each other (**Illustration 2.1**). Stochastic intra-tumor epigenetic heterogeneity is documented to occur in lymphoid and myeloid malignancies and is associated with inferior clinical outcome[189,260,261,267,268]. Intra-tumor heterogeneity is proposed to increase population fitness in tumors by enabling individual cells to derive new epigenetic subclones, providing the tumor with greater capacity to adapt and resist therapeutic challenges [269-271]. Intra-tumor heterogeneity has also been linked to decoupling of the relationship between promoter methylation and gene expression, leading to a more variable transcriptional landscape[267]. The generation of transcriptionally distinct epigenetic subclones, potentially containing regions of variable genomic stability, may be detrimental and explain both the prevalence and poor clinical

associations of cytosine methylation heterogeneity within lymphoid malignancies. In support of this hypothesis, fitness simulations have shown that, even in the absence of change in the mean phenotype, propensity toward phenotypic variability itself can substantially increase population fitness within a changing environment[272]. All mechanisms will need more formal examination before the final model of AICDA-mediated epigenetic regulation is articulared, but our work indicates for the first time a clear role epigenetic role for AICDA in B cell maturation during GC transition and as a driver for cytosine methylation heterogeneity.

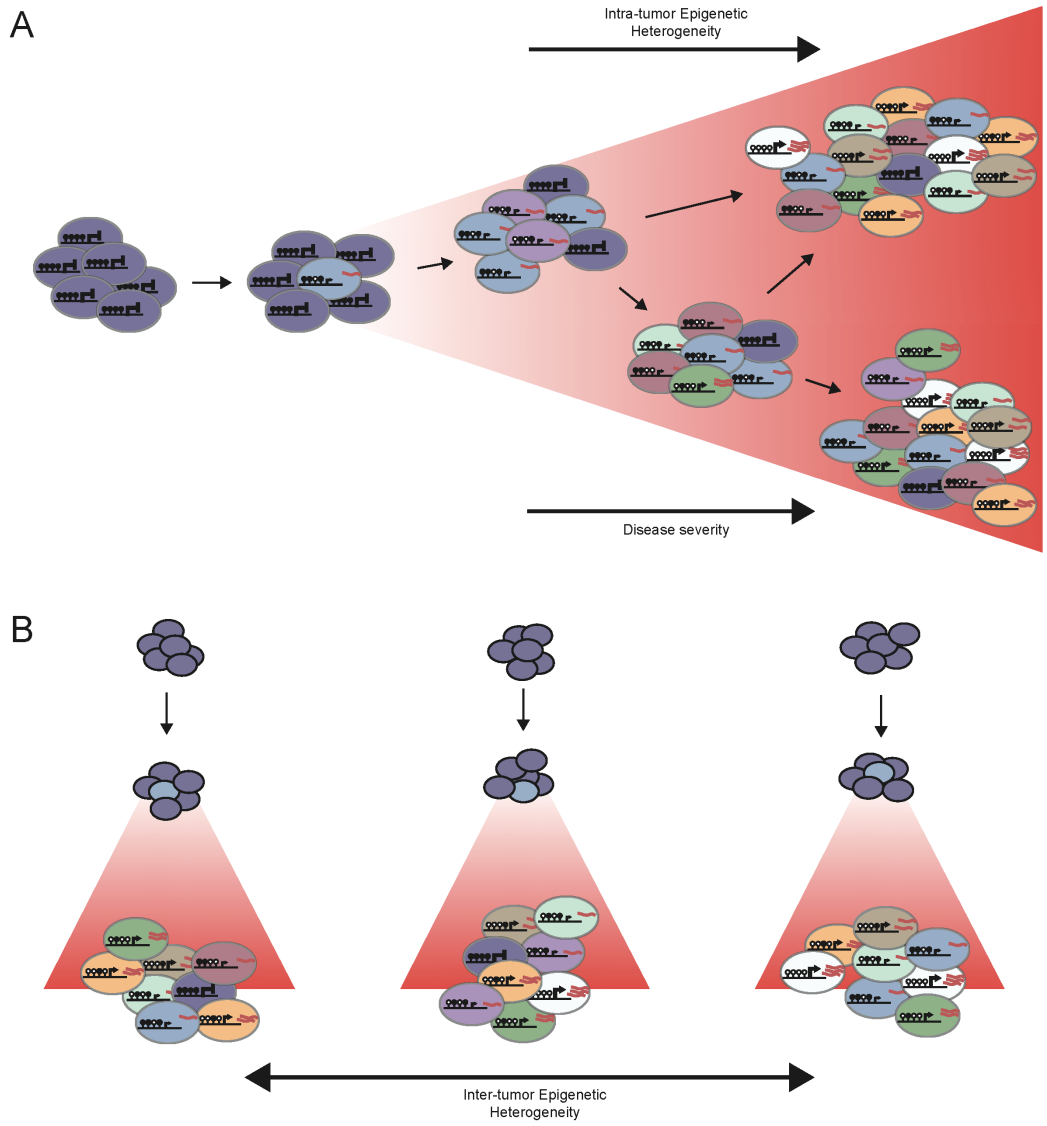

Illustration 2.1. Cytosine methylation heterogeneity reflects a more diverse population and greater disease severity. (A) Representative schematic showing the transition from low to high intra-population heterogeneity. Circles represent unmethylated (white) or methylated (black) cytosines and are believed to affect transcriptional regulation within cells. The color of cells indicates the respective epigenetic state. High intra-tumor heterogeneity indicates a diverse tumor subclone population and is associated with poor clinical outcome. (B) Epigenetic plasticity can manifest as inter-tumor heterogeneity, indicating deviation of individual tumors from 'the cell of origin' signature and from each other.

2.14 Materials and methods

Mouse and Human B Cell Isolation

Aicda$^{-/-}$ mice were a generous gift from T. Honjo. WT (BALB/c) mice were from The Jackson Laboratory. All animals were maintained according to the guidelines of the Research Animal Resource Center of the Weill Cornell Medical College, which approved all mouse procedures. 10- to 12-week-old WT or Aicda$^{-/-}$ mice were immunized intraperitoneally with NP-CGG ratio 20-25 (Biosearch Technologies) in alum (1:1) to induce GC formation. Mice were sacrificed at day 10 after immunization, spleens were dissected, and mononuclear cells were purified using Histopaque (Sigma) gradient centrifugation. Cell suspensions were enriched in B cells by positive selection with anti-B220 magnetic microbeads (Miltenyi Biotech). B cells were separated in NB (B220+GL7-FAS-DAPI-) and GCB (B220+GL7+FAS+DAPI-) using a BD FACSAria II sorter. Spleens of VavP-*Bcl2* and VavP-*Bcl2*+*Aicda* mice were dissected, cell suspensions were prepared, and mononuclear cells were purified as above.

Leftover human tonsils were obtained after routine tonsillectomies, performed at New York Presbyterian Hospital. All tissue collection was approved by the Weill Cornell Medical College Institutional Review Board. Tonsils were minced and mononuclear cells were isolated using Histopaque density centrifugation. NBs for were separated by positive selection using the AUTOMACS system (Miltenyi Biotech) after incubation with anti-IgD fluorescein isothiocyanate (FITC) (BD Pharmingen) followed by anti-FITC microbeads (Miltenyi Biotec). GCBs were separated by positive selection with anti-CD77 (AbD Serotec)

followed by mouse anti-IgM, IgG1 isotype (BD Pharmingen), and anti-mouse-IgG1 microbeads (Miltenyi Biotec).

***Ex vivo* activated B-cell cultures**

Mouse splenic NB were purified by negative selection with anti-CD43 magnetic beads (Miltenyi Biotec) and were stimulated with 25 μg/ml LPS (Sigma Aldrich), 25 ng/ml IL-4 (R & D Systems) and 1 μg/ml anti-CD40 (eBioscience). One day after culture, cells were infected with EV or AICDA-expressing retroviral vector pMIG, described previously[273]. The cDNA encoding mouse AICDA was generated by PCR amplification and was subsequently cloned into the multiple cloning site of pMIG[274]. Total cultures were analyzed after 96 h stimulation and GFP+IgG1+ cells were sorted using a BD FACSAria II sorter.

Flow Cytometry Analysis and Antibodies

Flow cytometry analysis of mouse NBs and GCBs was performed using the following fluorescent-labeled anti-mouse antibodies: APC-conjugated anti-B220 (BD Pharmingen), PE-Cy7-conjugated anti-CD95 (BD Pharmingen), FITC-conjugated anti-GL7 (BD Pharmingen), and PE-conjugated CXCR4 (eBioscience). Cell-cycle analysis was performed using the BrdU Flow Kit (BD Pharmingen), and antigen-specific GCBs (NP+GL7+CD95+B220+) was detected using PE-conjugated NP (Biosearch Technologies). *Ex vivo* stimulated B cells were stained with PE-Cy7-conjugated anti-B220 (eBioscience), PE-conjugated anti-IgD (BD Pharmingen), and APC-conjugated anti-IgG1 (BD Pharmingen). DAPI was used for the exclusion of dead cells. Data were acquired on a MACSQuant analyzer (Miltenyi Biotec) and analyzed using FlowJo 7.6.4 software (Tree Star).

Animal models

Bone marrow cells from 8-10-week-old BCL2 transgenic animals (VavP-*Bcl2*)[275] were harvested and transduced with viral supernatants containing either EV or AICDA-expressing retroviral vector pMIG. Two million bone marrow cells of each condition were injected into the tail veins of lethally irradiated recipient C57BL/6J mice. All mice were followed until any one of several criteria for euthanizing were met, including severe lethargy and more than 10% body weight loss in accordance with our Weill Cornell Medicine Institutional Animal Care and Use Committee–approved animal protocols. All animals were maintained according to the guidelines of the Research Animal Resource Center of Weill Cornell Medicine.

DLBCL Patient Samples

Human DLBCL samples were obtained as previously described by Pan *et al*[261]. In brief, genomic DNA (gDNA) was extracted from frozen solid tissue sections. The tumor purity of DLBCL samples was above 80-90% based on histological observation. Frozen tissue samples were first digested overnight with 0.5 mg ml−1 Proteinase K and 0.625% SDS in 4 ml Nucleic Lysis Buffer at 37 °C. After digestion, 1 ml of saturated NaCl was added to the samples and samples were shaken vigorously for 15 s before spun at 2,500 r.p.m. for 15 min. Supernatant was transferred to a new tube and mixed with two volumes of room temperature 100% ethanol. DNA was precipitated by centrifugation at maximum speed for 30 min, washed twice with 70% ethanol and finally dissolved in TE or nuclease-free water overnight at room temperature.

To assess *AICDA* expression differences between DLBCL subtypes, we analyzed 287 newly diagnosed and characterized DLBCL cases, in which individuals were treated with R-CHOP (given with curative intent) at the BC

Cancer Agency (Vancouver)[276]. These studies were approved by the Research Ethics Board at The University of British Columbia, British Columbia Cancer Agency (REB#H13-01478).

Enhanced reduced representation bisulfite sequencing (ERRBS)

To quantify the epigenetic diversity within the different groups, we calculated pairwise methylation distance between ERRBS profiles using rectilinear Manhattan distance normalized to 1e+06 CpGs for all CpGs that were represented between both samples. The value of the distribution of all pairwise distances within a group of samples defines the average methylation diversity of that group. To make no assumptions regarding the distribution of these pairwise distances, Wilcoxon rank sum tests were used to compare between conditions. Genomic distribution of CpGs was determined using the ChIPseeqerAnnotate module in the ChIPseeqer package[277]. Significance of genomic distributions of DMCs versus all represented CpGs was assessed using Fisher's exact test. DMC-enriched repeat elements were identified using 2-fold threshold for DMC fraction relative to all represented CpG fraction and a Bonferroni-adjusted binomial test $p<0.05$. Association of genes and DMRs was performed by identifying RefSeq genes that overlap DMRs within -2kb of TSS to transcription end site (TES).

Murine gDNA was extracted using the Puregene Gentra cell kit (QIAGEN) and eluted in TE. The gDNA quality was assessed using 1% agarose gel to assure no shearing. Quality of purified DNA was checked using an Agilent 2100 Bioanalyzer. 50 ng gDNA was bisulfite converted using the EZ DNA Methylation kit (Zymo Research). Base-pair-resolution DNA methylation analysis was performed on gDNA following the ERRBS protocol previously described[230]. To compare loss of highly methylated CpGs in VavP-*Bcl2+Aicda*

tumors, we performed a Fisher's Exact Test and classified 'high methylation' state as CpGs with mean methylation level greater than 70%. Delta mean methylation was calculated by subtracting the mean control (e.g. VavP-*Bcl2* tumors, WT GC B cells, AICDA-low DLBCL) methylation value from the mean experimental (e.g. VavP-*Bcl2*+*Aicda* tumors, *Aicda*$^{-/-}$ GC B cells, AICDA-high DLBCL) methylation value. Delta IQR among replicates were calculated by subtracting the IQR among control replicates from the IQR among experimental replicates. PCA was performed on mean-centered data. AICDA-perturbed CpGs were defined as CpGs with component loading factor greater than two standard deviations above mean loading factor. To quantify the degree of intra-sample heterogeneity, we calculated the minimum distance from the methylation value of each CpG to [0,1] (*i.e.,* the distance away from the closest homogenous unmethylated/methylated state). Using this metric, the maximum intra-sample heterogeneity is 0.5, reflecting a population state with half of loci being methylated and the other half being unmethylated. To compare intra-tumor heterogeneity between conditions without assumptions of distribution, we performed pairwise Wilcoxon rank sum tests. Genes over-representing signature CpGs were identified according to hypergeometric test (Benjamini-Hochberg-adjusted $P<0.05$) using all ERRBS-represented CpGs within the gene body (interval from +2kb of TSS to TES).

RNA sequencing

Total RNA was extracted from murine tissues or human DLBCL patient samples using Trizol (LifeTechnologies) and RNeasy isolation Kit (Qiagen). RNA concentration was determined using Qubit (LifeTechnologies) and integrity was verified using Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries were generated using the TruSeq RNA sample kit (Illumina). First-

strand synthesis was performed using random oligos and SuperscriptIII (Invitrogen). After second-strand synthesis, a 200-bp paired-end library was prepared following the Illumina paired-end library preparation protocol. Pair-end sequencing (PE50) was performed on Illumina HiSeq2000. RNA sequencing results were aligned to mm10 or hg19, respectively, using STAR[278] and annotated to RefSeq using the Rsubread package[279]. DLBCL subtype classification was performed using RNAseq profiles as described in Cardenas, *et al*[280].

Real-Time qPCR

cDNA synthesis from RNA was performed using the Verso cDNA Synthesis kit (Thermo Scientific). The expression was detected using the Green FastMix kit (Applied Biosystems) on a 7900HT Fast RT-PCR System (Applied Bio-systems). Gene expression was normalized to RPL13 using the DDC(t) method, and results were represented as fold expression compared to NBs.

Primer sequences for qPCR

Used for	Gene		Oligonucleotide (5'-3')
qPCR	*Aicda*	Fwd	GGAGAGATAGTGCCACCTCC
		Rev	TCTCAGAAACTCAGCCACGT
PCR	*Abcc4*	Fwd	GCCCCTAAGCTACCAGCTCT
		Rev	GCAGAACAGGGTCTCTCGTC
	Cd83	Fwd	ACGCTTGCTCCCTCTTTACA
		Rev	GCTCGGGGGAAATTTTACTT
	Rassf3	Fwd	ATGAGCCCGGTTTAATCCTC
		Rev	ATAGGTACGAAGCGCACCAC
	JH4	Fwd	GTCAAGGAACCTCAGTCACCGTCT
		Rev	CAGACCTCTCTAGACAGCAACTACC
	Sμ	Fwd	AATGGATACCTCAGTGGTTTTTAATGGTGGGTTTA
		Rev	GCGGCCCGGCTCATTCCAGTTCATTACAG

V(D)J rearrangement analysis

PCR products were cleaned-up using the MinElute PCR Purification kit (QIAGEN) and subsequently were purified on a gel using the Gel Extraction kit

(QIAGEN). Sequencing libraries were constructed from the purified PCR product by using Illumina TruSeq DNA Sample Preparation Kit v2 (Illumina). Each sample was tagged with a unique index and sequenced on the Illumina MiSeq platform producing 2x151bp paired-end reads. Paired-end sequence reads were mapped against the IGH mouse reference database available from the IMGT website (reference: imgt.org) using a modified nucleotide blast search. Sequences without a match in Ig heavy chain V regions were filtered out and reads with a corresponding match to a VH region were counted for the number of each unique rearrangement. We then performed statistical analysis to see whether there was a difference in the selection of Ig rearrangements between the samples by comparing the clonality. As described in Jiang *et al.*[281], clonality was measured using an adjusted measure of Shannon's entropy of the distribution of VH region selection counts. This gives some indication as to the shape of the distribution of VH regions in each sample and is influenced by the number of reads mapped and how clonal each sample is. The lower the entropy is, the more skewed towards a single VH region the distribution is (i.e., the more clonal it is).

Immunoblotting

Total cell extracts were prepared after treatment with lysis buffer (10 mM Tris [pH 7.4], 150 mM NaCl, 1% NP-40, 0.5% deoxycholate, 1% Triton X-100, 0.1% SDS, and 1mM EDTA) supplemented with PMSF (Sigma) and a protease inhibitor cocktail (Roche). Lysates were subjected to SDS-PAGE, transferred to a polyvinylidene fluoride membrane (Bio-Rad Laboratories), and blotted with anti-AID (L7E7, Cell Signaling Technology) or anti-actin (A5441, Sigma). Signals were detected with horseradish-peroxidase-conjugated

secondary antibodies (Santa Cruz Biotechnology) using the ECL system (Thermo Scientific).

CHAPTER THREE:

EZH2 TRANSCRIPTIONAL REPRESSION IN GC REACTION AND GC-DERIVED LYMPHOMAGENESIS*

3.1 EZH2 activity is key to germinal center formation

Previous studies have identified increased expression of EZH2 in GC B cells and demonstrated the importance of EZH2 to the growth of lymphoma cells[47,92-94]. However, the role and requirement for EZH2 and its mutant alleles in the formation of GCs and pathogenesis of B cell lymphomas is not well explored. Additionally, the mechanism by which EZH2 contributes to transformation, beyond association with global cellular abundance of H3K27me3, is unknown. To explore the function of WT and mutant EZH2 in mature B cells and lymphomas, we attempted to generate a conditional $EZH2^{-/-}$ mouse model. EZH2 deletion is lethal in early embryonic development[48], and its inducible knockout in early hematopoietic cells perturbs lymphoid differentiation at the pre-B cell stage[50]. We therefore crossed

---

* Béguelin W, Popovic R*, **Teater M***, Jiang Y, Bunting KL, Rosen M, Shen H, Yang SN, Wang L, Ezponda T, Martinez-Garcia E, Zhang H, Zheng Y, Verma SK, McCabe MT, Ott HM, Van Aller GS, Kruger RG, Liu Y, McHugh CF, Scott DW, Chung YR, Kelleher N, Shaknovich R, Creasy CL, Gascoyne RD, Wong KK, Cerchietti L, Levine RL, Abdel-Wahab O, Licht JD, Elemento O, Melnick AM. (2013) EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell 23(5):677-92*

* Béguelin W, **Teater M**, Gearhart MD, Goldstein RL, Cardenas MG, Rosen M, Shen H, Corcoran CM, Hamline MY, Gascoyne RD, Levine RL, Abdel-Wahab O, Licht JD, Shaknovich R, Elemento O, Bardwell VJ, Melnick AM. (2016) EZH2 and BCL6 cooperate to assemble CBX8-BCOR Polycomb complex to repress bivalent promoters, mediate germinal center formation and promote lymphomagenesis. *Cancer Cell 30(2):197-213*

conditional *Ezh2*$^{-/-}$ mice with the Cγ1cre strain, which expresses cre recombinase in GC B cells[282]. Upon reaching immunological maturity, *Ezh2*$^{fl/fl}$, *Ezh2*$^{-/-}$, and WT control mice were injected with T cell-dependent antigen sheep red blood cells (SRBC) to induce GC formation and sacrificed 10 days later, at which time the GC reaction is at its peak. EZH2 loss resulted in a marked reduction in the number of splenic GC (GL7+/FAS+/B220+) B cells (**Figure 3.1A**). Immunohistochemical analysis using peanut agglutinin (a GC B cell marker) revealed a reduction in the number ($p < 0.0001$) and size ($p = 0.001$) of GCs in *Ezh2*$^{-/-}$ versus *Ezh2*$^{+/+}$ mice, whereas there was no change in GCs in *Ezh2*$^{fl/fl}$ mice (**Figure 3.1B**). We then immunized a cohort of C57BL6 mice with SRBC followed by once daily treatment with 150 mg/kg/day GSK503, a specific EZH2 methyltransferase inhibitor, or vehicle. GSK503, but not vehicle, prevented the formation of GC after SRBC or NP-KLH immunization, phenocopying the *Ezh2* null phenotype. GSK503 treatment led to reduced numbers of GC B cells by flow cytometry (**Figure 3.1C**) and reduced number and volume of GCs by immunohistochemistry (**Figure 3.1D**). Collectively, these data demonstrate that EZH2 is required for the formation of GCs and that this function is dependent on its histone methyltransferase activity.

Given that EZH2 was essential for development of GC B cells, we surmised that gain-of-function EZH2 mutants might reinforce these actions, with implications for lymphomagenesis. To address this point, we generated mice conditionally expressing the *Ezh2*^{Y641N} lymphoma in GC B cells, using a Cγ1-cre background. We observed a 3-fold increase in the abundance of H3K27me3 in sorted GC B cells from *Ezh2*^{Y641N} versus nonrecombined mice

($p < 0.005$; **Figure 3.1E**), analogous to what is observed in mutant EZH2 DLBCL cell lines[88,93]. After immunization, *Ezh2*^{Y641N} mice displayed greatly increased numbers of GL7+/FAS+/ B220+ GC B cells ($p < 0.001$; **Figure 3.1F**) and increased number and size of GCs ($p < 0.0005$; **Figure 3.1G**), while cre-negative *Ezh2*$^{fl/fl}$ animals showed no such effects. These data demonstrate that Y641 mutation reinforces the GC phenotype-driving function of wild-type EZH2, resulting in an expansion of these proliferative and mutagenic cells. However, *Ezh2*^{Y641N} knockin mice did not develop B cell lymphomas (data not shown).

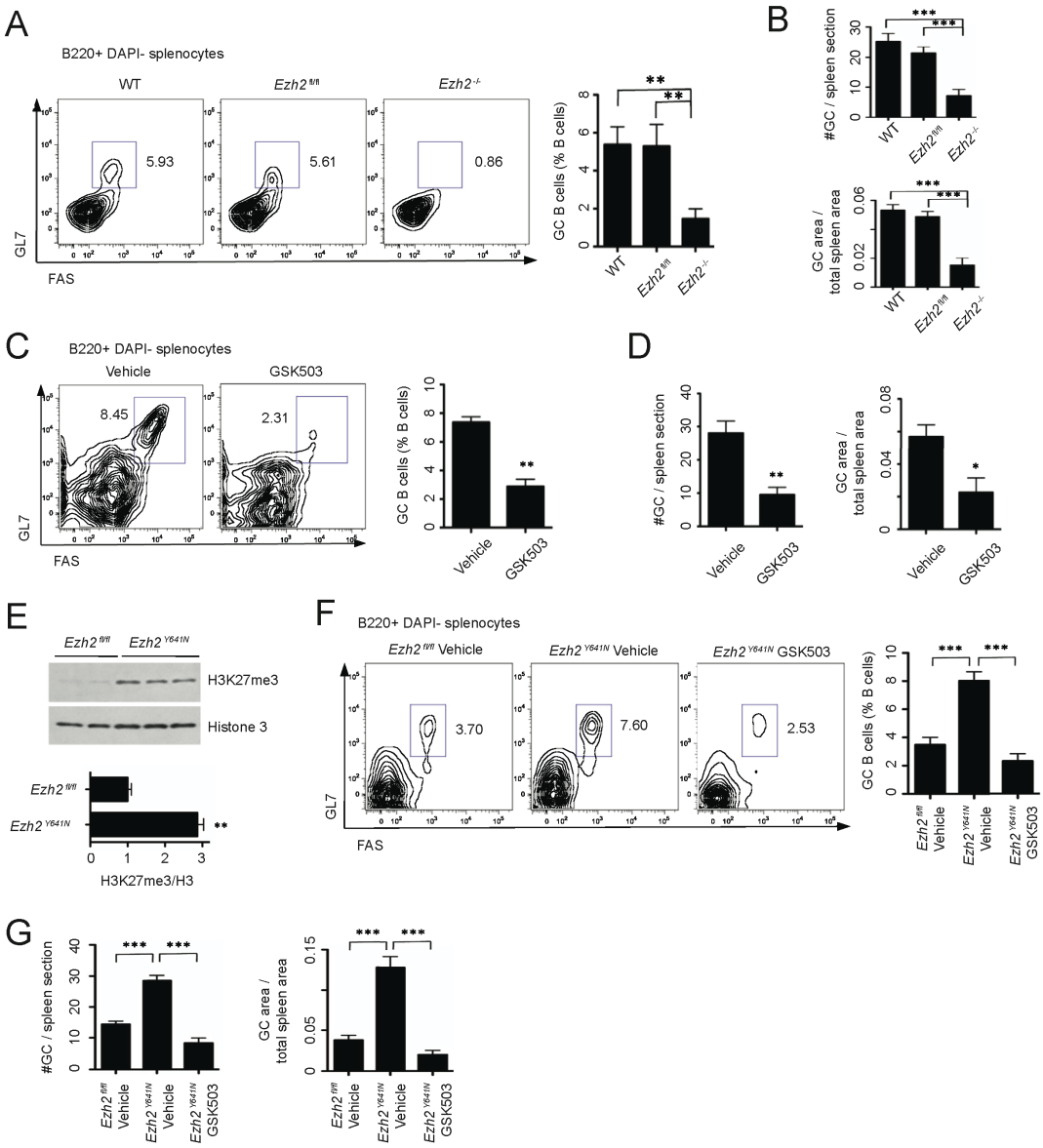

Figure 3.1. EZH2 activity is key to germinal center formation. (A-B) *EZH2*$^{fl/fl}$, *EZH2*$^{-/-}$, and WT C57BL6 control mice (n = 5 per group) were immunized with SRBC to induce GC formation. (C–D) C57BL6 mice were immunized with SRBC and treated with GSK503 (150 mg/kg/day) or vehicle (n = 5 per group). (E-G) *EZH2*$^{fl/fl}$ and *EZH2*^{Y641N} mice were immunized wtih NP-KLH and treated with GSK503 (150 mg/kg/day) or vehicle (n=6 per group). (A, C, F) Representative flow cytometric plots of splenic GC cells and quantification. (B, D, G) Quantitation of PNA staining. (E) Immunoblotting from whole cell lysates from sorted splenic GC B cells (GL7+/FAS+/B220+). Values are shown as mean ± SEM. t test, *p < 0.05, **p < 0.01, ***p < 0.001. Experiments and data prepared by W. Beguelin.

3.2 EZH2 mediates differentiation blockade in DLBCL cells

GCB-DLBCLs exhibit differentiation blockade, locking them into the GC phenotype[283]. Given that gain-of-function mutants induce an expansion in the population of GC B cells, we postulated that EZH2 might play a critical role in suppressing differentiation. To test this, we expressed wild-type EZH2 and $EZH2^{Y641}$ in murine BCL1, a cell line originating from a spontaneous murine lymphoproliferative disease that exhibits clonal somatic hypermutation and has been used to model GC B cell biology[284]. BCL1 cells were transduced with GFP-expressing retrovirus harboring FLAG-tagged $Ezh2^{Y641N}$, $Ezh2^{Y641F}$, or WT *Ezh2* and subsequent RNAseq was performed. Comparing RNAseq gene expression profiles in transduced BCL1 cells, we observed that $EZH2^{Y641N}$ and $EZH2^{Y641F}$ increased repression of genes that are normally expressed upon terminal differentiation into plasma and memory cells when compared to EZH2-WT-expressing cells (gene set enrichment analysis [GSEA] false discovery rate [FDR] q = 0.009; **Figure 3.2A**). We then induced differentiation in BCL1 cells using interleukin-2 (IL-2) and IL-5, which typically drive BCL1 cells toward a partial plasma cell phenotype accompanied by proliferation arrest[284]. Expression of $EZH2^{Y641N}$ or $EZH2^{Y641F}$ but not WT-EZH2 significantly impaired cytokine-induced proliferation arrest (t-test $p < 0.05$; **Figure 3.2B**) and more profoundly suppressed activation of *Prdm1*, a key plasma cell differentiation gene (t-test $p < 0.01$; **Figure 3.2C**).

Based on these data, we next examined whether EZH2 inhibition by GSK343 could induce differentiation in human DLBCL cells. In GCB-type DLBCL cell

lines (two mutant and one WT), EZH2 inhibitor treatment induced functional evidence of differentiation manifested by increased light chain and immunoglobulin production (**Figure 3.2D**) and morphologic changes consistent with plasma cell differentiation (**Figure 3.2E**). Gene expression analysis by quantitative PCR revealed upregulation of differentiation-related genes, including *PRDM1*, *IRF4*, and syndecan *CD138*, with greater effect generally noted in EZH2 mutant DLBCL cells ($p < 0.05$; **Figure 3.2F**). Analysis of RNAseq expression profiles induced by GSK343 found strong enrichment in gene sets involved in exit from the GC reaction, including genes induced by CD40, IRF4, IL10, and NFκB, and, notably, the ABC-DLBCL signature (which consists of activation of these post-GC genes), as well as genes involved in apoptosis and immune responses (**Figure 3.2G**). Memory and plasma cell differentiation genes were among genes preferentially repressed in GCB-type DLBCL patients with EZH2 mutations as compared to GCB-DLBCL patients with WT EZH2 (FDR $q < 0.001$; **Figure 3.2H**). Collectively, these data suggest that EZH2 maintains the GC phenotype by suppressing transcriptional programs required for exiting the GC reaction and terminal differentiation. Mutant EZH2 augments these same functions by reinforcing repression of B cell differentiation genes.

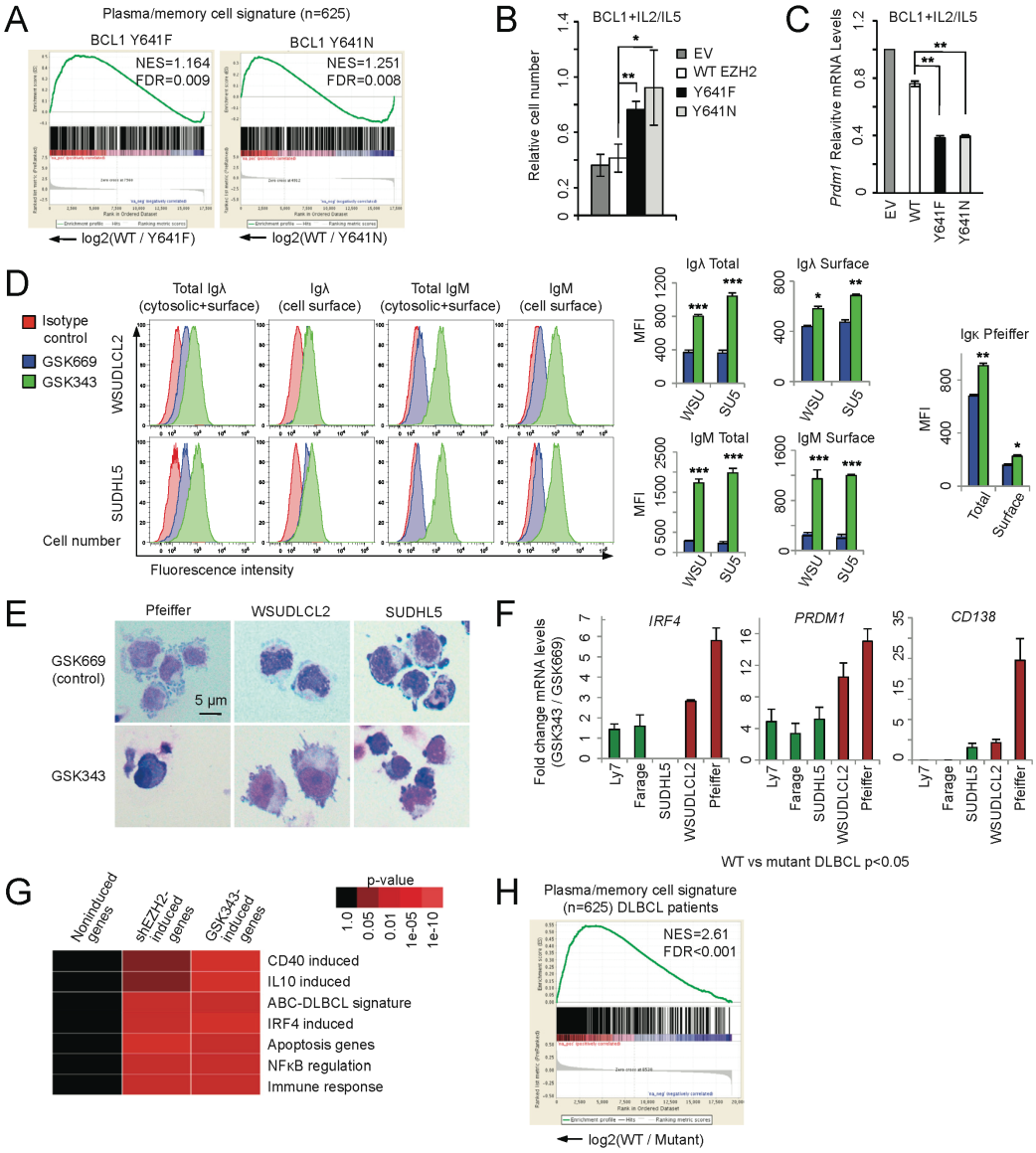

Figure 3.2. EZH2 mediates differentiation blockade. (A) GSEA in WT versus mutant EZH2 BCL1-transduced cells. (B) Cells were treated with 20 ng/ml IL2 and IL5 for 5 days and viability measured using trypan blue exclusion. (C) qRT-PCR of *Prdm1*. (D) Cells were treated with 2 μM GSK343 or control GSK669 for 7 days. Ig expression levels were examined by flow cytometry. (E) Representative pictures of cells treated as in (D). (F) qRT-PCR of indicated mRNAs in cells treated with 2 μM GSK343 for 7 days. Green, WT EZH2, red, mutant EZH2 cell lines. (G) Heat map of over-represented gene categories among genes up-regulated by *EZH2* shRNA or treatment with 2 μM GSK343 for 7 days in Ly7, Ly1, SUDHL5, Farage, WSDLCL2, and Pfeiffer cells. Enrichment measured using hypergeometric p-values. (H) GSEA in GCB-DLBCL patient samples with WT versus mutant EZH2. Values in (B-D) and (F) are mean of duplicate or triplicate ± SD. t-test, *p < 0.05, **p < 0.01, ***p < 0.001. BCL1 experiments performed by R. Popovic. DLBCL cell line experiments performed by W. Beguelin.

3.3 Mutant EZH2 aberrantly repressed gene via increased promoter H3K27me3

In order to gain insight into how mutant EZH2 might reinforce or alter normal EZH2-dependent transcriptional programs, we performed H3K27me3 ChIPseq in BCL1 cells after transduction with $EZH2^{Y641N}$, $EZH2^{Y641F}$, WT EZH2, or vector and determined how gene expression in these cells responded to EZH2 inhibitors. Cells transduced with either $EZH2^{Y641N}$ or $EZH2^{Y641F}$ showed a marked increase in the abundance of H3K27me3 at gene promoters compared to EZH2 WT-transduced cells (Wilcoxon $p < 2 \times 10^{-16}$ for both mutants; **Figure 3.3A**), and genes with abundant H3K27me3 were generally repressed ($p < 2 \times 10^{-16}$; data not shown). In addition, loci displaying increased H3K27me3 were more potently induced upon exposure of the BCL1- $EZH2^{Y641N}$/$EZH2^{Y641F}$ cells to GSK343 than EZH2 WT cells (Wilcoxon $p < 5 \times 10^{-11}$ for $EZH2^{Y641N}$ and $EZH2^{Y641F}$; **Figure 3.3B**). Hence, mutant EZH2 appears to function at least in part by increasing the burden of epigenetic repressive marks at gene promoters.

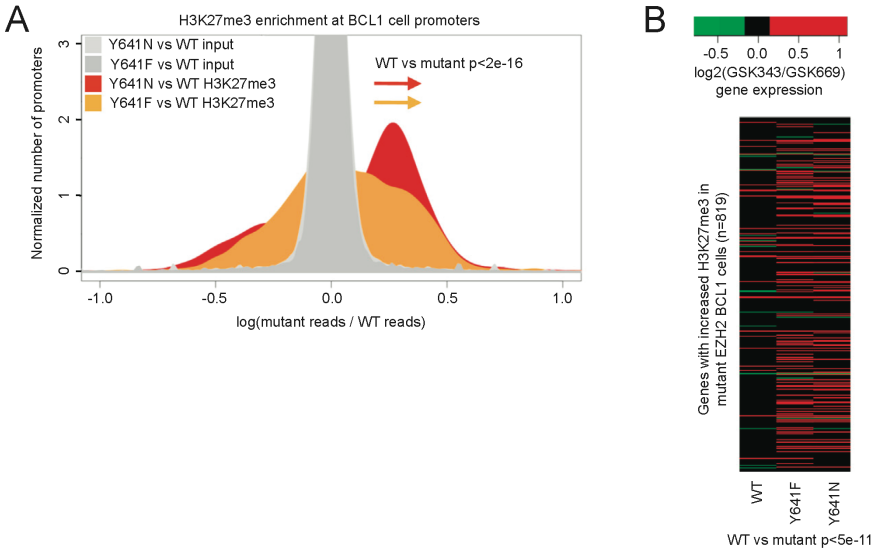

Figure 3.3. Mutant EZH2 aberrantly represses genes by increasing H3K27me3 at promoters. (A) Density plot showing log ratio of WT versus *EZH2* mutant ChIP-seq reads. Input reads in both mutant comparisons show no difference between WT and mutant and are shown as a narrow distribution centered on zero. H3K27me3 comparisons are shifted to right for both mutants, indicating increased H3K27me3 reads in mutant promoters compared to WT controls. (B) Heatmap showing change in expression of genes with increased promoter H3K27me3 in BCL1 cells after treatment with 0.5 μM GSK343 for 3 days.

3.4. Identification of H3K27me3 enrichment from ChIPseq experiments

Although unbiased genomic surveys of H3K27me3 can be made using ChIPseq, identification of H3K27me3-enriched loci can be problematic. Repressive histone modifications, such as H3K27me3, typically cover long, continuous regions and form heavily condensed heterochromatin[285,286]. Unlike transcription factors and histone modifications that span a small number of nucleosomes and show narrow and strong ChIPseq enrichment, the broad and low-level enrichment of H3K27me3 cannot be detected reliably by ChIPseq peak callers, such as MACS which is designed to detect in narrow windows[287].

In order to identify regions of H3K27me3 enrichment in NB and GCB, we first quantified H3K27me3 occupancy into read counts genome-wide in 1kb bins and transformed the counts into z-scores (**Figure 3.4A**). To remove single outlier bins with exceptionally high H3K27me3 ChIPseq read counts located adjacent to bins with read counts near or below the genome-wide mean, we fit a cubic spline to the z-scores within a sliding window of 1MB, defining the loci corresponding to spline-fitted z-score values above 1 as putative "broad domains" (**Figure 3.4B**). This approach identified broad regions of enrichment that showed good correspondence with observed H3K27me3 ChIPseq read density (**Figure 3.4C**). To validate that these called regions of H3K27me3 enrichment were associated with transcriptional repression, we compared the expression of genes in human GCB with and without H3K27me3 broad domain enrichment within their promoters (±2kb TSS). We found that genes not associated with broad domains exhibited a range of expression values, but genes with broad domains overlapping their promoters were silenced or restricted to low expression values (**Figure 3.4D**).

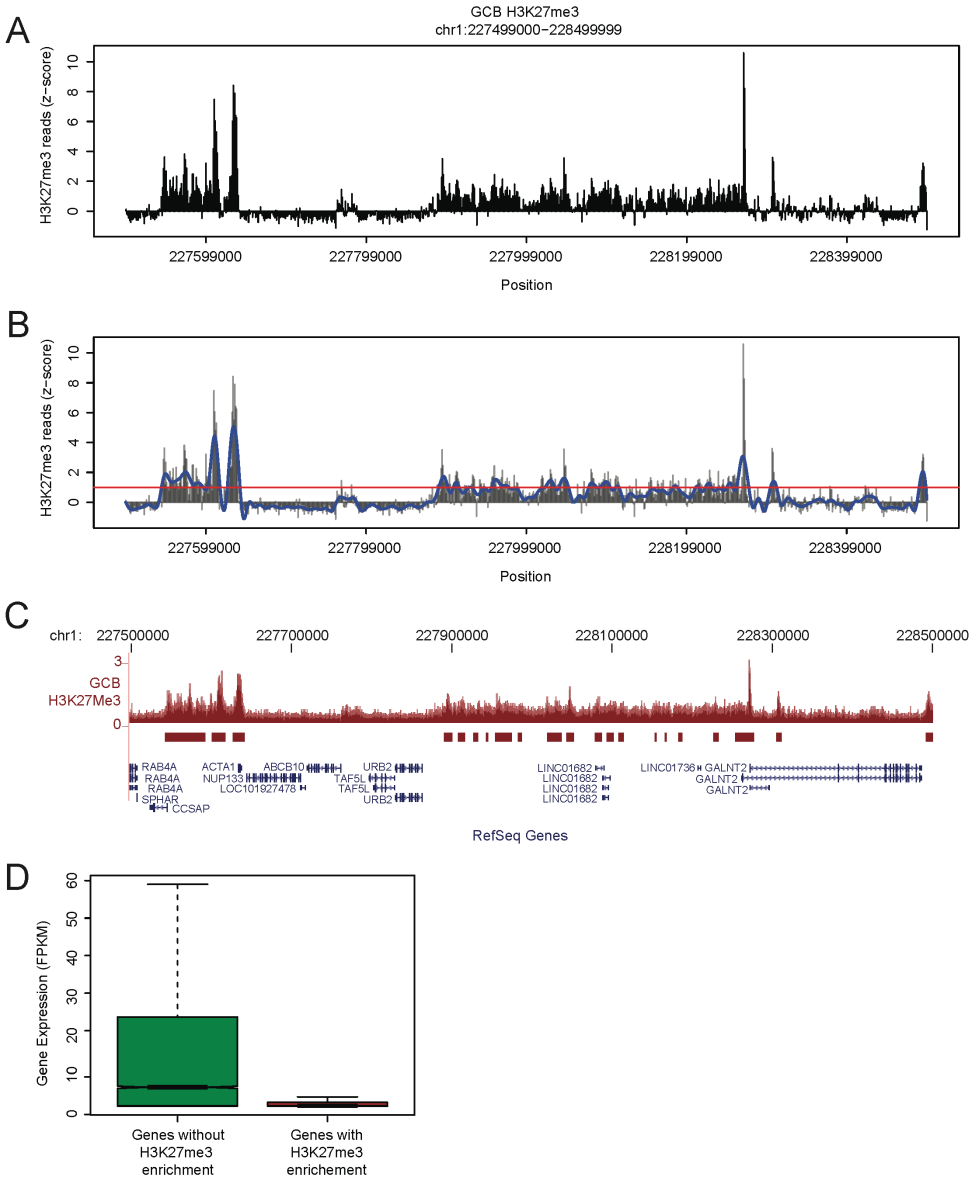

Figure 3.4. H3K27me3 enrichment within promoters is associated with low transcription. (A) Representative 1Mb interval showing the relative enrichment of H3K27me3 ChIPseq in GCB. Enrichment was quantified a z-scores of read counts in 1kb bins across the genome. (B) Smoothing spline (blue line) of relative enrichment data. Smoothed data above 1 were defined as broad domains. (C) Representative interval showing H3K27me3 read density and called peaks. (D) Box plot showing gene expression of genes without H3K27me3 broad domains overlapping their promoters and genes with promoters that intersect H3K27me3 broad domains.

3.5 EZH2 is linked to *de novo* formation of GCB-specific bivalent genes that are implicated in differentiation

To assess the transcriptional impact of mutant EZH2 upon the GCB-specific EZH2 targets, we identified genes associated with promoter H3K27me3 broad domains and are downregulated in GCB compared to NB cells. We found that these GCB cell-specific H3K27me3 targets (n=510) were further repressed in BCL1 cell transduced with $EZH2^{Y641F}$ or $EZH2^{Y641N}$ than in their BCL1- WT EZH2 counterparts (FDR $q < 0.001$; **Figure 3.5A**), suggesting that mutant EZH2 exaggerates the epigenetic silencing of normal GC B cell EZH2 targets.

In stem cells, a subset of genes that are H3K27 trimethylated by EZH2 are also marked by the activating chromatin modification H3K4me3. These so-called ‘‘bivalent (H3K4me3/H3K27me3) domains’’ are believed to represent a mechanism for poising key lineage transcription factors so that they can be either activated or repressed during subsequent differentiation[225]. We examined H3K27me3 and H3K4me3 distribution in purified primary human NB cells and GC B cells by ChIPseq. Notably, even though GC B cells are mature, committed cells (i.e., in contrast to stem cells), we find that they gain 1,026 new bivalent domains at promoters that are not found in NB cells (**Figure 3.5B**). Of these 910 (88%), bivalent promoters originate from H3K4me3-only promoters in NB cells (**Figure 3.5B**), indicating that most bivalent loci in GC B cells occur due to acquisition of H3K27me3, concordant with upregulation of EZH2 in these cells. ChIPseq profiles from seven human GCB-DLBCL cell lines showed that these bivalent promoters were preferentially bound by EZH2 (Anova $p < 1 \times 10^{-16}$; **Figure 3.5C**). In primary human GC B cells, bivalent genes were expressed at lower levels than loci with H3K4me3 alone but at

higher levels than genes marked only by H3K27me3 (Wilcoxon $p < 1 \times 10^{-300}$; **Figure 3.5D**), consistent with the proposed poised nature of bivalent genes and observations of minimal expression in stem cells[225]. GC B cell bivalent genes were highly enriched in gene sets associated with termination of the GC reaction, such as IRF4-induced, ABC-DLBCL, memory cell upregulated genes, as well as other GC B cell relevant genes, such as negative regulation of cell cycle (**Figure 3.5E**). The key regulatory transcription factors *IRF4* and *PRDM1* were among the GC B cell-specific bivalent genes, consistent with the notion that these genes are marked for dynamic activation upon exit from the GC reaction (**Figure 3.5F**) and that bivalent marks at this stage of B cell maturation function to transiently suppress terminal differentiation.

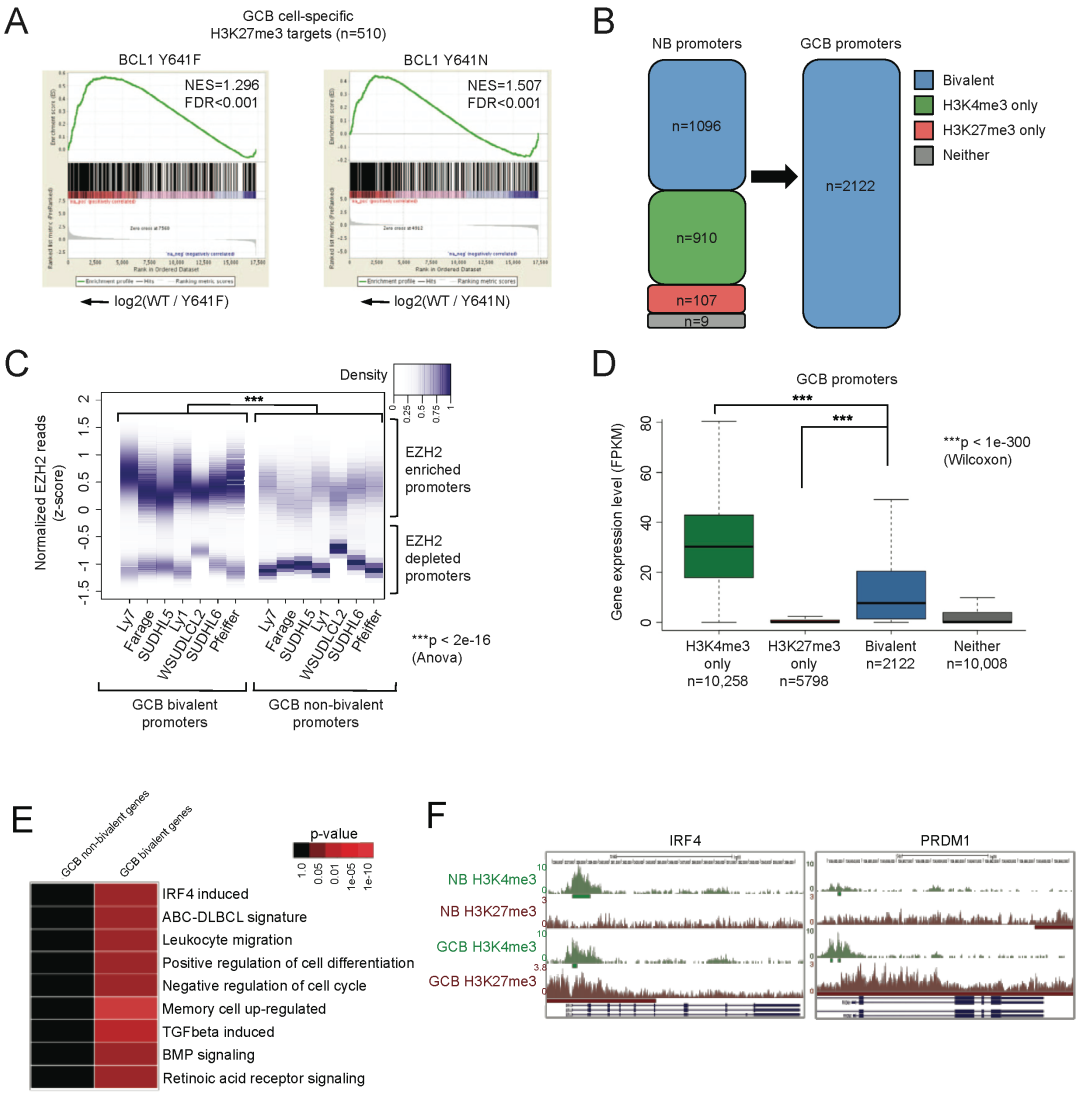

Figure 3.5. EZH2 generates GC B cell-specific bivalent genes involved in differentiation. (A) GSEA of GCB cell-specific H3K27me3 targets in WT versus mutant EZH2 BCL1 cells. (B) State of GCB cell bivalent promoters within NB. (C) Density strip representations of normalized EZH2 ChIPseq reads within DLBCL cell line promoters. (D) Boxplot of gene expression level of bivalent, H3K4me3 monovalent, H3K27me3 monovalent, and promoters not exhibiting either of these marks in GCB cells. (E-F) GSEA of GCB cell-specific bivalent genes in (E) Heatmap of over-represented gene categories among GCB cell bivalent genes. (F) *IRF4* and *PRDM1* gene loci showing H3K4me3 and H3K27me3 read density in NBs and GCBs. Green bars indicate location of H3K4me3 peaks and red bars represent locations of H3K27me3 broad domain enrichment.

We wondered whether somatic mutation of EZH2 might "lock" these bivalent promoters into a more repressed configuration, perhaps helping to explain the irreversible differentiation blockade induced by mutant EZH2. Indeed, we observed that GC B cell bivalent genes were even more highly repressed in murine BCL1 cells transduced with $EZH2^{Y641N}$ and $EZH2^{Y641F}$ than in WT EZH2-transduced cells (FDR q < 0.001; **Figure 3.6A**). GC B cell bivalent domain genes were also significantly more repressed in human GCB-DLBCL patients with EZH2 somatic mutations compared to GCB-DLBCL with WT EZH2 (FDR q < 0.001; **Figure 3.6B**). GC B cell bivalent genes involved in GC B cell differentiation, such as IRF4-induced genes, CD40-induced genes, and plasma/ memory cell genes, were especially enriched among genes differentially expressed in mutant EZH2 DLBCL patients versus GCB-DLBCL specimens with wild-type EZH2 (**Figure 3.6C**). To confirm that bivalent marks are indeed occurring at the same chromatin regions within GC- derived DLBCL cells, we performed ChIP re-ChIP assays. The key regulatory transcription factors necessary for GC exit, *IRF4*, and *PRDM1*, as well as the proliferation checkpoint gene promoter *CDKN1B*, all shown to be putative GCB bivalent genes by our ChIPseq experiments, were confirmed to be significantly co-occupied by H3K27me3 and H3K4me3 marks (**Figure 3.6D**). In contrast, *HOXA7* (which is silenced in mature B cells) was exclusively H3K27 trimethylated and *BCL6* (which is highly expressed in GC B cells) was almost exclusively H3K4 trimethylated. Collectively, these data suggest that, in normal GC B cell development, EZH2 reversibly suppresses terminal differentiation by forming or maintaining bivalent domains at specific loci that are also marked by H3K4me3, in addition to epigenetically silencing other genes through pure H3K27me3. In B cell lymphomas, mutant EZH2

reinforces silencing of these genes, perhaps by increasing the frequency with which these genes are H3K27 methylated among tumor cells or by more subtle stoichiometric effects, tipping the balance of H3K27 toward trimethylation and away from demethylation.

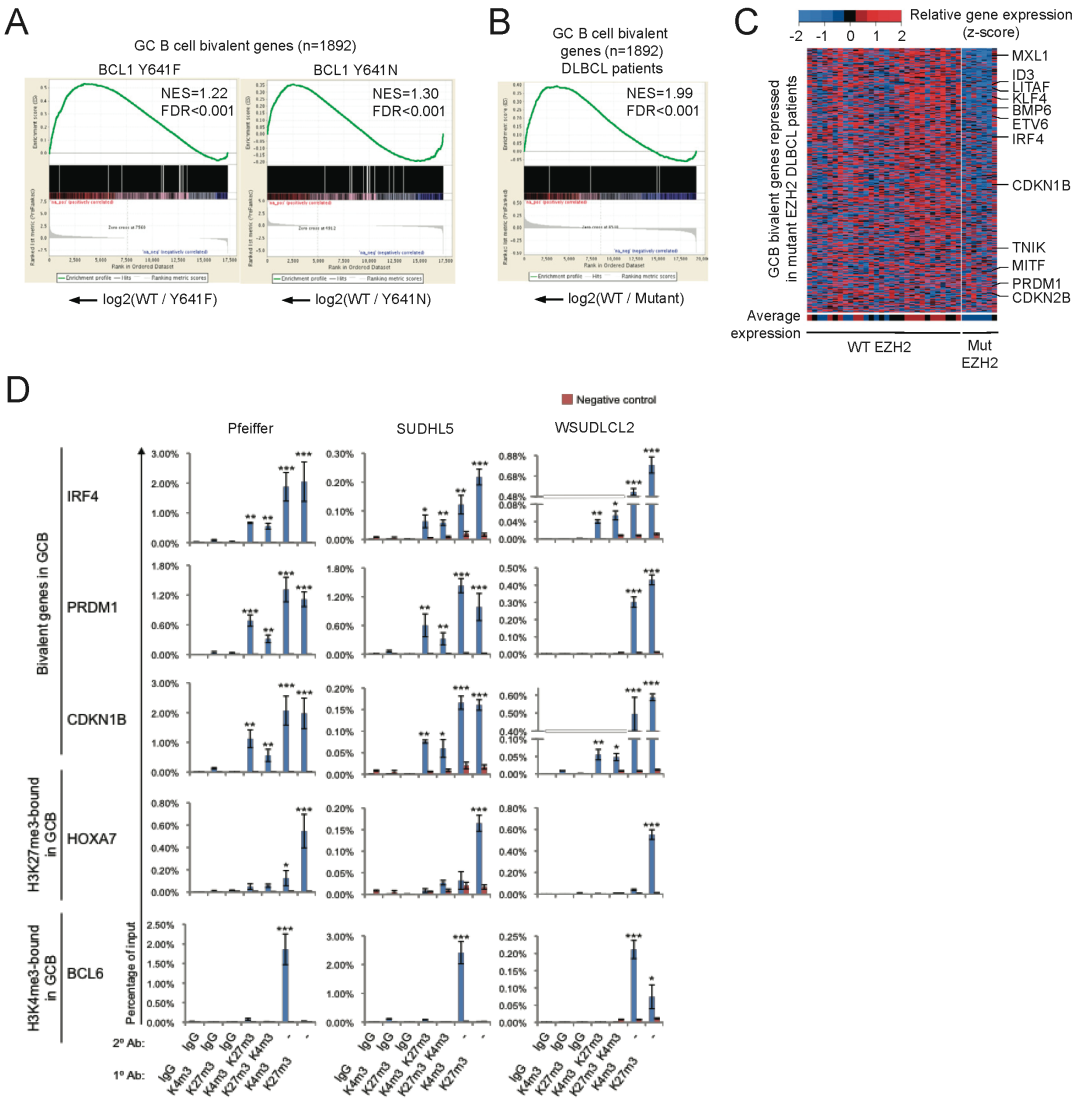

Figure 3.6. Mutant EZH2 further represses GC B cell bivalent genes. (A) GSEA of WT versus mutant EZH2 BCL1 cells and (B) GCB-DLBCL patient samples with WT versus mutant EZH2. (C) Heatmap of relative expression of GCB cell bivalent genes repressed in WT and mutant EZH2 GCB-DLBCL patient samples. (D) ChIP re-ChIP assayed in Pfeiffer, SUDHL5 and WSUDLCL2 cell lines using the indicated antibodies for the first ChIP (1° Ab) and the sequential ChIP (2° Ab). The promoter genes tested for co-occupancy are indicated in the left. As negative control qPCR was performed using primers for a region in chromosome 6 where no H3K27me3 or H3K4me3 marks were found by ChIPseq read density in GCB cells (red bars). Data are presented as mean percentage of input enrichment ± SD for triplicate experiments. t-test ***p<0.001, **p<0.01, *p<0.05 compared with negative control. ChIP re-ChIP experiments performed and data prepared by W. Beguelin.

3.6 EZH2 cooperates with BCL2 to generate GC-derived lymphomas

We next investigated whether mutant EZH2 might cooperate with other GC B cell lymphoma oncoproteins, such as BCL2, which is frequently translocated in patients with $EZH2^{Y641}$ mutations[84]. We transduced bone marrow of VavP-*Bcl2* mice with retrovirus expressing GFP and encoding WT EZH2, $EZH2^{Y641F}$, or GFP alone and transplanted them into lethally irradiated cohorts of ten recipients each (**Figure 3.7A**). Animals were immunized with SRBC every four weeks to ensure continuous formation of GCs. Macroscopic examination of spleens showed marked splenomegaly in $EZH2^{Y641F}$ versus EZH2 WT or empty vector (**Figure 3.7B**). Immunoblot analysis of splenic extracts showed similar levels of expression of EZH2 in the EZH2 WT and $EZH2^{Y641F}$ mice (**Figure 3.7C**). By contrast, H3K27me3 abundance was only elevated in splenocytes from $EZH2^{Y641F}$ mice, whereas H3K27me3 levels in EZH2 WT mice were similar to controls (**Figure 3.7C**). The livers of $EZH2^{Y641F}$ but not EZH2 WT mice were also significantly enlarged versus vector control (**Figure 3.7D**). All EV and WT EZH2 mice displayed evidence of follicular hyperplasia, as expected in *Bcl2* transgenic mice (**Figure 3.7E**). However, examination of spleens in $EZH2^{Y641F}$/*Bcl2* mice revealed disruption of splenic architecture by neoplastic-appearing B220+ B cells (**Figure 3.7E**).

Given that BCL2 and EZH2 can cooperate in lymphomagenesis, we hypothesized that BH3 mimetic drugs that block BCL2 function, such as Obatoclax and ABT737, might enhance the activity of EZH2 inhibitors. We exposed a panel of GCB-DLBCL cells to increasing concentrations of GSK343, in combination with ABT737 or Obatoclax. In almost every cell line tested, the concentration of GSK343 or GSK503 required to yield 90% growth

inhibition was reduced when cells were concomitantly treated with BH3 mimetics (**Figure 3.7F**). In order to determine the impact of EZH2 inhibitor combinatorial therapy in a preclinical model, we evaluated the action of GSK503 and Obatoclax alone or in combination in mice bearing human DLBCL cell line (SUDHL4 and SUDHL6) xenografts. Although both GSK503 and Obatoclax inhibited tumor growth alone, the combination of these inhibitors again more potently and significantly suppressed tumor xenograft growth (**Figure 3.7G**). EZH2 mutants therefore enable, accelerate, and maintain malignant transformation of GC B cells in cooperation with BCL2.

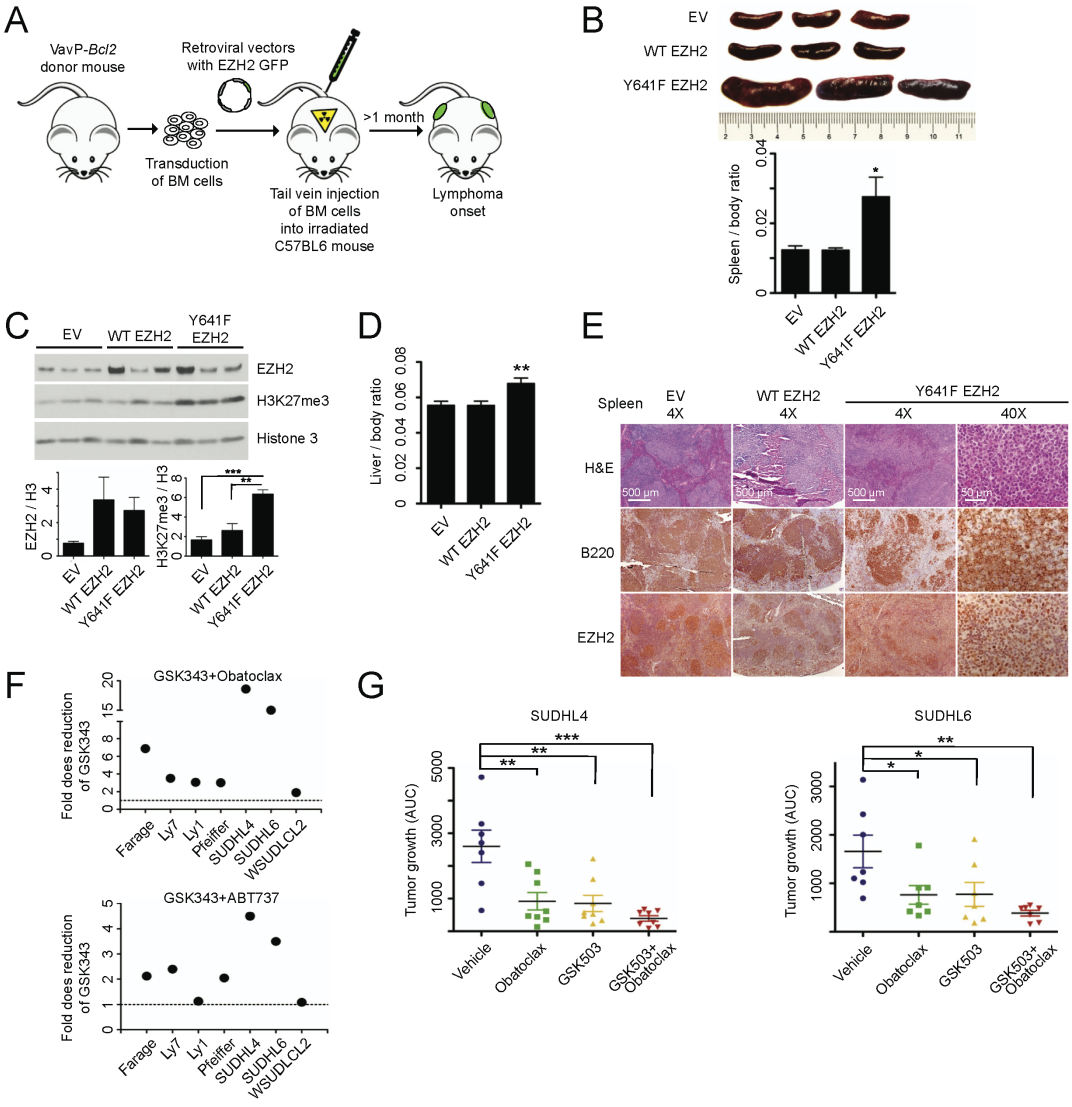

Figure 3.7. EZH2 cooperates with BCL2 to generate GC-derived lymphomas. (A) Bone marrow (BM) transplantation was performed using VavP-*Bcl2* transgenic donor mice. (B) Representative pictures of spleens from mice sacrificed 111 days after transplantation and quantification of the spleen weight (n=10 per group). (C) Immunoblotting from whole cell lysates from splenocytes of transplanted mice. (D) Quantification of the liver weight of transplanted mice (n=10 per group). (E) Splenic tissue from transplanted mice was stained with heamtoxylin and eosin (H&E), B220, and EZH2. (F) Dose reduction plot for Obatoclax and ABT737 at GI_{90} after exposure of cells to increasing concentrations of GSK343 for 6 days. Data represent mean of triplicate experiments. (G) Area under the curve (AUC) of the tumor growth curves for 20 days in SUDHL4 and SUDHL6 xenografted mice treated with vehicle (n=7), Obatoclax (2 mg/kg/day, n=8 in SUDHL4 and n=7 in SUDHL6), GSK503 (150 mg/kg/day; SUDHL4, n=8 and SUDHL6, n=7), or the combination of Obatoclax and GSK503 (SUDHL4, n=8 and SUDHL6, n=7). The p values were calculated by t-test. Values in (B-D) and (H) are mean ± SEM. t test, *p < 0.05, **p < 0.01, ***p < 0.001. Experiments and data were prepared by W. Beguelin.

3.7 EZH2 targeted therapy preferentially affects GCB-DLBCL, but not ABC-DLBCL

EZH2 gain-of-function somatic mutations are restricted to GCB- type DLBCLs[288]. Moreover, whereas EZH2 is a critical mediator of the GC B cell phenotype, which is reflected by the phenotype of GCB-DLBCL, it also represses genes and pathways that drive the phenotype of ABC-DLBCLs (e.g., *IRF4* and NFκB pathway genes). We questioned whether these GCB-specific functions would translate to a specific role for EZH2 in the pathogenesis and therapeutic targeting of GCB versus ABC subtypes of DLBCL. Along these lines, we noted that GCB EZH2 target genes were significantly more repressed in GCB- than in ABC-DLBCLs, although even more significantly repressed when comparing ABC versus mutant EZH2 DLBCLs (**Figure 3.8A**). Gene set enrichment analysis revealed that EZH2 target genes, including GCB bivalent genes were overrepresented among genes upregulated in ABC versus GCB DLBCLs (FDR q < 0.001; **Figure 3.8B**), with enrichment scores even higher when comparing ABC versus mutant EZH2 DLBCLs (FDR q < 0.001; **Figure 3.8C**). Hence, EZH2 target genes, including those with bivalent marks at the GC stage of development, are expressed at relatively higher levels in ABC-DLBCL cells, suggesting that EZH2 does not play a key role in their regulation in this form of lymphoma associated with the transformation of a postgerminal center B cell.

If EZH2 is not critical for the suppression of bivalent genes that drive proliferation once a cell exits the GC, then the biological effects of EZH2 inhibition would be predicted to be significantly different for ABC- versus GCB-DLBCL cells and ABC-DLBCL cells might be predicted to be relatively

insensitive to EZH2 inhibitors. Therefore, we treated a panel of ABC- and GCB-DLBCL cell lines with increasing concentrations of GSK343. Strikingly, the drug concentrations required to inhibit 50% of growth (GI_{50}) for GCB-DLBCL cell lines were in the 0.5–20 mM range, while, for ABC-DLBCL cells, no significant inhibition of cell growth was observed (t-test p = 0.0004; **Figure 3.8D**). Exposure of EZH2 WT GCB-DLBCL cells to a fixed dose of 10 mM GSK343 led to 30%–75% reduction in viable cell number, while EZH2 mutant GCB-DLBCL cells were inhibited 50%–99%, suggesting a trend toward increased sensitivity. By contrast, this dose of drug led to no killing of ABC lymphoma cells (t-test p = 0.0001 versus GCB; **Figure 3.8E**). EZH2 inhibitors completely demethylated H3K27me3 in ABC-DLBCL cells, indicating that resistance is biological and not due to failure of the drug to inhibit its target (**Figure 3.8F**). Collectively, these data suggest that EZH2 inhibitors may be useful for GCB DLBCL either with or without EZH2 mutations but are likely to be ineffective for ABC-type DLBCL.

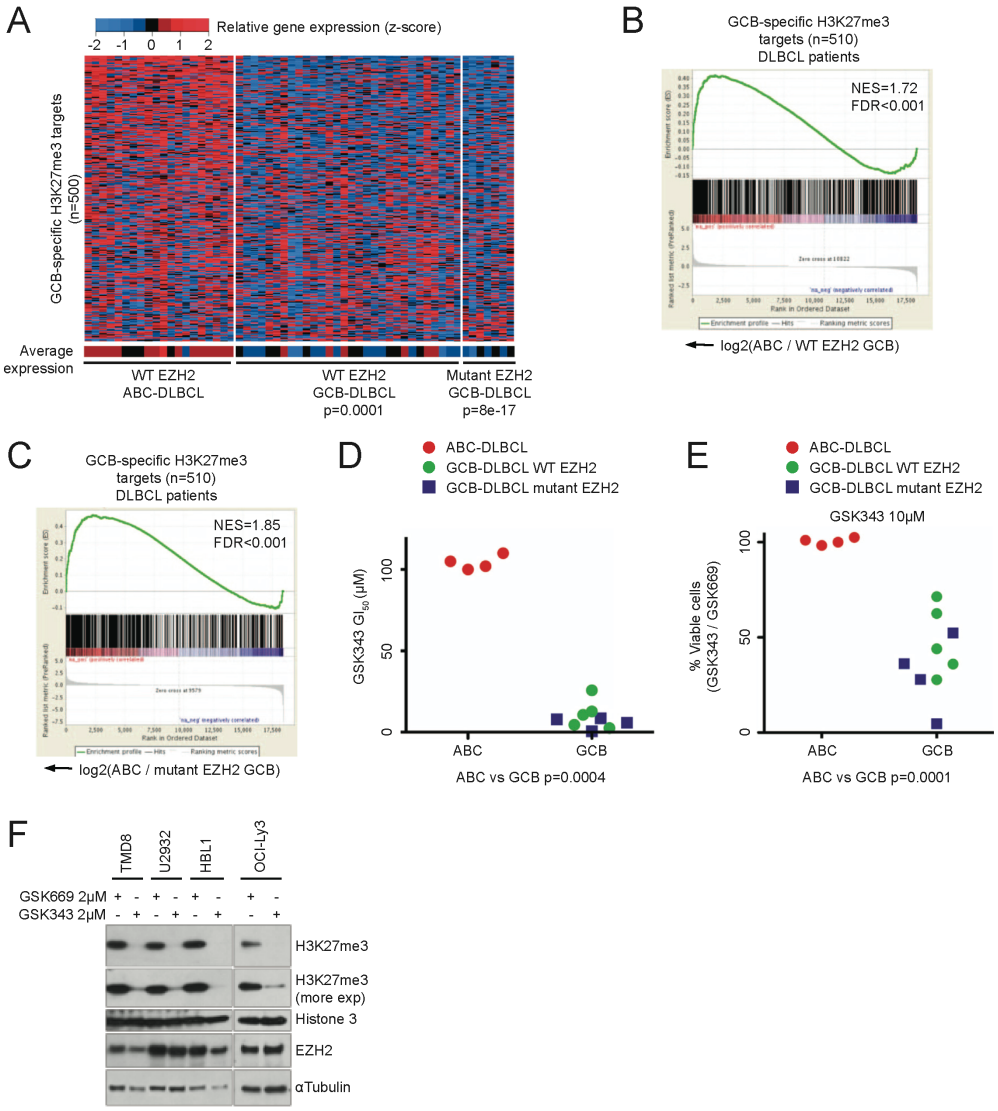

Figure 3.8. EZH2-targeted therapy preferentially affects GCB- but not ABC-DLBCL cells. (A) Heatmap of relative expression of GCB-specific H3K27me3 target genes repressed in mutant EZH2 GCB-DLBCL patient samples. (B-C) GSEA in ABC-DLBCL versus WT EZH2 GCB-DLBCL (B) and mutant GCB-DLBCL (C). (D-E) Four ABC-DLBCL (HBL1, Ly3, U2932, TMD8) and nine GCB-DLBCL cell lines (five WT EZH2: Ly7, Ly19, Farage, Ly18, SUDHL5, and four mutant EZH2: Pfeiffer, WSUDLCL2, SUDHL4, SUDHL6) were exposed to increasing concentrations of GSK343 and GSK669 for 6 days. (D) GI_{50} of GSK343 relative to GSK669. (E) Cell viability at 10 μM GSK343. Data represent mean with 95% confidence interval for duplicate. (F) Immunoblotting with the indicated antibodies was performed in whole cell lysates from ABC-DLBCL cell lines treated with 2 μM GSK343 or control GSK669 for 6 days. Alpha tubuline and histone 3 were used as loading controls. Cell line experiments were performed and data prepared by W. Beguelin.

3.8 EZH2 is required for BCL6 to drive GC hyperplasia

The canonical mechanism by which EZH2 represses transcription is through recruitment of PRC1 complexes. However, GC centroblasts lack canonical core PRC1 components such as PCGF2/MEL18 and PCGF4/BMI1[52], raising the question of how EZH2 coordinates repression in this context. The critical dependency of GC B cells on EZH2 thus provides an opportunity to explore key determinants of its noncanonical and context-specific mechanisms of action. Various other modes of action of EZH2 have been proposed, including potential cooperation with sequence-specific transcription factors[205,289]. Along these lines, the parallels between EZH2 and BCL6 are especially intriguing[290,291] and prompted us to explore whether and how these proteins might cooperate to control transcriptional repression and mediate the GC phenotype.

The similar effects of BCL6 and EZH2 on the GC phenotype prompted us to evaluate whether BCL6 and EZH2 cooperate in the development of GCs. To explore this question, we crossed conditional *Ezh2*$^{fl/fl}$ knockout mice[50] with the Cγ1-cre strain, which expresses CRE recombinase in established GC B cells[282]. These animals were crossed to IμBcl6 mice, which maintain constitutive BCL6 expression in GC B cells[290]. *Ezh2*$^{fl/fl}$;Cγ1-cre, IμBcl6, *Ezh2*$^{fl/fl}$;Cγ1-cre;IμBcl6, and *Ezh2*$^{fl/fl}$ control mice were immunized with SRBC to induce GC formation and sacrificed 10 days later, at which time the GC reaction is at its peak. Notably, deletion of *Ezh2* from GC B cells not only abrogated the BCL6-induced hyperplastic phenotype but also resulted in profound reduction in GC B cells (FAS+/GL7+/B220+, t-test $p < 0.001$; **Figure 3.9A and B**). Immunohistochemical analysis using GC B cell marker peanut

agglutinin further revealed a reduction in the number and size of GCs in *Ezh2*$^{fl/fl}$;Cγ1-cre;IμBcl6 versus *Ezh2*$^{fl/fl}$ controls (t-test $p < 0.05$; **Figure 3.9C**). To determine if the requirement for EZH2 is dependent on its enzymatic function, we next immunized IμBcl6 mice with SRBC followed by daily treatment for 9 days with the EZH2 inhibitor GSK503 or vehicle. We found that GSK503 prevented GC hyperplasia in IμBcl6 mice, manifesting fewer GC B cells by flow cytometry (t-test $p < 0.001$; **Figure 3.9D**) and reduced number and volume of GCs by immunohistochemistry (t-test $p < 0.001$; **Figure 3.9E**). Collectively, these data show that constitutive expression of BCL6 is unable to drive GC hyperplasia in the absence of EZH2 protein or its catalytic activity.

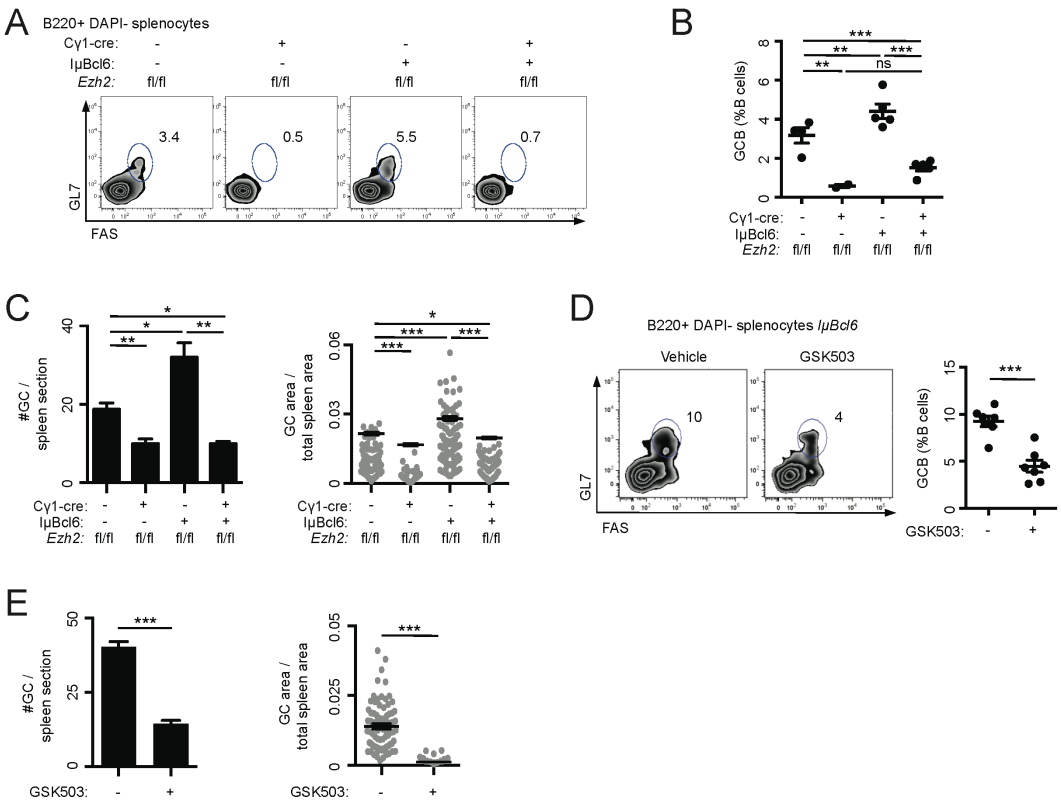

Figure 3.9. EZH2 is required for BCL6 to drive GC hyperplasia. (A-C) *Ezh2*$^{fl/fl}$, *Ezh2*$^{fl/fl}$;Cγ1-cre, IμBcl6, and *Ezh2*$^{fl/fl}$;Cγ1-cre;IμBcl6 mice (n=5 per group) were immunized with SRBC to induce GC formation and sacrificed 10 days later. (A) Flow cytometry plot of one representative mouse spleen per group. The gated area indicates the percentage of GC B cells (GL7+FAS+) within live B cells (B220+DAPI-). (B) Average of GC B populations of each group of mice quantified by flow cytometry. Each dot represents the percenatge of GC B cells within splenic live B cells of one mouse. (C) Quantification of peanut agglutinin (PNA) staining of splenic tissue. The #GC/spleen section (left) indicates the count of all GCs per spleen section. GC area/total area (right) indicates the quantified area of each individual GC divided by the total area of the spleen section. (D-E) IμBcl6 mice were immunized with SRBC, treated daily with GSK503 (150 mg/kg/day, n=7) or vehicle (n=7) and sacrificed 10 days after immunization. (D) Representative flow cytometry plot of splenic GC B cells (left) and quantification (right). (E) Quantification of PCA staining of splenic tissue. Values are shown as means ± SEM. t-test, *p < 0.05, **p < 0.01, ***p < 0.001. Data and figures prepared by W. Beguelin.

3.9 BCL6 is required for mutant EZH2 to drive GC hyperplasia

We performed the reciprocal experiment to determine whether BCL6 is required for hyperactive mutant EZH2^{Y641} to drive lymphoid hyperplasia. Because *Bcl6* constitutive knockout has a complex and lethal phenotype[292],

we first generated conditional *Bcl6*$^{fl/fl}$;Cγ1-cre mice. As expected, conditional deletion of *Bcl6* resulted in profound reduction in GC B cells, underlining that wild-type (WT) EZH2 alone is not sufficient to drive GC formation (**Figure 3.10A, B, C**). To determine whether BCL6 was also required to support the function of hyperactive mutant EZH2^{Y641}, we generated an additional conditional allele, *Ezh2(Y641F)*^{fl}, that expresses *Ezh2*^{Y641F} from the endogenous *Ezh2* locus when activated by CRE. We assessed the GC reaction in the offspring of *Bcl6* conditional KO mice crossed with *Ezh2(Y641F)*$^{fl/WT}$;Cγ1-cre strain. *Bcl6*$^{fl/fl}$;*Ezh2(Y641F)*$^{fl/WT}$;Cγ1-cre mice failed not only to develop EZH2 driven GC hyperplasia, but also exhibited profound reduction of GC B cells (t-test $p < 0.001$ versus mutant EZH2 and $p < 0.05$ versus WT EZH2; **Figure 3.10A and B**). We also found significant reduction in the number and size of GCs as shown by immunohistochemistry (t-test $p < 0.001$; **Figure 3.10C**).

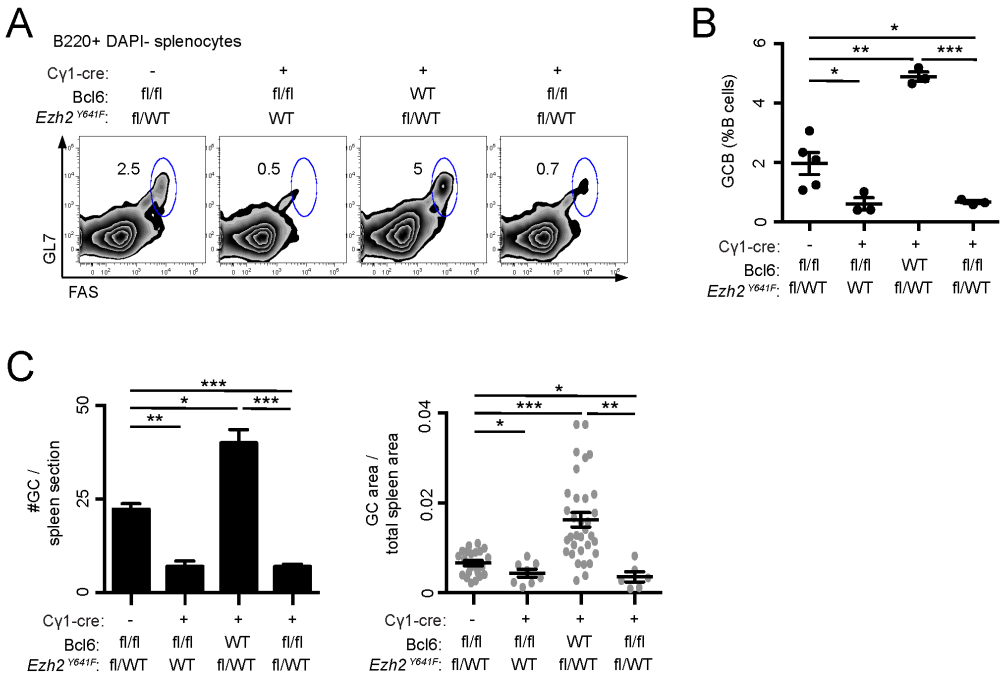

Figure 3.10. BCL6 is required for mutant EZH2 to drive GC hyperplasia.(A-C) *Bcl6*$^{fl/fl}$, *Bcl6*$^{fl/fl}$;Cγ1-cre, *Ezh2(Y641F)*$^{fl/WT}$;Cγ1-cre and *Bcl6*$^{fl/fl}$;*Ezh2(Y641F)*$^{fl/WT}$;Cγ1-cre mice (n=3 to 5 per group) were immunized with SRBC and sacrificed 10 days later. (A) Flow cytometry plot of one representative mouse spleen per group. The gated area indicates the percentage of GC B cells (GL7+FAS+) within live B cells (B220+DAPI-). (B) Average of GC B populations of each group of mice quantified by flow cytometry. Each dot represents the percentage of GC B cells within splenic live B cells of one mouse. (C) Quantification of peanut agglutinin staining of splenic tissue. The #GC/spleen section (left) indicates the count of all GCs per spleen section. GC area/total area (right) indicates the quantified area of each individual GC divided by the total area of the spleen section. Values are shown as mean ± SEM. t-test, *p < 0.05, **p < 0.01, ***p < 0.001. Data and figures prepared by W. Beguelin.

3.10 EZH2 and BCL6/BCOR complexes are both required for repression of key *de novo* GC B cell bivalent promoters

This data suggest a functional dependency between EZH2 and BCL6. EZH2 mediates its effects in GC B cells in part through *de novo* formation of bivalent promoters. BCL6 represses promoters mainly by recruiting the co-repressor

protein BCOR[41]. To explore potential mechanistic links between EZH2 and BCL6, we examined the genomic distribution of H3K27me3, H3K4me3, BCL6, and BCOR in purified primary human NB cells and GC B cells using ChIPseq data[41]. We observed significant overlap of GC *de novo* bivalent promoters with BCL6 and BCOR (hypergeometric test, $p = 2.9 \times 10^{-20}$; **Figure 3.11A**). In contrast, BCL6 and BCOR were excluded from monovalent H3K27me3 genes (depletion $p < 1 \times 10^{-76}$; **Figure 3.11B**). BCL6/BCOR-occupied GC *de novo* bivalent genes were significantly enriched for pathways involved in GC exit and terminal differentiation, including genes induced by IRF4 and T cell cytokines, genes highly expressed in ABC-DLBCL signature compared with GCB-DLBCL (which includes GC exit genes), and genes associated with immune responses (**Figure 3.11C**). Among BCL6-BCOR-occupied *de novo* bivalent genes were key proliferation checkpoint (*CDKN1A*, *CDKN1B*) and B cell differentiation (*PRDM1*, *IRF4*) genes (as exemplified in **Figure 3.11D and E**). Bivalent genes without BCL6/BCOR complexes were not preferentially linked to these pathways. We next compared RNAseq gene expression profiles of NB versus GC B cells and found that *de novo* bivalent genes bound by BCL6 and BCOR at their promoters are significantly more repressed in GC B compared with NB cells (GSEA FDR, $q < 0.001$; **Figure 3.11F**). To determine whether these BCL6-BCOR-bound *de novo* bivalent genes were actively repressed by EZH2 and BCL6, we examined RNAseq profiles of GC-derived DLBCL cell lines treated with an EZH2 inhibitor or a BCL6 inhibitor that disrupts BCL6-BCOR interaction; or EZH2 small hairpin RNAs (shRNAs) or BCL6 small interfering RNAs (siRNAs)[41]. In all cases, we observed significant de-repression of BCL6-BCOR *de novo* bivalent genes. In contrast, *de novo* bivalent genes lacking BCL6/BCOR were only de-repressed by EZH2

shRNA or inhibitors (**Figure 3.11G**). These data suggest functional cooperation between EZH2 and BCL6, specifically at genes where BCL6 recruits BCOR, since EZH2 alone or BCL6 alone is not sufficient to maintain repression of these bivalent genes.

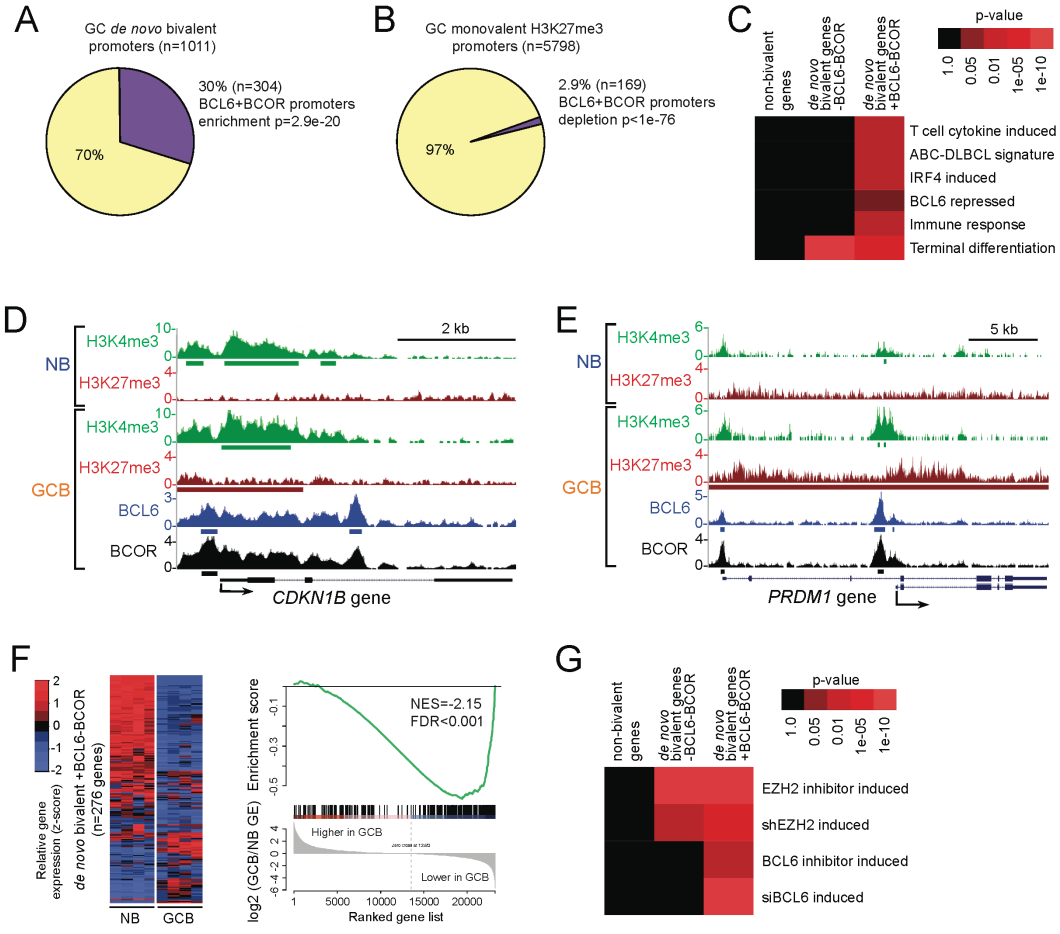

Figure 3.11. EZH2 and BCL6-BCOR complexes are both required to repress key *de novo* GC B cell bivalent promoters. (A-B) Percentage and number of *de novo* bivalent promoters (A) and H3K27me3 monovalent promoters (B) overlapping ChIPseq peaks of BCL6 and BCOR in GC B cells. (C) Heatmap of over-represented gene signatures among genes with BCL6+BCOR-occupied *de novo* bivalent promoters compared with non-bivalent genes and *de novo* bivalent genes without BCL6 or BCOR. Enrichment measured using hypergeometric p-values. (D-E) *CDKN1B* (D) and *PRDM1* (E) gene loci showing H3K4me3, H3K27me3, BCL6, and BCOR CHIPseq read density and enriched loci (bars) in NB and GCB. (F) Heatmap of the relative gene expression levels and GSEA of *de novo* bivalent genes with BCL6+BCOR in 4 NB and 4 GCB samples. NES, normalized enrichment score; FDR, false discovery rate. (G) Heatmap of enrichment for *de novo* bivalent genes with and without BCL6+BCOR occupied promoters among genes induced by EZH2 inhibitors (2 μM GSK343 for 7 days and 25 μM FX1 for 12h) or shRNAs for EZH2 (7 days) or siRNA for BCL6 (2 days) in OCI-Ly1, OCI-Ly7, SUDHL5, SUCHL6, Farage, WSUDLCL2, and Pfeiffer DLBCL cell lines. Enrichment measured using hypergeometric p-values.

Transcriptional repression of H3K27me3-marked chromatin is mediated by PRC1 complexes. However, we found that genes for canonical PRC1 components *BMI1* (*PCGF4*), *PHC1*, and *PHC3* are repressed and downregulated in GC B cells compared with NB cells (**Figure 3.12A**). The canonical PRC1 component PCGF2 (MEL18) was absent in both GC B and NB cells (data not shown). However, BCOR forms an alternative noncanonical complex with certain PRC1 subunits[216,293]. Noncanonical PRC1 genes *BCOR*, *PCGF1*, *KDM2B*, *SKP1*, and *USP7* are upregulated in GC B cells, similar to PRC2 (**Figure 3.12A**). *BMI1* downregulation and BCOR upregulation in GC B cells were confirmed by RT-qPCR (**Figure 3.12B**). Immunoblot analysis likewise revealed that GC B cells express higher protein levels of BCOR, KDM2B, and PCGF1 than NB cells (**Figure 3.12C**). Remarkably, the core canonical PRC1 component BMI1 is among the *de novo* bivalent genes bound and repressed by BCL6-BCOR complexes (**Figure 3.12D**). The promoter of PCGF2, on the other hand, is marked by only H3K27me3 in both GCB and NB cells, suggesting it is highly repressed (**Figure 3.12D**). Taken together, these data indicate that the noncanonical PRC1-BCOR complex may represent the dominant PRC1 in GC B cells.

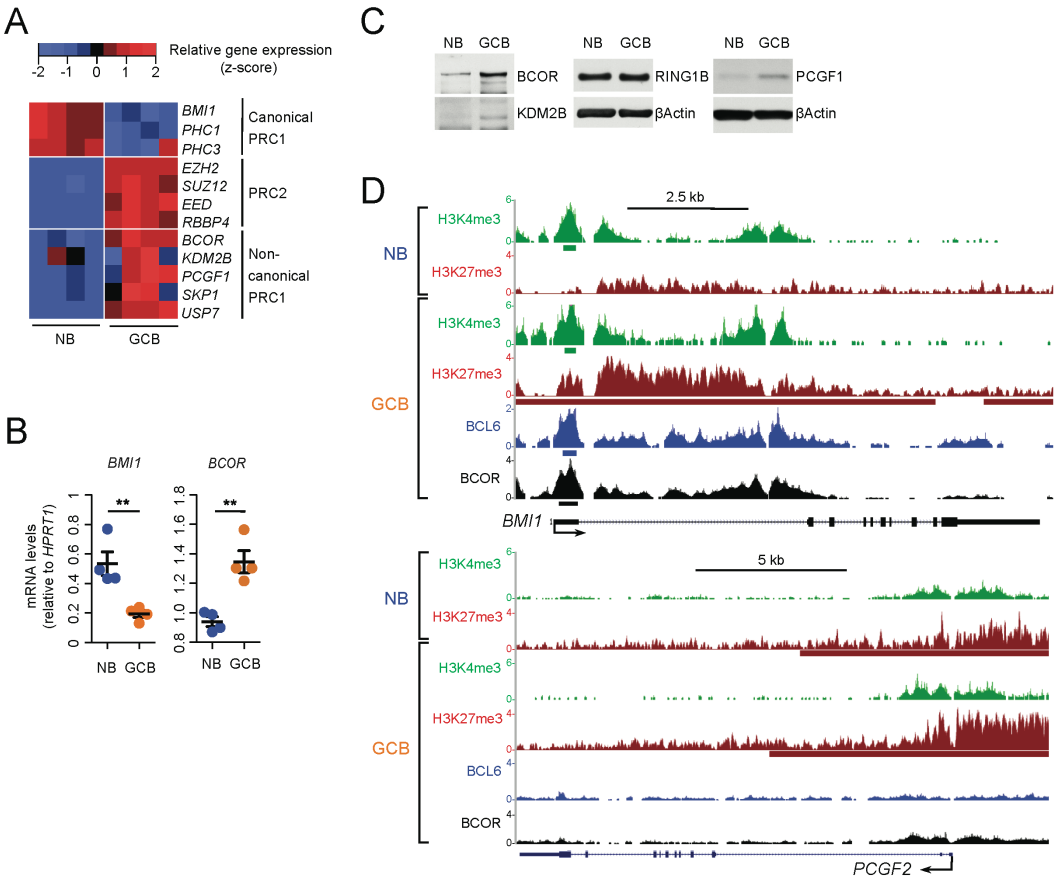

Figure 3.12. Noncanonical PRC1-BCOR complex may represent the dominant PRC1 in GC B cells. (A) Heatmap of the gene expression level of canonical and noncanonical PRC1 and PRC2 components in n=4 NB and n=4 GCB samples. (B) RT-qPCR of *BMI1* and *BCOR* in NB and GCB samples. (C) Immunoblotting of whole lysates from NB and GCB samples. (D) *BMI1* (top) and *PCGF2* (bottom) gene loci showing H3K4me3, H3K27me3, BCL6, and BCOR ChIPseq read density in naive B cells (NB) and GC B cells (GCB). Green, red, blue, and black bars: H3K4me3 peaks, H3K27me3 broad domains, BCL6 peaks, and BCOR peaks, respectively. t-test *p < 0.05, **p < 0.01, ***p <0.001. RT-qPCR and immunoblotting were performed and data prepared by W. Beguelin.

3.11 Mutant EZH2 fails to induce GC hyperplasia in absence of BCL6/BCOR

The above data suggest that BCOR, like BCL6, may be required for the transcriptional and biological effects of EZH2. Thus, to determine whether

BCOR is required for GC formation, we used a conditional *Bcor* allele, $Bcor^{fl}$ together with the Cγ1-cre allele. $Bcor^{fl/Y}$;Cγ1-cre mice failed to form GCs after immunization, similar to the case of *Ezh2* or *Bcl6* deletion (**Fiugre 3.13A**). To evaluate if mutant EZH2 can drive GC formation or hyperplasia in the absence of BCOR, we performed immunization experiments in mice bred for simultaneous conditional knockout of *Bcor* and conditional knockin of $Ezh2(Y641F)^{fl/WT}$ with the Cγ1-cre allele. *Bcor* deletion in $Bcor^{fl/Y}$;$Ezh2(Y641F)^{fl/WT}$;Cγ1-cre mice resulted in marked depletion in the number of GC B cells (t-test $p < 0.001$; **Figure 3.13B**) and significant reduction in the number and volume of GCs by immunohistochemistry (t-test $p < 0.001$; **Figure 3.13C**)

These data suggest that BCL6 and BCOR are each required for the actions of EZH2 in established GC B cells, but do not address whether it is the interaction between BCL6 and BCOR that mediates this effect. To address this point, we used a *Bcl6* allele that encodes a mutant BCL6 protein unable to bind to BCOR ($Bcl6^{BTBmut}$)[43]. As previously reported, homozygous $Bcl6^{BTBmut}$ mice were unable to form GCs (**Figure 3.13D**). $Bcl6^{BTBmut}$ homozygous mice crossed with $Ezh2(Y641F)^{fl/WT}$;Cγ1-cre rescued the GC hyperplasia phenotype and again abrogated GC formation, as shown by flow cytometry (t-test $p < 0.001$; **Figure 3.13E**) and immunohistochemistry (t-test $p < 0.001$; **Figure 3.13F**). Collectively, these results demonstrate that WT and gain-of-function mutant EZH2 require a functional BCL6-BCOR complex to drive formation of GCs and GC hyperplasia, respectively.

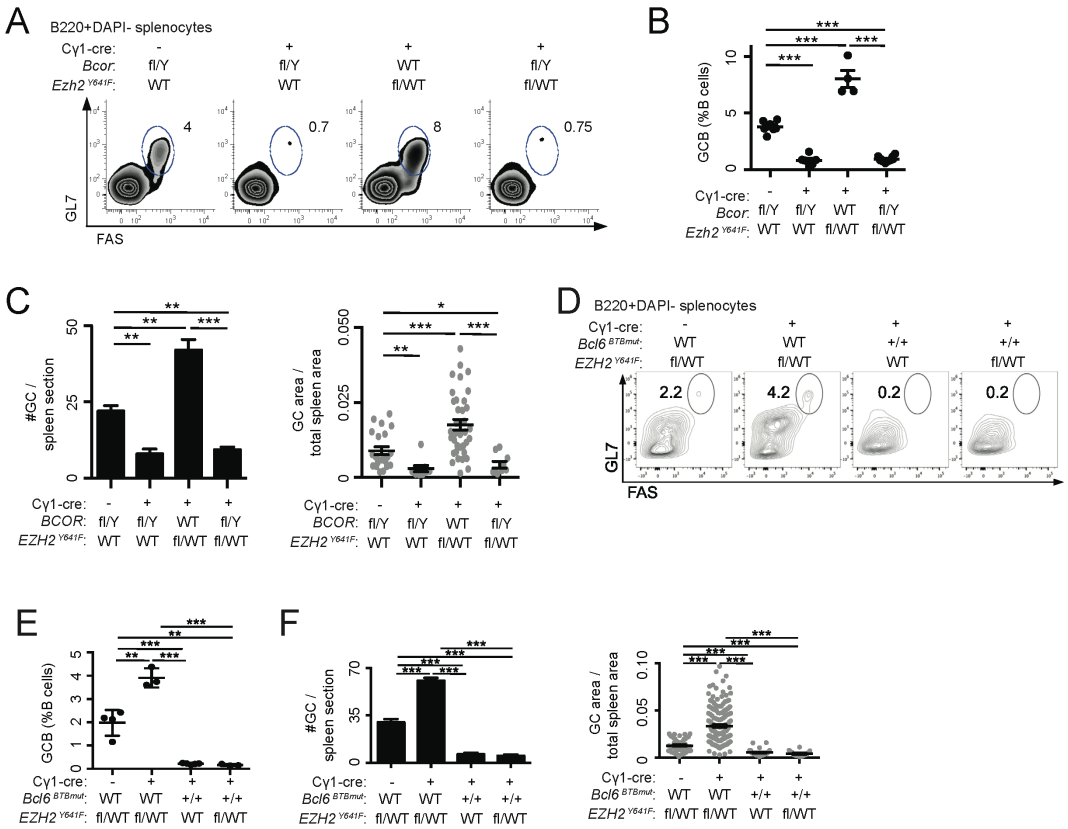

Figure 3.13. Mutant EZH2 fails to induce GC hyperplasia in absence of BCOR and BCL6. (A-C) *Bcor*$^{fl/Y}$, *Bcor*$^{fl/Y}$;Cγ1-cre, *Ezh2(Y641)*$^{fl/WT}$;Cγ1-cre, and *Bcor*$^{fl/Y}$;*Ezh2(Y641)*$^{fl/WT}$;Cγ1-cre mice (n=4-6 per group) were immunized with SRBC and sacrificed 10 days later. As *Bcor* is present on the X chromosome, male mice only have one floxed *Bcor* allele. Y indicates Y chromosome. (A) Representative flow cytometry plot of splenic GC B cells. (B) Quantification of GC B cells by flow cytometry. (C) Quantification of GC number per spleen section (left) and GC area per total spleen area (right) based on peanut agglutinin (PNA) staining. (D-F) *Ezh2(Y641)*$^{fl/WT}$, *Ezh2(Y641)*$^{fl/WT}$;Cγ1-cre, *Bcl6*^{BTBmut};Cγ1-cre, *and Bcl6*^{BTBmut};*Ezh2(Y641)*$^{fl/WT}$;Cγ1-cre mice (n=4 per group) were immunized with SRBC and sacrificed 10 days later. (D) Representative flow cytometry plot of splenic B cells. (E) Quantification of GC B cells by flow cytometry. (F) Quantifications of GC number (left) and area (right) based on PNA staining. Experiments performed and data prepared by W. Beguelin.

3.12 PRC1-BCOR complex requires both PRC2 and BCL6 for stable association and repression of bivalent promoters

We next evaluated the mechanism through which EZH2, BCL6, and BCOR cooperate to mediate the GC phenotype. First, we examined whether BCL6

could interact with EZH2. Using sequential ChIP (ChIP re-ChIP), we evaluated whether these proteins were bound to the same loci and found that EZH2 is co-recruited at the same bivalent promoters as BCOR and BCL6 as shown for the *CDKN1B* and *PRDM1* loci (**Figure 3.14A**). However, we were unable to detect any interaction either with endogenous or transfected proteins (data not shown). Hence, BCL6-BCOR and PRC2 complexes co-localize on chromatin without direct BCL6-EZH2 contact.

To determine whether PRC2 and BCL6-BCOR functionally cooperate on chromatin, we treated GC-derived DLBCL cells with the EZH2 inhibitor GSK343 or the inactive compound GSK669 and then evaluated recruitment of PRC2 components EZH2, EED, and SUZ12; PRC1-BCOR complex components BCOR, RING1B, KDM2B, and PCGF1; BCL6; and the PRC2 histone mark H3K27me3, and the RING1B mark H2AK119ub. We performed qChIP for these proteins in four independent GC-derived DLBCL cell lines at six key bivalent promoters (*CDKN1B, PRDM1, IRF4, CDKN1A, ARID3A*, and *ARID3B*), as well as a negative control region. EZH2 inhibitor caused a significant reduction in both recruitment of PRC2 and BCOR complex components, along with concordant loss of H3K27me3 and H2AK119ub (**Figure 3.14B**). In contrast, BCL6 binding was unaffected. This suggests that, even though EZH2 does not directly interact with the PRC1-BCOR complex, stable association of PRC1-BCOR complex with chromatin still requires PRC2 activity. BCL6 occupancy alone is not sufficient to maintain maximal PRC1-BCOR recruitment or transcriptional repression of these genes, and BCL6 recruitment does not require H3K27me3. In reciprocal experiments, we treated the same cell lines with FX1 to block the interaction

between BCL6 and BCOR. In this case, qChIP revealed loss of PRC1-BCOR complex recruitment with no effect on PRC2 occupancy (**Figure 3.14C**). We observed depletion of both H2AK119ub and H3K27me3, consistent with loss of BCOR complex as well as impairment of PRC2 function. Collectively, these data suggest a model whereby BCL6 and EZH2 must cooperate to mediate the stable recruitment of the noncanonical PRC1-BCOR complex to bivalent promoters in GC B cells. Neither BCL6 nor PRC2 alone are sufficient to optimally tether this complex or fully repress expression of these target genes.

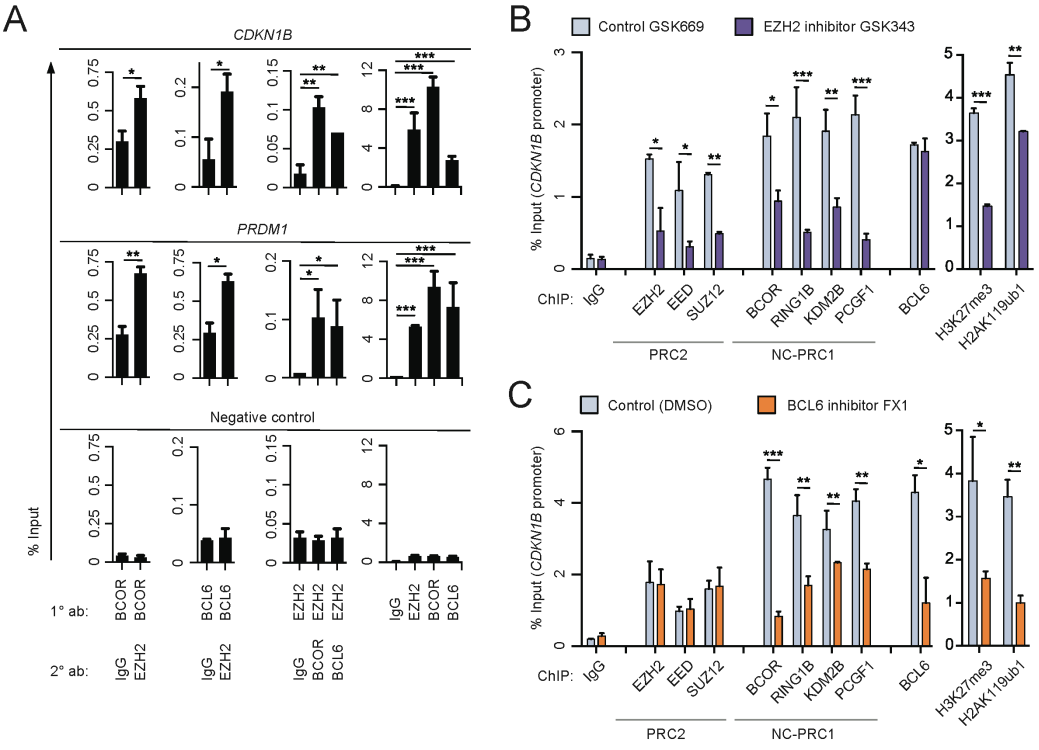

Figure 3.14. PRC1-BCOR complex requires both PRC2 and BCL6 for stable association with and repression of bivalent promoters. (A) ChIP re-ChIP in the bivalent promoters of CDKN1B and PRDM1 in OCI-Ly7 cells using the indicated antibodies for the first ChIP (1° ab) and the sequential ChIP (2° ab). As negative control, qPCR was performed using primers for a region in chromosome 6 where no BCOR, EZH2, or BCL6 enrichment was found by CHIPseq read density in GC B cells. (B-C) qChIP in CDKN1B promoter of SUDHL6 cells treated with (B) 2 μM GSK343 or control GSK669 for 72hr, or (C) 25 μM BCL6 inhibitor FX1 or vehicle for 6hr. NC-PRC1, noncanonical PRC1. Values shown as mean of triplicate or quadruplicates ± SD. t-test, *p < 0.05, **p < 0.01, ***p < 0.001. Experiments performed and data prepared by W. Beguelin.

3.13 CBX8-mediated recruitment of BCOR complex to H3K27me3-marked bivalent genes is required for EZH2 function

BCOR recruitment occurs through direct binding of BCOR to the BCL6 BTB domain[294-296]. However, it is not known how BCOR is recruited to bivalent promoter regions through PRC2. Canonical PRC1 complexes contain chromobox homolog (CBX) histone reader proteins that bind to H3K27me3. Hence, we wondered whether CBX proteins might also mediate the PRC2-dependent recruitment of BCOR complex to bivalent promoters in the GC B cell context. We examined RNAseq gene expression profiles to identify CBX family proteins potentially relevant to the GC B cell context (**Figure 3.15A**). Among these, CBX8 was the most differentially upregulated CBX family member in GC B cells. We confirmed CBX8 upregulation in purified GC B cells and NB cells using qPCR and immunoblots (**Figure 3.15B and C**).

To determine whether CBX8 might form part of the PRC1-BCOR complex, we performed BCOR tandem affinity purifications followed by mass spectrometry in HEK293 cells and identified CBX8 as a co-purifying protein (data not shown). PCGF1 was also reported enriched in CBX8 purifications in HeLa cells[297] and differentiating embryonic stem cells[298], and CBX8 was also associated to a KDM2B-BCOR complex[299]. To confirm these results, we developed an insect cell reconstitution system for BCOR complexes and showed that CBX8 can be incorporated into BCOR complex and

immunoprecipitated with BCOR complex components (**Figure 3.15D**). A previously described co-structure of CBX7 and RING1B and associated mutational analysis[300] allowed us to identify residues that might inhibit the CBX8-RING1B interaction. Mutation of these residues on RING1B (Y262A) or CBX8 (I375D) resulted in failure to incorporate CBX8 into the BCOR complex (**Figure 3.15D**). To further investigate whether CBX8 associates with BCOR in GC B cells, we performed co-immunoprecipitation experiments for the endogenous CBX8 and BCOR proteins in two DLBCL cell lines. CBX8 antibody enriched for BCOR, as did the reciprocal experiment with BCOR and CBX8 immunoblot (**Figurer 3.15E**). Most importantly, we confirmed endogenous CBX8 association with BCOR in purified GC B cells from human tonsils (**Figure 3.15F**). CBX8 is thus an integral component of the PRC1-BCOR complex in GC B cells in a RING1B- dependent manner.

To evaluate the functional relevance of CBX8, we first examined whether it was recruited to bivalent promoters. CBX8 binding was observed using qChIP assays in four DLBCL cell lines at the same six bivalent promoters evaluated earlier, but not at a negative control locus (**Figure 3.15G**). To determine if CBX8 binding is linked to H3K27me3, we treated these cell lines with the EZH2 inhibitor GSK343 or GSK669 control. In all cases, EZH2 inhibition resulted in profound loss of CBX8 recruitment (**Figure 3.15G**). Next, to determine whether CBX8 was necessary for BCOR recruitment, we depleted CBX8 from GC-derived DLBCL cells using two independent shRNA or a control shRNA. Both shRNAs induced significant reduction of CBX8 protein

(**Figure 3.15H**) and resulted in decreased BCOR recruitment to bivalent promoters, as well as reduction of the PRC1-BCOR complex catalyzed H2AK119ub histone mark (**Figure 3.15I**).

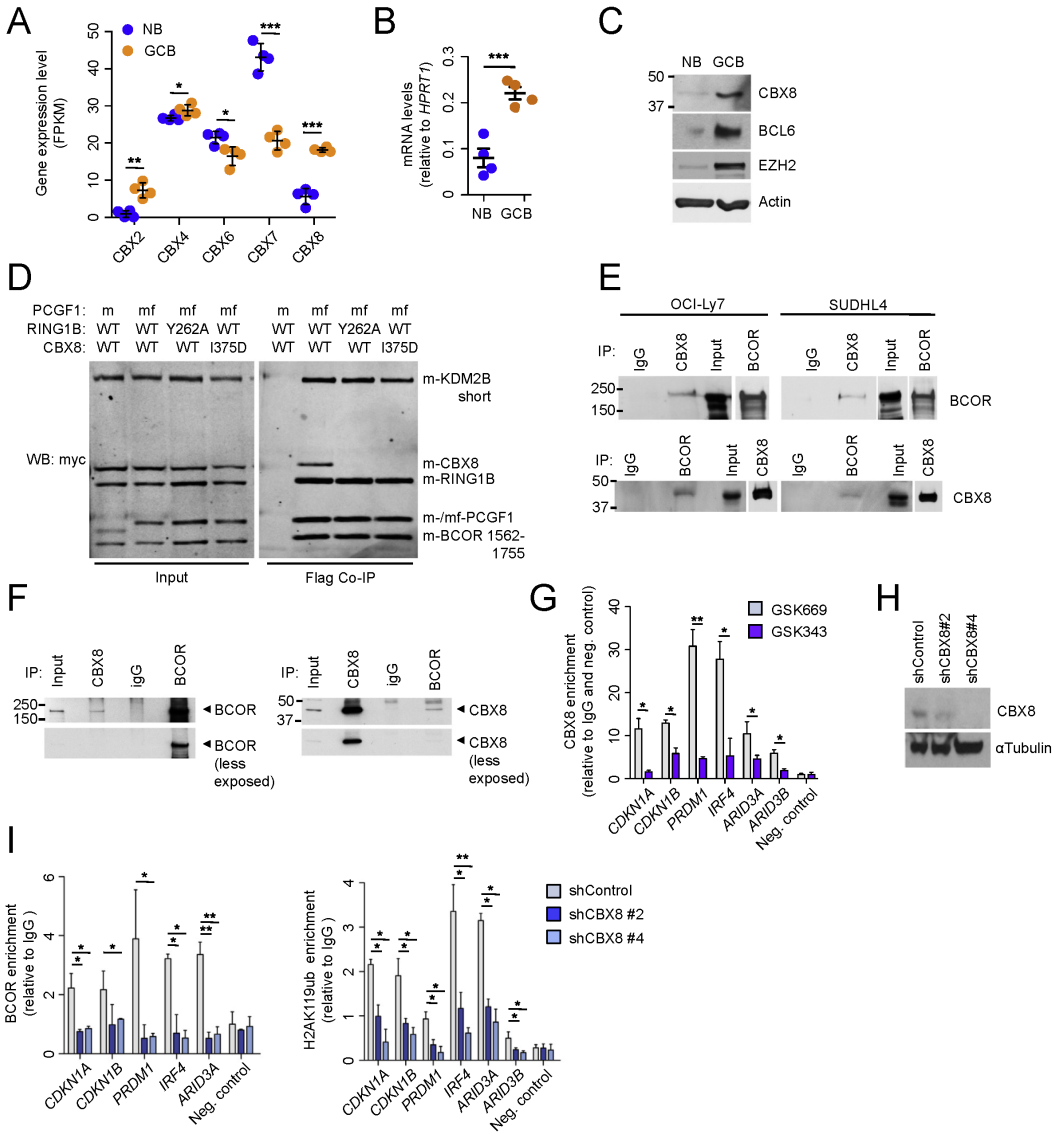

Figure 3.15. CBX8 recruits BCOR complex to H3K27me3 marked bivalent genes and is required for the biological actions of EZH2. (A) Expression level of CBX genes in n=4 NB and n=4 GCB samples. Values are means ± SEM. (B) RT-qPCR of CBX8 in n=4 NB and n=4 GCB samples. Values are means ± SEM. (C) Immunoblotting of whole-cell lysates from NB and GCB cell samples. (D) Myc immublotting of Flag IPs using whole-cell extracts from Sf9 insect cells transfected with expression vectors for the indicated proteins. m, amino terminal myc epitope tag; f, flag tag. (E) Immunoblotting of immunoprecipitation (IP) using whole-cell lysates from indicated cell lines. (F) Immunoblotting of IP using whole-cell lysates from primary human tonsillar GCB. (G) CBX8 qPCR in OCI-Ly7 cells treated with 2 μM GSK343 or GSK669 for 72hr. (H) Immunoblotting of whole-cell lysates from OCI-Ly7 cells expressing two independent CBX8 shRNAs or control. (I) BCOR and H2AK119 qChIP in cells from (H). Values in (G) and (I) shown as means of triplicates or quadruplicates ± SD. t-test, *p < 0.05, **p < 0.01, ***p < 0.001. Experiments performed and data prepared by W. Beguelin.

3.14 CBX8 phenocopies EZH2 loss-of-function *in vitro* and *in vivo*

We observed that CBX8 was required for repression of key EZH2 bivalent genes including *CDKN1A*, *PRDM1*, and *IRF4*, since CBX8 shRNA resulted in their de-repression (**Figure 3.16A**). To determine whether loss of CBX8 mimics the effects of loss of EZH2, we examined the phenotype of DLBCL cells after CBX8 depletion relative to shRNA control in DLBCL cell lines. In all cases, CBX8 loss results in significant growth suppression (t-test $p < 0.001$; **Figure 3.16B**). We also observed induction of the plasma cell genes *PRDM1*, *TP73*, and *CD138* by qPCR (**Figure 3.16A**). Plasma cell differentiation was further demonstrated by decreased B cell surface marker CD20, increased plasma cell marker CD138, and surface expression of immunoglobulin heavy and light chains using flow cytometry (**Figure 3.16C and D**). Morphologically, the DLBCL cells exhibited the characteristic features of plasma cell differentiation, including basophilic cytoplasm, eccentric more condensed nuclei, and prominent Golgi apparatus (**Figure 3.16E**).

These data suggested that CBX8 is a required component of the PRC1-BCOR complex in GC B cells. To confirm whether this is truly the case, we generated *Cbx8*$^{fl/fl}$;Cγ1-cre mice[301] and performed immunization experiments to induce GCs. *Cbx8* deletion resulted in marked depletion of GC B cells (t-test $p < 0.001$; **Figure 3.16F and G**) and significant reduction in the number and volume of GCs (t-test $p < 0.001$; **Figure 3.16H and I**). Together these results indicate that CBX8 is the component of the PRC1-BCOR complex that tethers the complex to chromatin downstream of the actions of EZH2, thus

enabling repression of bivalent promoters and mediating the actions of EZH2 on GC formation.

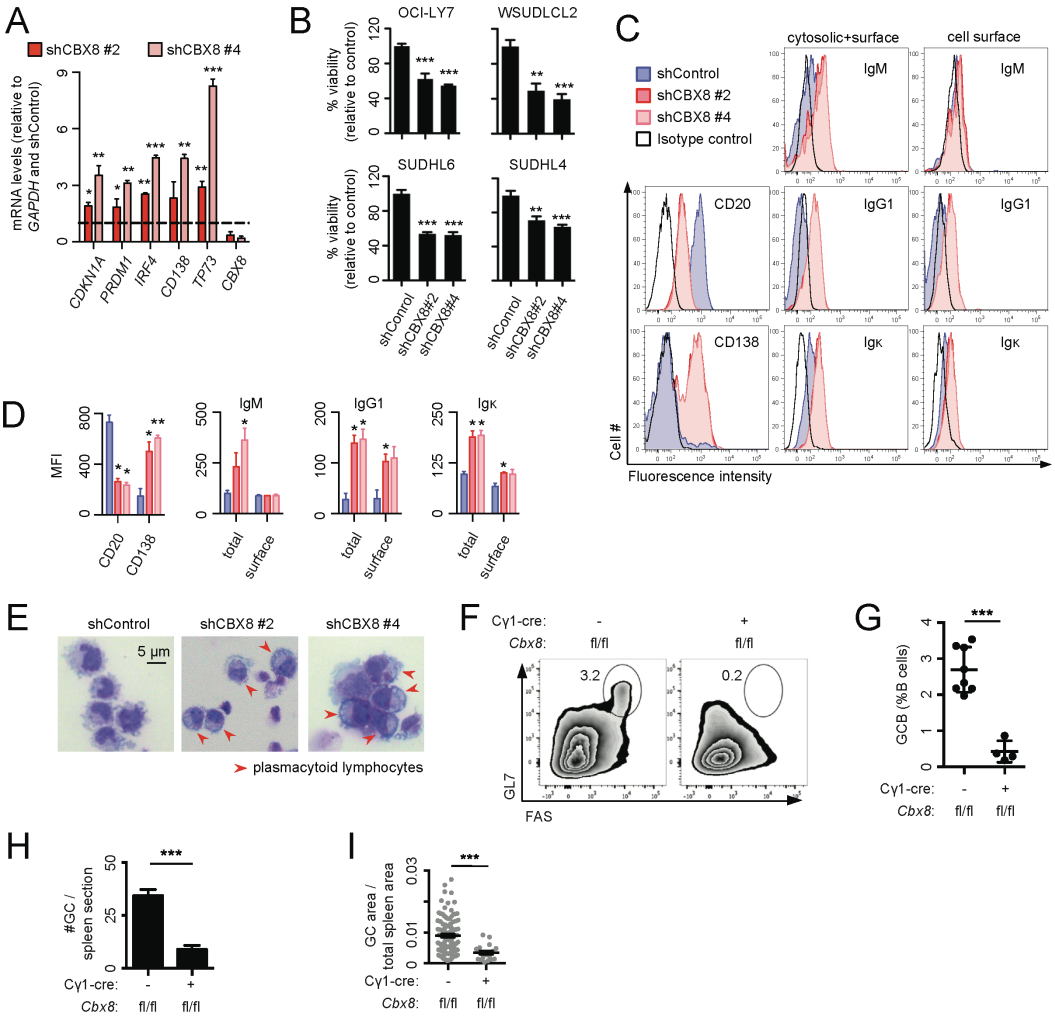

Figure 3.16. CBX8 loss phenocopies EZH2 loss-of-function *in vitro* and *in vivo*. (A) RT-qPCR of the indicated mRNAs from OCI-Ly7 cells expressing two independent CBX8 shRNAs or control used in Figures 3.13H and 3.13I. p-Values shown are compared with the shControl. (B) Viability of OCI-Ly7, WSUDLCL2, SUDHL6, and SUDHL4 cells were evaluated 7 days after shCBX8 infection using cell titer blue. (C) WSUDLCL2 cells were infected with two independent CBX8 shRNAs or control for 5 days, and CD20, CD138, and immunoglobulin (Ig) expression levels were examined by flow cytometry. (D) Quantification of mean fluorescence intensity (MFI) from (C) (n=3). (E) Representative images of WSUDLCL2 cells infected as in (C). (F–I) *Cbx8*$^{fl/fl}$ (n=8) and *Cbx8*$^{fl/fl}$;Cγ1-cre mice (n=4) were immunized with SRBC and sacrificed 10 days later. (F) Representative flow cytometry plot of splenic GC B cells (G) Quantification of GC B cells by flow cytometry as in Figure 3.8B. (H-I) Quantification of GC number (H) and area (I) based on peanut agglutinin staining. Values in (A) and (B) are shown as means of triplicates or quadruplicates ± SD. Values in (D), (G-I) are means ± SEM. t-test, *p < 0.05, **p < 0.01, ***p < 0.001. Experiments performed and data prepared by W. Beguelin.

3.15 Mutant EZH2 and constitutive BCL6 cooperate to induce lymphomagenesis

Both $EZH2^{Y641}$ mutation and BCL6 constitutive expression induce GC hyperplasia. Having established the mechanistic basis for cooperation and interdependence of EZH2 and BCL6 in repressing critical GC B cell genes, we next examined whether their combined gain-of-function alleles might cooperate to drive the transformation of GC B cells to form DLBCLs. BCL6 is constitutively expressed in the GCB-DLBCLs in which EZH2 somatic mutations occur. Therefore, we crossed IμBcl6 with *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre mice to engineer BCL6 constitutive expression and mutant EZH2 activity in GC B cells. The breeding resulted in four different allele combinations: IμBcl6 alone, *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre alone, *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre;IμBcl6, and control littermates. Bone marrow of these four groups was next transplanted into lethally irradiated recipient mice (**Figure 3.17A**). Animals were immunized with SRBC every 3 weeks to ensure continuous formation of GCs and were observed for survival. *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre;IμBcl6 mice showed significant acceleration of lethality compared with IμBcl6 and *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre mice ($p = 0.001$; **Figure 3.17B**). A second cohort of mice (control n = 4, *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre n=4, IμBcl6 n=5, *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre;IμBcl6 n=12) was euthanized 223 days after transplant for more detailed phenotypic analysis. Macroscopic examination of spleens and lymph nodes showed massive splenomegaly in *Ezh2(Y641F)$^{fl/WT}$;*Cγ1-cre;IμBcl6 versus the other groups (t-test $p < 0.05$;

Figure 3.17C). Histopathologic examination indicated that, whereas all *Ezh2(Y641F)*$^{fl/WT}$; Cγ1-cre;IμBcl6 mice had developed a B cell lymphoma (FL or DLBCL n=10/12) or pre-neoplastic lymphoid hyperplasia (n=2/12), by contrast, none of the other groups showed either phenotype at this time point (**Figure 3.17D**). These data suggest that mutant EZH2 and BCL6 cooperate to induce and accelerate the development of DLBCL-like disease.

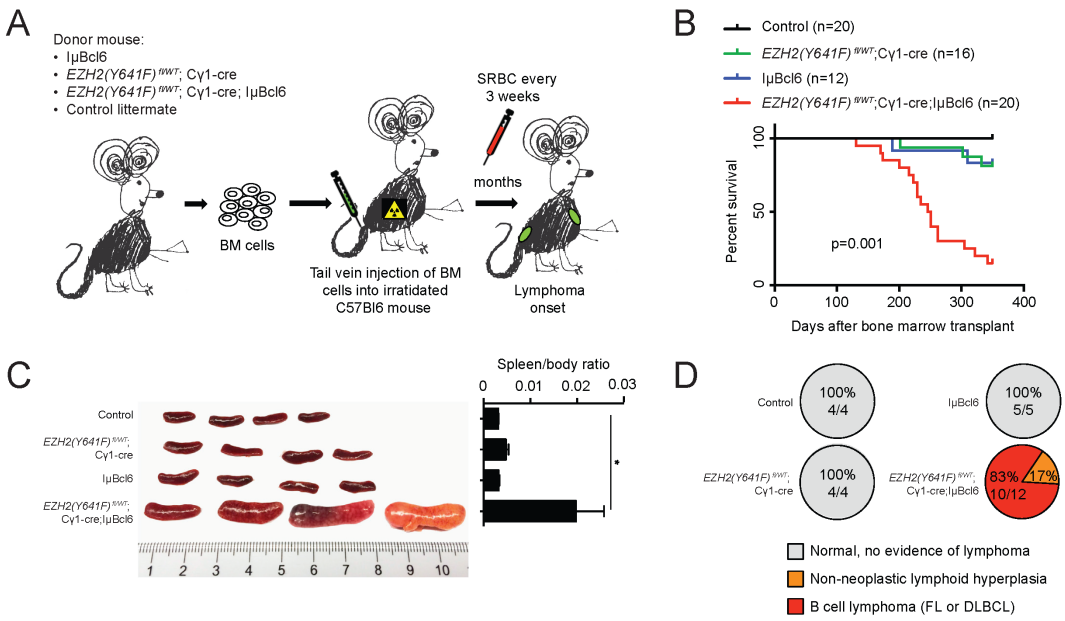

Figure 3.17. Mutant EZH2 and constitutive BCL6 cooperate to induce lymphomagenesis. (A) Bone marrow transplantation was performed using IμBcl6, *EZH2(Y641F)*$^{fl/WT}$;Cγ1-cre, *EZH2(Y641F)*$^{fl/WT}$; Cγ1-cre;IμBcl6, and control negative littermate donor mice. BM, bone marrow. (B) Survival curve of transplanted mice. (C) Representative pictures of spleens from mice sacrificed 223 days after transplantation and quantification of spleen weight (control, n=4; *EZH2(Y641F)*$^{fl/WT}$;Cγ1-cre, n=4; IμBcl6, n=5; *EZH2(Y641F)*$^{fl/WT}$;Cγ1-cre;IμBcl6, n=12). (D) Percentage and number of mice from cohort used in (C) that did or did not develop lymphoma. Values are mean ± SEM. t-test, *p < 0.05, **p < 0.01, ***p < 0.001. Experiments were performed and data prepared by W. Beguelin. Artwork generated by Heidi Beguelin.

3.16 BCL6 and EZH2 Inhibitors Cooperate to Kill DLBCLs and Suppress Tumor Xenografts and Primary Human DLBCL Growth

Both BCL6 and EZH2 inhibitors are proposed as potential therapies for patients with B cell lymphomas. BCL6 and EZH2 are both constitutively expressed and required to maintain the growth of GCB-type DLBCLs, regardless of whether EZH2 is mutated. Given that BCL6 and EZH2 cooperate to induce maximal repression of their key target promoters, we asked whether simultaneously targeting both proteins might yield enhanced anti-lymphoma activity. We exposed a panel of GCB-DLBCL cells to increasing concentrations of GSK343 in combination with FX1 (along with their respective controls). In almost every case, the concentration of GSK343 required to yield 50% growth inhibition was reduced when cells were concomitantly treated with FX1 (**Figure 3.18A**). We also observed that EZH2-BCL6 bivalent target genes were significantly further de-repressed by treatment with the combination GSK343 and FX1 versus the single drugs (t-test $p < 0.01$; **Figure 3.18B**). Hence, targeting both arms of PRC1-BCOR tethering through EZH2 and BCL6 results in more powerful target gene de-repression with corresponding greater biological activity against lymphoma cells.

To determine the impact of combinatorial BCL6-EZH2 targeted therapy in a preclinical model, we evaluated the action of the EZH2 inhibitor GSK126 and FX1 alone or in combination at submaximal doses in mice bearing established

human DLBCL cell line (SUDHL6 and WSU-DLCL2) xenografts. Although both EZH2 and BCL6 inhibitors suppressed tumor growth as single agents, the combination more potently and significantly suppressed lymphoma growth *in vivo* as demonstrated by growth curves and tumor weight (**Figure 3.18C and D**). These data suggest that targeting BCL6 and EZH2 together may provide the basis for rational combinatorial therapies for GC-derived B cell lymphomas.

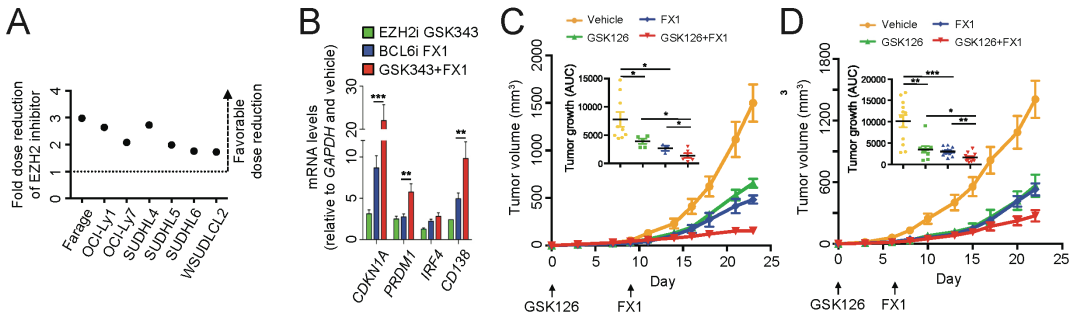

Figure 3.18. Combinatorial targeting of EZH2 and BCL6 yields enhanced anti-lymphoma effect. (A) Dose reduction plot for GSK343 at 90% growth inhibition after exposure of cells to increasing concentrations of GSK343 for 6 days and FX1 for 2 days. Data represent means of triplicate experiments. (B) RT-qPCR of the indicated mRNAs from OCI-Ly7 treated with 2 mM GSK343 for 72 hr, 25 mM FX1 for 12 hr, or the combination. Values are mean of triplicates ±SD. (C-D) Tumor growth curves and area under the curve (AUC) for SUDHL6 (C) and WSUDLCL2 (D) xenografted mice treated with vehicle (SUDHL6, n=9; WSUDLCL2, n=11), GSK126 (80 mg/kg/day, SUDHL6, n=5; WSUDLCL2, n=10), FX1 (12 mg/kg/day, SUDHL6, n=4; WSUDLCL2, n=10), or the combination of GSK126 and FX1 (SUDHL6, n=6; WSUDLCL2, n=12). Values in (C-D) are mean ± SEM. t test, *p < 0.05, **p < 0.01, ***p < 0.001. Experiments performed and data prepared by W. Beguelin.

3.17 Notes on analysis

ChIPseq peak calling for broad and lowly enriched histone modifications is difficult and is often unreliably detected. To overcome this, many studies have

examined ChIPseq read counts within gene promoter regions as a continuous quantification for occupancy. Unfortunately, this approach does not allow for the identification of “occupied” versus “non-occupied” promoters and can hinder characterization of different chromatin features within the data, *e.g.* bivalent domains. By identifying continuous regions with H3K27me3 ChIPseq enrichment greater than one standard deviation above the genome-wide mean, we were able to use the underlying structure of ChIPseq reads to determine loci where H3K27me3 exhibits clusters of 1kb bins with high enrichment. As our approach smoothed the data to remove highly enriched bins with inconsistently enriched neighbors, we have selected for long, continuous regions and it is possible that our approach may have missed shorter and more acute enrichment over a few nucleosomes. Despite such limitations, our H3K27me3 broad domain calling is supported both by association with gene repression (**Figure 3.4D**), EZH2 occupancy (**Figure 3.5C**), and ChIP re-ChIP validation of enrichment (**Figure 3.6D**). Furthermore, the identification of H3K27me3-occupied promoters allowed for key subsequent findings, including the establishment of GC B cell-specific *de novo* bivalent domains and cooperation of BCL6 and EZH2 to repress targets.

3.18 Discussion

Our work shows that EZH2 is a master regulator of the GC B cell phenotype, and that this function is aberrantly reinforced by mutant EZH2 lymphoma disease alleles (**Illustration 3.1**). We find that EZH2 mediates its effects in GC B cells by repressing target genes involved in proliferation checkpoints (e.g., *CDKN1A*) and exit from the GC and terminal differentiation (e.g., *IRF4*

and *PRDM1*). For immunoglobulin affinity maturation to occur, GC B cells must maintain their phenotype long enough to transit repeated rounds of division and somatic hypermutation. Hence, in the absence of EZH2 function, which is needed to support the GC B cell phenotype, mice display defective immunoglobulin affinity maturation. Normally, EZH2 levels decrease as B cells exit the GC reaction, enabling expression of genes that mediate terminal differentiation[47]. However, in the presence of somatically mutated EZH2, suppression of GC exit genes and checkpoints persists, resulting in hyperplasia, and the presence of other oncogenic hits may enable transformation to GCB-type DLBCL. An alternative route leading to GCB-DLBCL could involve overexpression or aberrant maintenance of WT EZH2. Indeed, the highest quartile WT EZH2-expressing GCB-DLBCL display a trend toward increased repression of EZH2 targets and tend to cluster together with EZH2 mutant patients (data not shown). The role of EZH2 in lymphomas is in large part to maintain or exaggerate (*i.e.*, to hijack) the same EZH2 transcriptional program required for development of normal GCs and immunoglobulin affinity maturation. Hence, GC B cells exhibit both oncogene and nononcogene addiction to EZH2.

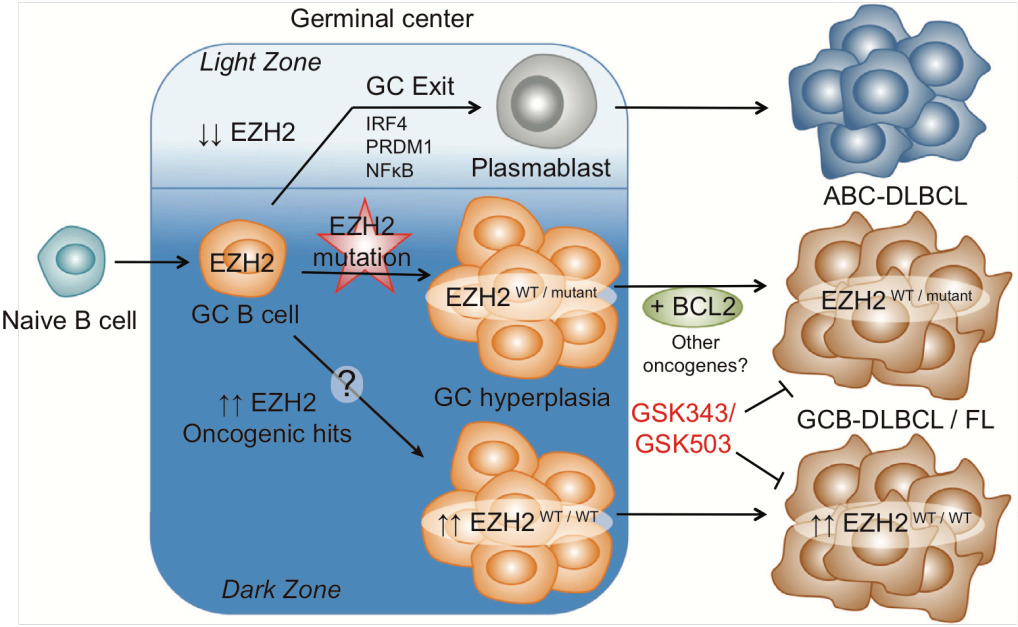

Illustration 3.1. EZH2 mutations promote lymphoid transformation. When naive B cells are activated, EZH2 expression is highly induced. EZH2 is required for GC formation and Ig affinity maturation. EZH2 levels decrease as B cells exit the GC reaction, enabling expression of genes that mediate terminal differentiation (*e.g.* IRF4, PRDM1, and NFκB). B cells in this "exiting" phase, such as plasmablasts, are believed to give rise to ABC-DLBCLs. However, the occurrence of EZH2 somatic mutations aberrantly sustains repression of proliferation checkpoint and differentiation genes, resulting in GC hyperplasia, and the presence of other oncogenic hits, such as BCL2, enables transformation to GCB-DLBCL or FL. A possible alternative route leading to GCB-DLBCL could involve overexpression or aberrant maintenance of WT EZH2 expression. GCB-DLBCLs and FLs, but not ABC-DLBCLs, require EZH2 to maintain their proliferation and survival. Thus, EZH2 methyltransferase inhibitors suppress GCB-DLBCLs, but not ABC-DLBCLs.

The actions of EZH2 appear linked in part to de novo formation of bivalent chromatin domains, whereby genes marked by H3K4me3 in naive follicular B cells acquire H3K27me3 in GC B cells, concordant with upregulation of EZH2. In stem cells, bivalent domains are hypothesized to maintain genes in a repressed but poised conformation, which can be subsequently dynamically activated or repressed according to lineage-specific differentiation

programs[225,302]. GC B cells represent a potentially unique situation where bivalent chromatin marks at specific genes are gained during differentiation, contrary to ES and tissue stem cell differentiation where bivalency is lost. We cannot exclude, however, that other cell types and tumors that overexpress EZH2 might also form bivalent domains at promoters. GC bivalent promoters are enriched in key gene sets involved in exit from the GC, such as IRF4- and CD40-induced genes, and genes upregulated in memory and plasma cells. As long as EZH2 maintains H3K27me3 at these loci, these genes are expressed at low levels, maintaining the GC phenotype. As B cells exit the GC reaction and are selected for terminal differentiation, EZH2-mediated repression of these genes is terminated. By contrast, EZH2 mutants disrupt the equilibrium of bivalent domains, enabling aberrant, persistent epigenetic silencing of genes and, in turn, allowing persistence of the GC B cell phenotype, facilitating lymphoid transformation. Consistent with this hypothesis, we found that bivalent domain genes are aberrantly repressed with in EZH2 mutant DLBCL patient specimens. These data, along with the recent identification of frequent loss-of-function mutations in the histone methyltransferase protein KMT2D in B cell lymphomas[84,86], suggest that the balance of H3K4me3 and H3K27me3 is disrupted and represents a therapeutic target in lymphoid malignancies. Mutations in KMT2D and EZH2 in lymphoma are not mutually exclusive[84], suggesting these mutations may cooperate to deregulate bivalent domains in GC B cells or that they also have independent roles in lymphomagenesis.

EZH2 mutants may also mediate their actions through additional mechanisms. We and others have noted that H3K27 ChIPseq profiles as well as gene

expression profiles induced by EZH2 inhibitors are variable between cell lines[93], suggesting that hyperactive EZH2 might lead to epigenetic instability and stochastic aberrant epigenetic silencing of different gene sets. On the other hand, variability may be due to the genetic diversity of cell lines and, under the stress of continuous passage *in vitro*, the tendency to drift apart epigenetically. This has been demonstrated in the case of cytosine methylation profiles, which differ considerably between lymphoma cell lines and primary lymphoma specimens[189]. By contrast, primary human DLBCL specimens with mutant EZH2 showed a robust signature, consisting of greater repression of EZH2 target genes, including bivalent domains and with similarity to the EZH2 mutant signature induced in the isogenic BCL1 experimental model. Moreover, analysis of individual DLBCL cell lines after exposure to GSK343 showed upregulation of GC B cell EZH2 targets, including bivalent genes as well as phenotypic effects consistent with de-repression of GC B cell EZH2 target genes. Notably, these studies underline that EZH2-mediated epigenetic effects are reversible in lymphomas, which is consistent with data in the prostate cancer field also showing that suppression of EZH2 can result in reactivation of genes with tumor-suppressing activity[303,304].

Expression of mutant EZH2 alone in GC B cells was insufficient to induce development of DLBCL. In this way, EZH2 mutation appears analogous to many of the somatic mutations in acute myeloid leukemia (AML), which induce a myeloproliferative phenotype when expressed in murine hematopoietic stem cells but, when expressed together, cooperate to form AML[305]. Mutant EZH2 induces a lymphoproliferative phenotype with expansion of the proliferative

GC B cell compartment. Constitutive expression of the GC B cell oncoproteins BCL6 and BCL2 also manifest GC hyperplasia and a partially penetrant GC B cell lymphoma phenotype[290,306]. Hence, even though normal GC B cells exhibit features of partially transformed cells, such as suppression of proliferative checkpoints and attenuated DNA damage response, multiple oncogenic hits are still required for overt lymphomagenesis. This concept is supported by mutational profiling studies, revealing multiple concurrent somatic mutations in DLBCL specimens[84,86], as well as our transplantation studies, demonstrating cooperation between mutant EZH2 and BCL2 in accelerating lymphomagenesis in mice. Knowledge of the genetic composition of lymphomas and how these cooperate to transform B cells affords the opportunity to rationally design combinatorial therapies. The enhanced anti-lymphoma activities of GSK343 and GSK503 in combination with anti-BCL2 therapies support this notion and point toward design of clinical trials geared toward the underlying biology of DLBCL and a reduced reliance on relatively nonspecific cytotoxic chemotherapy.

While previous reports have focused on the role of EZH2 inhibitors against mutant-EZH2 DLBCL[92-94], our data indicate that EZH2 is a relevant target beyond those tumors. We find that GCB-type DLBCLs are dependent on wild-type EZH2 for their proliferation and survival, regardless of somatic mutation, although response to EZH2 inhibitors is slightly delayed as compared to mutant EZH2 DLBCL cells. This finding is in agreement with the absolute requirement of normal GC B cells for EZH2 and indicates that EZH2 is a lineage factor to which GCB-DLBCLs are addicted (**Illustration 3.1**). By contrast, ABC-DLBCLs do not require EZH2 to maintain their proliferation and

survival. Indeed, the target genes repressed by EZH2 consist of many of same genes that define the ABC-DLBCL sub- type. These data provide a mechanism by which increased EZH2-mediated repression of target genes can impair B cell differentiation and demonstrate that therapeutic targeting of EZH2 in GCB-DLBCLs can induce differentiation and abrogate proliferation of GCB-DLBCLs with mutant or wild-type EZH2. We thus provide the basis for the expanded clinical translation of EZH2 inhibitors for the treatment of GCB-type DLBCLs. Clinical studies with pharmacologic EZH2 inhibitors will determine if this approach can improve outcomes for lymphoma patients. Although many GCB-DLBCL cases can be cured with combination chemoimmunotherapy regimens, such treatments involve the use of toxic drugs that carry a lifelong risk of developing second malignancies. The current standard therapy for B cell lymphomas (R-CHOP) wipes out the entire bone marrow of patients, and the B cell lineage is completely eradicated for months by rituximab (anti-CD20 antibody). In comparison, the transient suppression of high-affinity antibody formation by EZH2 inhibitors seems less likely to be clinically significant. By targeting the oncoproteins that drive and define the GCB-DLBCL phenotype, it may be possible to reduce our reliance on cytotoxic drugs to eradicate this disease.

Additionally, we show that the GC phenotype and lymphomagenesis are mediated through cooperative and mutually interdependent actions of EZH2 together with the transcriptional repressor BCL6. Our data suggest a scenario whereby, within early GC B cells, BCL6, EZH2, CBX8, BCOR, and other noncanonical PRC1 components are upregulated while canonical PRC1 components are repressed. This allows the formation of a PRC1-BCOR

complex containing CBX8. BCL6, a sequence-specific transcription factor, binds to gene promoters mostly through direct binding of its cognate DNA consensus site. At the same time, EZH2 is recruited, via still largely unclear mechanisms, to a subset of the promoters of cell cycle- and differentiation-associated genes bound by BCL6 and mediates H3K27 methylation of nucleosomes that were previously marked as active with H3K4me3 in mature resting B cells. EZH2 and BCL6 appear to arrive at these genes independently and do not physically interact. What ensues is a form of combinatorial tethering, whereby the presence of BCL6 and H3K27me3 at bivalent chromatin formed by EZH2 is required for the stable recruitment of the BCOR-CBX8 noncanonical PRC1 complex (**Illustration 3.2**). However, neither BCL6 binding to BCOR nor CBX8 binding to H3K27me3 alone is sufficient to maintain the association of the BCOR complex. Notably, binding of BCOR to BCL6 occurs with surprisingly low affinity (~20 mM) given the extended binding surface between these two proteins. Among CBX proteins, CBX8 has relatively lower binding affinity for H3K27me3[307] and may only result in metastable binding. Apparently, both of these independent protein interactions are required to sustain and stabilize the association of BCOR with this particular set of key GC B cell promoters. Thus engaged, the PRC1-BCOR-CBX8 complex then mediates PRC1 functions, including H2AK119 ubiquitylation and, in turn, putatively contributing to repression of transcription at these loci. Notably, CBX8 has not been previously implicated in B cell biology or lymphomagenesis. The fact that CBX8 loss-of-function recapitulates the effects of EZH2 loss in normal and malignant GC B cells speaks to its critical function in this cellular context.

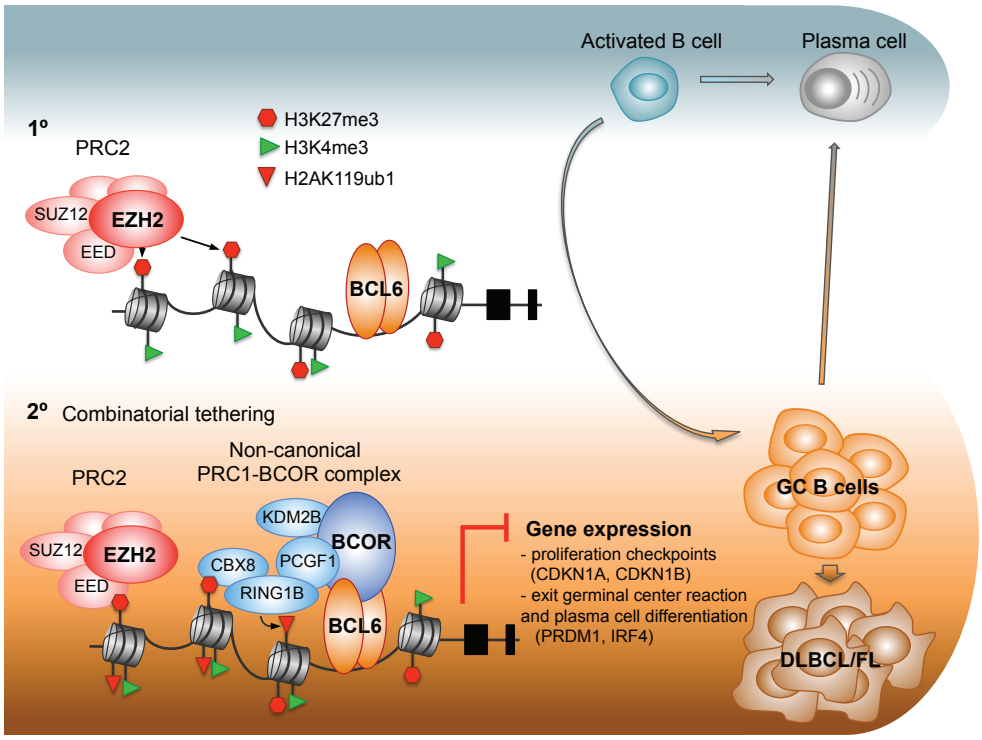

Illustration 3.2. Combinatorial tethering mediates noncanonical PRC1-BCOR complex. EZH2 and BCL6 cooperate to recruit a noncanonical PRC1-BCOR complex containing CBX8 to repress expression of differentiation-associated genes in GCBs and promote lymphomagenesis. Targeting both BCL6 and EZH2 for inhibition elicits strong anti-lymphoma activity in DLBCL.

Our proposed model of combinatorial tethering conceptually links the proposed “instructive” versus “sampling“ modes of action for PRC2 functionality[308]. In this instance, neither instructive (transcription factor-directed) nor sampling (PRC2-directed) modes are sufficient to direct PRC1 recruitment and instead the complexes must cooperate for cell context-specific gene repression. Thus, PRC2-mediated formation of bivalent chromatin at specific promoters in B cells provides a required link to support a stoichiometrically weak interaction between transcription factors and their co-repressors. By the same token, BCL6 binding to BCOR is insufficient to

sustain and stabilize its activity and requires the independent action of EZH2. Combinatorial tethering may also help to explain certain puzzling aspects of transcriptional repression. For example, BCL6 is known to bind to many promoters, yet only represses the subset of these where it forms a complex with BCOR[41]. Perhaps the combined actions of BCL6 and H3K27me3 together represent a combinatorial code that limits the formation of competent repression complexes only to sites relevant to GC B cells. Indeed, the genes where this combinatorial mechanism occurs are critical for the GC phenotype, such as *CDKN1A*, *CDKN1B*, *IRF4*, and *PRDM1*. In contrast, BCL6 and BCOR are mostly excluded from monovalent H3K27me3 domains, and hence are not involved in repressing these regions.

It remains possible that PRC1, PRC2, and BCL6 function to mutually sustain and stabilize their respective binding in bivalent promoters. Along these lines, Kalb *et al.* identified a positive feedback loop in which H2Aub promotes PRC2 binding and H3K27 trimethylation, and H3K27me3, in turn, promotes binding of canonical PRC1[309]. H2A ubiquitylation mediated by noncanonical complexes was shown to facilitate recruitment of PRC2 in embryonic stem cells engineered to contain a Polycomb tethering sequence[310]. Recruitment of PRC2 was also reported to occur through canonical PRC1-mediated H2A ubiquitylation[311]. Indeed, our data hint at additional aspects of Polycomb functionality. For example, the finding that disruption of the BCL6-BCOR interaction reduces EZH2-mediated H3K27me3 without loss of PRC2 complex binding could be linked to EZH2 requirement for H2A ubiquitylation; or to impaired PRC2 function due to increasing H3K36 methylation because of loss of KDM2B[312]. Indeed, H3K36me2 was increased after treating three different

GCB-DLBCL cell lines with GSK343 or after disrupting BCL6-BCOR interaction with FX1 (data not shown). Collectively, the combinatorial tethering model expands notions on how transcription factors and Polycomb complexes can work integratively to direct gene-specific repression. It is important to also underline that these results do not explain all the effects of BCL6, EZH2, and BCOR in B cells. For example, BCOR binds to many promoters independent of the presence of BCL6 and it was recently shown in embryonic stem cells that KDM2B, a subunit of the PRC1-BCOR complex, can contribute to recruitment of noncanonical PRC1 complexes via binding of its CXXC motif to unmethylated CpG islands[313,314]. Our analysis of BCOR distribution in GC B cells suggests that a similar mechanism may be at play at different sets of target genes independent of the BCL6-EZH2 mechanism described herein[41].

Constitutive expression of BCL6, as well as somatic mutation of EZH2, can prevent the resolution and sustain the GC phenotype, potentially explaining how they induce lymphomas. We have shown that this is partially linked to their common action in the combinatorial tethering of the noncanonical PRC1-BCOR complex to bivalent chromatin domains formed during the humoral immune response. The enhanced anti-lymphoma activity observed by combining EZH2 with BCL6 inhibitors is likely, at least in part, due to more profound disruption of bivalent gene repression, as the combination results further increased expression of these transcripts. Administration of BCL6 and EZH2 inhibitors may thus constitute a mechanism-oriented rational combinatorial therapy by disabling both arms of the PRC1-BCOR tethering mechanism.

3.19 Material and methods

Murine models

Conditional *Ezh2* knockout mice (*loxP*-flanked Ezh2 allele, *Ezh2*$^{fl/fl}$) were a generous gift of Dr. Alexander Tarakhovsky, The Rockefeller University[50]. By crossing *Ezh2*$^{fl/fl}$ with the transgenic *Cγ1cre* strain (The Jackson Laboratory, 010611), we generated heterozygous *Ezh2*$^{fl/WT}$ mice, which were crossed to yield *Ezh2*$^{fl/fl}$ mice. As control group, we used *Ezh2*$^{fl/fl}$ *Cγ1cre* negative littermates. *Ezh2(Y641F)*$^{fl/WT}$, *Bcl6*$^{fl/fl}$, and *Bcor*^{fl} mice were crossed with *Cγ1cre* strain. The *Cγ1cre* negative littermates were used as control groups. *IμBcl6* were obtained from Dr. Ricardo Dalla-Favera, Columbia University[290]. The conditional *Bcor* allele (*Bcor*^{fl} which is on the X chromosome), which contains loxP sites flanking *Bcor* exons 9 and 10, was generated by homologous recombination (M.Y. Hamline, C.M. Corcoran, J.A. Wamstad, M.D. Gearhart, I. Miletich, J. Feng, M. Hemberger, P.T. Sharpe, and V.J. Bardwell, manuscript in preparation). CRE-mediated deletion results in a premature stop codon and a *Bcor* null allele. Conditional *Bcl6* knockout mice (*loxP*-flanked Bcl6 allele, *Bcl6*$^{fl/fl}$) were generated by Taconic[315]. *Cbx8*^{fl} were obtained from Dr. Haruhiko Koseki, Riken Center, Japan[301]. *Bcl6*^{BTBmut} mice were developed as previously described[43]. All knockout, knock-in and transgenic mice were used for assessment of the germinal center formation, which were induced with SRBC.

Germinal center assessment in mice

The Research Animal Resource Center of the Weill Cornell Medical College of Medicine approved all mouse procedures. Age- and sex-matched C57BL6 mice were immunized intraperitoneally at 8 to 12 weeks of age with 0.5 ml of a 2% sheep red blood cell (SRBC) suspension in PBS (Cocalico Biologicals), or

100 µg of highly substituted NP-KLH (NP-25 Keyhole Limpet Hemocyanin, Bioresearch Technologies) in alum (Thermo Scientific), and sacrificed after 10 or 14 days, respectively.

For GSK503 and FX1 experiments: drug or vehicle (20% captisol for GSK503, and 30% PEG-300 + 3% dextrose + 5% Tween-80 for FX1) was injected intraperitoneally starting the following day after induction of GC by SRBC and administered daily at a concentration of 150 mg/kg/day GSK and 50 mg/kg/day FX1 for 9 consecutive days after which the mice were sacrificed (day 10).

Flow cytometry analysis

Analysis of splenocytes: single-cell suspensions from mouse spleens were stained using the following fluorescent-labeled anti-mouse antibodies: PE-Cy7 conjugated anti-B220, PE-Cy7 conjugated anti-CD21, PE conjugated anti-CD23, APC conjugated anti-IgM (eBioscience), APC conjugated anti-B220, PE conjugated anti-FAS, APC conjugated anti-CD38, FITC conjugated anti-GL7, PE conjugated anti-IgD (BD Bioscience), AlexaFluor488 conjugated anti-EZH2, AlexaFluor488 conjugated IgG1κ isotype control (BD Biosciences). DAPI was used for the exclusion of dead cells.

Analysis of cell lines: cells were stained using the following fluorescent-labeled anti-human antibodies: PE conjugated anti-CD20, PE-Cy5.5 conjugated anti-CD20, FITC conjugated anti-CD138, PE-Cy5.5 conjugated anti-IgM, PE-Cy7 conjugated anti-CD27, APC conjugated anti-IgG1, APC conjugated anti-annexinV (all from BD Bioscience), multimix for plasma cells (FITC conjugated anti-CD19, PE conjugated anti-Ig lambda, APC conjugated anti-Ig kappa) and isotype control multimix (Dako). To evaluate total levels (cytosolic+cell surface) of Igλ, Igκ, IgM and IgG1, cells were permeabilized with BD

Cytofix/Cytoperm Fixation/Permeabilization Solution Kit (BD Biosciences). Data were acquired on MacsQuant flow cytometer (Miltenyi Biotec) and analyzed using FlowJo software package (TreeStar).

Immunohistology and diagnosis

Mice organs were fixed in 4% formaldehyde and embedded in paraffin. Deparaffinized slides were antigen retrieved in citrate buffer pH 6.4 and endogenous peroxidase (HRP) activity was blocked by treating the sections with 3% hydrogen peroxide in methanol. Indirect immunohistochemistry was performed with antispecies-specific biotinylated secondary antibodies followed by avidin–horseradish peroxidase or avidin-AP, and developed by Vector Blue or DAB color substrates (Vector Laboratories). Sections were counterstained with hematoxylin if necessary. The following antibodies were used: biotin-conjugated anti-PNA (Vector Laboratories), biotin-conjugated anti-B220 (Invitrogen RM2615), EZH2 (Cell Signaling 5246), Ki67 (Vector VP- K451), CD3 (Vector VP-RM01). Slides were scanned using a Zeiss Mirax Slide Scanner and photomicrographs were examined with Panoramic Viewer software. ImageJ 1.44o software (NIH) was used to quantify germinal center areas.

H&E and IHC stained sections were examined by a board certified veterinary pathologists and lesions were classified based on morphologic features and marker expression, according to the classification system for murine lymphoid neoplasms of the hematopathology subcommittee of the Mouse Models for Human Cancers Consortium, United States National Cancer Institute[316].

ELISA

Murine serum samples were collected 14 days after NP-KLH immunization and immunoglobulin levels were analyzed by ELISA. Sera were tested for the

binding of NP-specific IgG1, IgG2a, IgG2b and IgG3 antibodies (SouthernBiotech) to NP4-BSA coated plates.

Immunoblotting

Lysates from splenocytes and DLBCL cells were prepared using 20 mM Tris, pH 8, 135 mM NaCl, 1% NP-40, 10% glycerol, 1 mM PMSF, and complete protease inhibitor cocktail (Roche) lysis buffer. Lysates for nuclear fractions of BCL1 cells were obtained using the Nuclear Complex Co-IP Kit (Active Motif). Protein lysates were resolved by SDS-PAGE, transferred to PVDF membrane, and probed with the indicated primary antibodies: EZH2 (Active Motif 39933 and BD 612666), H3K27me3 (Millipore 07-449 and 17-622), EED (Millipore 09-774), SUZ12 (Santa Cruz sc-67105), FLAG (Sigma F1804), p21 (Cell Signaling 2947), pan-Histone 4 (Abcam ab7311), pan-Histone 3 (Millipore 07-690), αTubulin (Sigma), rabbit polyclonal antibodies were raised against BCOR, KDM2B, and PCGF1 using glutathione S-transferase (GST) fusions of human BCOR(C) (1035-1230), human KDM2B (726-817), and human PCGF1 (128-189) and subsequently affinity purified (Gearhart et al., 2006); RING1B (Bethyl A302-869A), CBX8 (Bethyl A300-882A), BCL6 (Santa Cruz sc-7388), BCL6 (Santa Cruz sc-858), H2AK119ub1 (Cell Signaling 8240), αTubulin (Sigma), βActin (Sigma A5441), myc (Santa Cruz 9E10). Membranes were then incubated with a peroxidase-conjugated correspondent secondary antibody and detected using enhanced chemiluminescence. Densitometry values were obtained by using ImageJ 1.44o software (NIH).

Immunoprecipitation

Lysates from DLBCL cells were prepared using 25 mM Tris, pH 8, 150 to 250 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate, 1 mM PMSF, and complete protease inhibitor cocktail (Roche) lysis buffer. Two µg

of antibodies were added to the precleared sample and incubated overnight at 4°C. The complexes were purified using protein-A beads (Roche) followed by elution from the beads using SDS load buffer.

Full length human CBX8, PCGF1, RNF2, the short isoform of KDM2B (derived from the second promoter after the sequences encoding the JmjC domain) and the carboxy-terminus of BCOR (amino acids 1562-1755) were cloned into myc- or myc-flag-tagged versions of the pIEx-4 expression vector (EMD Millipore #71235). Mutations were made in CBX8 and RNF2 using PCR subcloning and confirmed by sequencing. Sf9 insect cells were transiently co-transfected as indicated using Insect GeneJuice Transfection Reagent (EMD Millipore #71259). Cells were harvested after 68 hours in 0.5 ml lysis buffer containing 1X phosphate buffered saline, 10% glycerol, 0.1% NP-40, 1 mM dithiothreitol (DTT), 2 mM sodium fluoride, 0.2 mM PMSF, and complete protease inhibitor cocktail without EDTA (Roche #11873580001). The extracts were sonicated for 10 seconds at 25% power with a stepped micro tip, centrifuged at 12,000 g and supplemented with 2 mM MgCl2and 25 U / sample Benzonase (EMD Millipore #70746). Lysates were incubated with Anti-Flag M2 Agarose Affinity Gel (Sigma #A2220) for 2 hours at 4°C with gentle agitation. The affinity gel beads were washed in lysis buffer and boiled in 1x LDS sample buffer (Thermo Fisher #NP0007) supplemented with 5% beta-mercaptoethanol. Proteins were resolved on a 4-12% Bis-Tris NuPage Gel in MOPS buffer (ThermoFisher #NP0322 and #NP0001) and transferred to 0.45μM nitrocellulose membrane (NitroBind #1215471). Blots were incubated with anti-Myc (1:500 dilution 9E10 Santa Crux #SC-40) and visualized with either Alexa Fluor 680-conjugated Goat Anti-Mouse IgG (H+L)(input samples, 1:500 dilution Jackson ImmunoResearch #115-625-166)

or TrueBlot Anti-Mouse Ig DyLight 680 (immunoprecipitates, 1:250 dilution Rockland #18-4417-32) using an Odyssey Scanner (Li-Cor Biosciences). For HEK293 immunoprecipitations the WT and mutant form of CBX8 were subcloned into an EF1a driven His- CBP-3XFlag lentiviral vector that co-expresses eGFP. Virus particles were produced using an empty vector control and the two CBX8 constructs in HEK293 cells and subsequently used to transduce freshly seeded HEK293 cells with nearly 100% efficiency based on eGFP. Cells were expanded for 5 days and immunoprecipitations were carried out as above Sf9 insect cell experiments.

Cell lines

The DLBCL cell lines OCI-Ly1 ($EZH2^{Y641N}$) and OCI-Ly7 (WT EZH2) were grown in Iscove's medium supplemented with 10% FBS and penicillin G/streptomycin; the DLBCL cell lines Farage (WT EZH2), WSU- DLCL2 ($EZH2^{Y641F}$), Pfeiffer ($EZH2^{A677G}$), SUDHL6 ($EZH2^{Y641N}$), SUDHL5 (WT EZH2) and SUDHL4 ($EZH2^{Y641S}$) were grown in RPMI medium supplemented with 10% FBS, penicillin G/streptomycin, l-glutamine, and HEPES. Sf9 insect cells (Novagen #71104-3) were maintained in suspension in logarithmic phase in serum-free medium at 28°C and 150 rpm agitation.

ChIP, ChIP re-ChIP and qPCR

ChIP was performed as previously described[45]. Briefly, 1e+08 cells were fixed with 1% formaldehyde, lysed, and sonicated (Branson Sonicator; Branson) leading to a DNA average size of 200 bp. Five µg of antibodies anti-EZH2 (Active Motif 39901), RING1B (Bethyl A302-869A), CBX8 (Bethyl A300-882A), BCL6 (Santa Cruz sc-858), H3K27me3 (Abcam 6002), H2AK119ub1 (Cell Signaling 8240), H3K36me2 (Active Motif 39255), EED (Millipore 09-774), SUZ12 (Santa Cruz sc-67105), BCOR, KDM2B, $PCGF1^{293}$ or control IgG

(Millipore) were added to the precleared sample and incubated overnight at 4°C. The complexes were purified using protein-A beads (Roche) followed by elution from the beads and decrosslinking. DNA was purified using PCR purification columns (QIAGEN). For ChIP re-ChIP experiments, chromatin immunoprecipitates were eluted with DTT and then subjected to a second round of immunoprecipitation with the indicated antibodies or IgG. ChIP and ChIP re-ChIP DNA was amplified by real-time quantitative PCR using SyberGreen (Applied Biosystems) on 7900HT Fast Real-Time PCR System (Applied Biosystems).

Plasmids and shRNAs

EZH2 cDNA was cloned into pRetroX-ZsGreen vector (Clontech) and Y641X mutations were made using QuickChange Lighting Site-Directed Mutagenesis Kit (Stratagene) following manufacturer's recommendation. Retroviruses were produced by transfection of amphotropic 293T cells with appropriate plasmids and FuGENE 6 Transfection reagent (Roche). shRNAs were delivered by lentivirus infection, which were produced by transfection of 293T cells with the vector pLKO.1. For shRNA anti-EZH2 we used the pLKO.1-YFP vector, and infected cells were identified by YFP expression by flow cytometry. For CBX8 we used pLKO.1-puro, and infected cells were selected by puromycin treatment (1μg/mL). Mature antisense sequences of shRNA used to knockdown EZH2 and CBX8 were: shEZH2#2: 5'-TTTGGTCCCAATTAACCTAGC-3', shEZH2#3: 5'-TAATGGGATGACTTGTGTTGG-3', shCBX8#2: 5'-AAAGTTTGAGGTCACGTCCGT-3', shCBX8#4: 5'-TTACTTTCCTTAATGGTGACG-3'.

Proliferation and differentiation assays

Cells were plated in triplicate (100,000 BCL1 cells/mL and 500,000-800,000 DLBCL cells/mL) and live cells were counted at indicated days using either trypan blue exclusion or flow cytometry using annexinV and DAPI exclusion. The experiments were performed with two different biological replicates and data is presented as average of six counts +/- standard deviation. Every four days, cells were replated at initial concentration with fresh media and inhibitors if necessary. To induce differentiation, BCL1 cells were treated with 20ng/mL of recombinant mouse IL-2 and IL-5 (R&D Systems). Cells were counted at day 3 and were replated in the presence of cytokines for additional two days.

Growth inhibition and combination of drugs

DLBCL cell lines were grown at concentrations sufficient to keep untreated cells in exponential growth over the complete drug exposure time. Cell viability was determined using a fluorometric resazurin reduction method (CellTiter-Blue, Promega) and trypan blue automatic method (TC10, BioRad). Fluorescence (Ex560nm / Em590nm) was determined with the Synergy4 microplate reader (BioTek). The number of viable cells was calculated by using the linear least-squares regression of the standard curve. The fluorescence was determined for three replicates per treatment condition and normalized to their respective controls. To plot dose-effect curves CompuSyn software (Biosoft) was used, and drug concentrations that inhibits the growth of the cell lines by 50% compared to control (GI50) were determined. Data were presented as the mean GI50 with a 95% confidence interval for duplicate experiments. In the experiments using combination of drugs, DLBCL cell lines were exposed to 5 concentrations of GSK343 or GSK503 for 6 days followed by 5 concentrations ABT-737 (Selleck) or Obatoclax (Selleck) for additional 48h and analyzed for cell viability as before. To quantify the effect of the

sensitization, we calculated the dose reduction index (DRI) at GI90 using CompuSyn software. The DRI is a measure of how many fold the dose of each drug in a combination may be reduce at a given effect level compared with the doses of each drug alone and is based on the equation DRI = (Dx)1 / (D)1, were (Dx)1 represent the dose of drug 1 for a given effect x and were (D)1 represent the dose of drug 1 given in combination to reach the same effect x.

Microarray analysis with DLBCL patient samples

DLBCL patient samples were hybridized to Affymetrix HG-U133 Plus 2.0 microarrays. Affymetrix data were extracted, normalized and summarized with the RMA method from Bioconductor's 'affy' package, using the default settings, and remapped to the newest annotated RefSeq genes[317].

BCOR and PCGF1 tandem affinity purifications

For mammalian expression, stable cell lines of HEK293 cells were generated by infection with ProtA2-TEV- CBP-Flag-Pcgf1[293], His-CBP-3XFlag-Bcor(A)-HA, and His-CBP-3XFlag-HA retroviruses. Bcor isoform a cDNA was tagged at the N-terminus with tandem His tag, calmodulin binding peptide, and three copies of the Flag tag, and at the C-terminus with HA. This or just the affinity tags were then cloned into a modified version of pLentiLox3.7 (containing a shortened version of the EF promoter and beta globin 5` untranslated region and a GFP blasticidin fusion gene) creating His-CBP-3XFlag-Bcor(A)-HA, and His-CBP- 3XFlag-HA encoding retroviruses. All nucleotides of inserts were verified by sequencing. Cells were cultured in Dulbecco's minimal essential medium (Cellgro) with 5% calf serum (Biosource) and 5 µg/ml blasticidin (or 1 µg/ml puromycin for the Pcgf1 cells). Nuclear extracts were supplemented with 0.1% NP-40, 0.1% Tween, 2 mM EGTA, and 0.5 mM EDTA and incubated

with M2-agarose (Sigma) overnight. Beads were washed in 50 mM Tris pH 8, 1 mM MgCl2, 1 mM imidazole, 0.1% NP-40, 20% glycerol, 2 mM DTT, 2 mM PMSF, 2 mM EGTA, 0.5 mM EDTA, protease inhibitor cocktail (Complete EDTA-free; Roche), and 350 mM KCl (TGN350). Complexes were eluted with 30% yields using 2 mg/ml Flag peptide, substituting 2 mM CaCl2 for EGTA and EDTA in the TGN350 buffer, and recaptured with calmodulin-sepharose (GE HealthCare). Calmodulin beads were washed with TGN350 and stripped of protein using 100 mM Tris pH 8.5 with 8 M urea, heated at 37oC for 1 hour. For visualization, protein samples were resolved on a 4-12% NuPAGE Novex Bis-Tris gel in MOPS running buffer (Invitrogen) and visualized by silver staining (SilverQuest; Invitrogen). For mass spectrometry analysis, complexes were submitted for trypsinization and analysis at the University of Minnesota Center for Mass Spectrometry and Proteomics.

Mass Spectrometry data analysis

Mass spectrometry analysis of H3K27 methylation was performed as described previously[318]. The amount of H3K27me3 mark was calculated as a percentage of a peptide encompassing amino acids 27-40 from histone H3. Samples were analyzed using liquid chromatography mass spectrometry (LC-MS) on an Orbitrap LTQ mass spectrometer. Proteins were identified by searching human protein databases with Sequest. Peptide identification and protein coverage were calculated using Sequest and Scaffold. Proteins identified in the corresponding empty vector control purification were eliminated from each purification list. In addition, proteins previously found to bind affinity tags non-specifically were eliminated from each purification list[319].

Proliferation assays

Cells were plated in triplicate (500,000-800,000 DLBCL cells/ml) and live cells were counted at indicated days using a fluorometric resazurin reduction method (CellTiter-Blue, Promega). The experiments were performed with two different biological replicates and data is presented as average of six counts +/- standard deviation. Every three days, cells were replated at initial concentration with fresh media.

B cell purification and characterization

Human B cell populations were affinity-purified from de-identified human tonsillectomy specimens using standard protocols[45] with approval from the Human Research Protections Programs, Division of Research Integrity of the Weill Cornell Medical College, in accordance with the Declaration of Helsinki. NB and GC B cell purity was determined by flow cytometry analysis of surface IgD (BD Pharmingen), CD77 (AbD Serotech) and CD38 (BD Pharmingen).

DLBCL patient samples

Patient-deidentified leftover tissues were obtained at diagnosis from patients with *de novo* DLBCL in Vancouver at the British Columbia Cancer Agency. Cases were selected on the basis of the presence of at least 80% of the neoplastic cells within the tumor section. Patient samples were co-cultured with irradiated HK stromal cells and were kept in culture for no longer than 5 days, given that primary DLBCL cells do not proliferate or maintain their viability for very long in culture.

RT-qPCR

RNA was prepared using Trizol extraction (Invitrogen). cDNA was prepared using cDNA synthesis kit (Thermo Scientific) and detected by fast SyberGreen (Applied Biosystems) on 7900HT Fast Real-Time PCR System (Applied Biosystems). We normalized gene expression to HPRT1 or GAPDH and

expressed values relative to control using the ΔΔCT method. Results were represented as fold expression with the standard deviation for 2 series of triplicates.

Primers used for qPCR

Used for	Gene		Oligonucleotide (5'-3')
cDNA	*IRF4*	Fwd	AGAAGAGCATCTTCCGCATC
		Rev	CCTTTAAACAGTGCCCAAGC
	PRDM1	*Fwd*	*CTACCCTTATCCCGGAGAGC*
		Rev	*GCTCGGTTGCTTTAGACTGC*
	CD138	*Fwd*	*GAGCAGGACTTCACCTTTGA*
		Rev	*TTCGTCCTTCTTCTTCATGC*
	CDKN1A	*Fwd*	*GGAAGACCATGTGGACCTGT*
		Rev	*TAGGGCTTCCTCTTGGAGAA*
	TP73	*Fwd*	*CTCCACCTTCGACACCATGT*
		Rev	*GGACACCTTGATCTGGATGG*
	BMI1	*Fwd*	*AATCCCCACCTGATGTGTGT*
		Rev	*GGTCTGGTCTTGTGAACTTGG*
	BCOR	*Fwd*	*CGATGCCTATAGCGATGTGTT*
		Rev	*TCCGAAAGCAGTAGCCAGTT*
	PCFG1	*Fwd*	*GAGACACAGCCACTGCTCAA*
		Rev	*ATTCCCGAATCCGTTTCTCT*
	PCGF2	*Fwd*	*CATCGGACTACACGGATCAA*
		Rev	*CACACATGGGGCAGTATTTG*
	CBX8	*Fwd*	*GCATGGAATACCTCGTGAAA*
		Rev	*CTCAAAGGCTGCGAGCAAG*
	GAPDH	*Fwd*	*CGACCACTTTGTCAAGCTCA*
		Rev	*CCCTGTTGCTGTAGCCAAAT*
	HPRT1	*Fwd*	*AAAGGACCCCACGAAGTGTT*
		Rev	*TCAAGGGCATATCCTACAACAA*
	Bcor ex2-3	*Fwd*	*ATGCTTTCTGCAACCCCTCT*
		Rev	*AGGGGAGTCTCCTCCCTCAG*
	Bcor ex8-9	*Fwd*	*AGAGAAGCCTGGCAGGAAA*
		Rev	*GCTTGGCTGAGTCTGCTTTT*
	Bcor ex14-15	*Fwd*	*TTGCTGAAAGCTCCCTCTTG*
		Rev	*TAAAACTCCGCCTCTGCAAT*
	Ezh2	*Fwd*	*ATCTGAGAAGGGACCGGTTT*
		Rev	*GCTGCTTCCACTCTTGGTTT*
	Gapdh	*Fwd*	*CTGCACCACCAACTGCTTAG*
		Rev	*GGATGCAGGGATGATGTTCT*

ChIP	*IRF4*	Fwd	CGACTCCCACCCCATCTG
		Rev	CGACAGTCCGGTTAGCTCAT
	PRDM1_1	Fwd	GTCCGGTGAGCACAAAATTC
		Rev	TCTTCCCCTCTTTTTAGGAGGT
	PRDM1_2	Fwd	CAGAATTCACCCAGCCTTGT
		Rev	CAGGCGGTAAACACCAGAAT
	CDKN1A	Fwd	CAGTGGACCTCAATTTCCTCA
		Rev	AAAACGATGCACCTCTCTGC
	CDKN1B_1	Fwd	CGAAGAGTTAACCCGGGACT
		Rev	AGTAGAACTCGGGCAAGCTG
	CDKN1B_2	Fwd	CAGGTTTGTTGGCAGCAGTA
		Rev	AGGAGGAGATCCATTGGTTG
	HOXA7	Fwd	GCTAAAAAGCGCGTTCACAT
		Rev	GCTCCGTCCAAAAGAAAATG
	BCL6	Fwd	GCAGTGGTAAAGTCCGAAGC
		Rev	AGCAACAGCAATAATCACCTG
	ARID3A	Fwd	TGGAGACTTCACTCCCCACT
		Rev	CTACCCTCCCTCCTCTCTGG
	ARID3B	Fwd	CCAACCTCCGCAATAGAAAA
		Rev	ACTGAGTTGTGGGAGGAAGC
	Negative control region_1	Fwd	TAGCTGGGAAGCTGGGACTA
		Rev	GGTTTCCTTGCCCTAAAAGG
	Negative control region_2	Fwd	AACCTGCAAAACATGGTTATTT
		Rev	AATTTGCCCAAACAGCAAGT

ChIPseq and mRNAseq library preparation and Illumina sequencing processing

ChIPseq and RNAseq libraries were prepared using the Illumina ChIPseq and TruSeq RNA sample kits, respectively, according to the manufacturer. Libraries were validated using the Agilent Technologies 2100 Bioanalyzer and Quant-iTTM dsDNA HS Assay (Life Technologies), and 8-10 pM sequenced on HiSeq2000 sequencer as follows: ChIPseq, 1 x 50; mRNAseq, 2 x 50. RNA sequencing results were aligned to mm10 using STAR[278] and annotated to RefSeq using the R subread package[279]. Differentially expressed genes

were identified using the EdgeR package GLM[320] with thresholds of fold-change >1.5 and p<0.01, adjusted for multiple testing using Benjamini-Hochberg correction. ChIPseq experiments from human B cells, human cell lines and murine cells were aligned to the hg18, hg19 and mm9 genome, respectively using ELAND. H3K4me3, BCL6 and BCOR ChIPseq reads were called into peaks using the ChIPseeqer framework (p<10-15 and fold-change threshold 2)[277] and H3K27me3 and EZH2 ChIPseq reads were quantified in 1kb bins genome-wide, identifying regions of enrichment as consecutive bins with read counts greater than one standard deviation of the genome-wide mean.

ChIPseq and mRNAseq analysis GCB-specific H3K27me3 targets: H3K27me3 ChIPseq read counts were determined for all regions +/- 2kb of RefSeq TSS for naive B cells and GC-B cells, normalizing for total ChIPseq reads. Promoter regions with 1.5- fold greater H3K27me3 ChIPseq reads in GCB cells and corresponding RNAseq gene expression 1.5-fold greater in GCB cells relative to naive B cells were identified. **GSK343-responsive genes:** RNAseq was performed in OCI-Ly7, OCI-Ly1, SUDHL5, Farage, WSU-DLCL2 and Pfeiffer cells treated with 2 μM GSK343 or GSK669 for 7 days. Gene expression ratios were determined in cells treated with GSK343 compared to cells treated with GSK669 and were identified using Cufflinks. GSK343- responsive genes were identified as genes that were upregulated 1.2-fold in four cell lines. **shEZH2-responsive genes:** RNAseq was performed in OCI-Ly7, OCI-Ly1, SUDHL5, Farage and WSU- DLCL2 cells transduced with shRNA for *EZH2* or control shRNA for 7 days. Gene expression ratios were determined in cells transduced with *EZH2* shRNA compared to cells transduced with control shRNA and were identified using Cufflinks. shEZH2-

responsive genes were identified as genes that were upregulated 1.2-fold in four cell lines. **Plasma/Memory Cell Signature**: RNAseq was performed in three replicates of plasma cells, three replicates of memory cells, and four replicates of GCB cells. The plasma/memory cell signature was determined as the union of genes upregulated 1.5-fold in plasma cells or memory cells compared to GCB cells. **Genes with increased H3K27me3 in mutant EZH2 BCL1 cells**: H3K27me3 ChIPseq read counts were determined for all regions +/- 2kb of RefSeq TSS for BCL1 cells transduced with WT EZH2, Y641F EZH2, or Y641N EZH2, normalizing for total ChIPseq reads. Promoter regions with 1.5-fold greater H3K27me3 ChIPseq reads in mutant BCL1 cells were identified. Gene expression ratios were determined in BCL1 transduced cells treated with 0.5 μM GSK343 compared to cells treated with 0.5 μM GSK669 for 3 days and were identified using Cufflinks. A 0.5 μM dose was used since it was sufficient to at least partially reverse H3K27me3, whereas 2.5 μM were used in functional assays since this corresponds to the IC50 for this drug.

Gene Set Enrichment Analyses

Enrichment of gene sets was performed using the GSEA algorithm as described in Subramanian *et al*.[321]. Enrichment for repression in mutant EZH2 samples was assessed using the GSEA algorithm against a gene list pre-ranked for log2 ratio of expression from wild-type EZH2 samples to expression from mutant EZH2 samples. Enrichment for responsiveness to EZH2 inhibition was assessed against a gene list pre-ranked for log2 ratio of expression after GSK343 treatment to expression after GSK669 treatment.

Gene Category Enrichment Analysis

Unsupervised pathway analysis was performed using information-theoretic pathway analysis approach as described in Goodarzi *et al*[322]. Briefly, pathways that are informative about non-overlapping gene groups were identified. Pathways annotations were used from the Biological Process annotations of the Gene Ontology database (http://www.geneontology.org) and signature categories from the Staudt Lab Signature database[323]. Only human-curated annotations were used from the Gene Ontology database and only pathways with 5 genes or more, and with 300 genes or less were evaluated. This pathway analysis estimates how informative each pathway is about the target gene groups, and applies a randomization-based statistical test to assess the significance of the highest information values. We use the default significance threshold of $p<0.005$. We estimated the false discovery rate (FDR) by randomizing the input profiles iteratively on shuffled profiles with identical parameters and thresholds, finding that the FDR was always less than 5%. For each informative pathway, we determined the extent to which the pathway was over-represented in the target gene group, using the hypergeometric distribution, as described in Elemento *et al*[324]. A supervised analysis was also performed by identifying genes upregulated 1.5-fold among plasma cells (three RNAseq replicates) or memory cells (three RNAseq replicates) compared to GCB cells (four RNAseq replicates). The significance of over-representation among these supervised categories was determined using the hypergeometric distribution.

GCB-DLBCL patient clustering dendrogram

GCB-DLBCL patient samples were separated into three groups: samples with wild-type EZH2 and EZH2 expression in the top quartile, samples with wild-

type EZH2 and EZH2 expression in the bottom quartile, and samples with mutant EZH2. Gene expression profiles of GCB cell bivalent genes showing repression in GCB-DLBCL with mutant EZH2 were analyzed using Euclidian distance and Ward's minimum variance method.

Bone marrow transplantation

Murine bone marrow transplantation assays were performed as described previously[325]. Briefly, bone marrow cells from 6-8 week old male donors were harvested and, for Figure S8C-S8I, cells were transduced with viral supernatants containing either pRetroX-IRES-ZsGreen1 empty vector, pRetroX-EZH2WT-IRES-ZsGreen1, or pRetroX- EZH2Y641F-IRES-ZsGreen1. Either 750,00 or 1e+06 bone marrow cells of each type were injected into the tail veins of lethally irradiated female C57BL6 mice. The donor mice from **Figure 3.7** were BCL2 transgenic animals (VavP-Bcl2)[306]. The donor mice used in **Figure 3.17** were IμBcl6[290], *Ezh2(Y641F)*$^{fl/WT}$;Cγ1-cre, *Ezh2(Y641F)*$^{fl/WT}$;Cγ1-cre;IμBcl6 and control littermates. Nonlethal mandibular bleeds were performed monthly after transplantation to assess disease severity in all mice. With the exception of mice euthanized at specific time points, all mice were followed until any one of several criteria for euthanizing were met, including severe lethargy, more than 10% body weight loss, and palpable splenomegaly that extended across the midline, in accordance with our Weill Cornell Medical College and Memorial Sloan-Kettering Cancer Center Institutional Animal Care and Use Committee–approved animal protocols. Animal care was in strict compliance with institutional guidelines established by the Weill Cornell Medical College, the Memorial Sloan-Kettering Cancer Center, the Guide for the Care and Use of Laboratory

Animals (National Academy of Sciences 1996)[326], and the Association for Assessment and Accreditation of Laboratory Animal Care International.

IgVH rearrangement analysis

RT-PCR to evaluate IgVH rearrangements was performed on cDNA of B220 enriched splenocytes with a set of forward primers that anneal to the framework region of the most abundantly used IgVH gene families and reverse primers located in the JH1-4 gene segments[327].

Mice xenotransplant

Six- to eight-week old male SCID mice housed in barrier environment were subcutaneously injected in the left flank with 1e+07 human DLBCL cells (SUDHL6 and WSU-DLCL2). Tumor volume was monitored every other day using electronic digital calipers in 2 dimensions. Tumor volume was calculated using the following formula: tumor volume (mm3) = (smallest diameter2 × largest diameter)/2. When tumors reached a palpable size (<50 mm3 after 17 days of cell injection), the mice were randomized to 2 different treatment arms. One group was injected intraperitoneally with vehicle (20% captisol) and the other group received GSK503 or GSK126 administered daily at a concentration of 80 mg/kg/day for 7 to 9 consecutive days. On day 7 or 9 of treatment the mice from each group were randomized to 2 arms, generating 4 different treatment arms. Two groups were treated with 12 mg/kg/day FX1 or RI-BPI for additional 15 days. Drugs were administered in a concurrent schedule by 2 intraperitoneal injections 4–6 hours apart. All mice were euthanized when at least 2 out of 10 tumors reached 20 mm in any dimension (equivalent to 1–1.5 grams), which was generally on day 20 to 24 of the treatment schedule.

EZH2 small molecule inhibitors

GSK343 was synthesized as described in Verma *et al*[328]. GSK126 was synthesized as described in McCabe et al.[93], and GSK503, as described in (Beguelin et al., 2013).

BCL6 inhibitors

FX1 small molecule was synthesized as described in Cardenas et al.[280], based on compound 79-6[329].

REFERENCES

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

2. Berek, C., Berger, A. & Apel, M. Maturation of the immune response in germinal centers. *Cell* **67,** 1121–1129 (1991).

3. Jacob, J., Kelsoe, G., Rajewsky, K. & Weiss, U. Intraclonal generation of antibody mutants in germinal centres. *Nature* **354,** 389–392 (1991).

4. Cyster, J. G. B cell follicles and antigen encounters of the third kind. *Nat. Immunol.* **11,** 989–996 (2010).

5. Gonzalez, S. F. *et al.* Trafficking of B cell antigen in lymph nodes. *Annu. Rev. Immunol.* **29,** 215–233 (2011).

6. Garside, P. *et al.* Visualization of specific B and T lymphocyte interactions in the lymph node. *Science* **281,** 96–99 (1998).

7. Okada, T. *et al.* Antigen-engaged B cells undergo chemotaxis toward the T zone and form motile conjugates with helper T cells. *PLoS Biol.* **3,** e150 (2005).

8. Coffey, F., Alabyev, B. & Manser, T. Initial clonal expansion of germinal center B cells takes place at the perimeter of follicles. *Immunity* **30,** 599–609 (2009).

9. Allen, C. D. C., Okada, T. & Cyster, J. G. Germinal-center organization and cellular dynamics. *Immunity* **27,** 190–202 (2007).

10. Allen, C. D. C. *et al.* Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat. Immunol.* **5,** 943–952 (2004).

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

12. Victora, G. D. *et al.* Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter. *Cell* **143,** 592–605 (2010).

13. Victora, G. D. *et al.* Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. *Blood* **120,** 2240–2248 (2012).

14. McHeyzer-Williams, L. J., Milpied, P. J., Okitsu, S. L. & McHeyzer-Williams, M. G. Class-switched memory B cells remodel BCRs within secondary germinal centers. *Nat. Immunol.* **16,** 296–305 (2015).

15. Weigert, M. G., Cesari, I. M., Yonkovich, S. J. & Cohn, M. *Variability in the lambda light chain sequences of mouse antibody*. (Nature, 1970).

16. Berek, C. & Milstein, C. Mutation drift and repertoire shift in the maturation of the immune response. *Immunol. Rev.* **96,** 23–41 (1987).

17. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102,** 553–563 (2000).

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

19. Zan, H. & Casali, P. Regulation of Aicda expression and AID activity. *Autoimmunity* **46,** 83–101 (2013).

20. Honjo, T., Nagaoka, H., Shinkura, R. & Muramatsu, M. AID to overcome the limitations of genomic information - Nature Immunology. *Nat. Immunol.* (2005).

21. Schwickert, T. A. *et al.* In vivo imaging of germinal centres reveals a dynamic open structure. *Nature* **446,** 83–87 (2007).

22. Dominguez-Sola, D. *et al.* The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry. *Nat. Immunol.* **13,** 1083–1091 (2012).

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

24. Vallabhapurapu, S. & Karin, M. Regulation and function of NF-kappaB transcription factors in the immune system. *Annu. Rev. Immunol.* **27,** 693–733 (2009).

25. Sen, R. Control of B lymphocyte apoptosis by the transcription factor NF-kappaB. *Immunity* **25,** 871–883 (2006).

26. Shaffer, A. L. *et al.* Signatures of the immune response. *Immunity* **15,** 375–385 (2001).

27. Basso, K. *et al.* Tracking CD40 signaling during germinal center development. *Blood* **104,** 4088–4096 (2004).

28. Meyer-Hermann, M. *et al.* A theory of germinal center B cell selection, division, and exit. *Cell Rep* **2,** 162–174 (2012).

29. Ye, B. H. *et al.* The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. *Nat. Genet.* **16,** 161–170 (1997).

30. Dent, A. L., Shaffer, A. L., Yu, X., Allman, D. & Staudt, L. M. Control of inflammation, cytokine expression, and germinal center formation by BCL-6. *Science* **276,** 589–592 (1997).

31. Hatzi, K. & Melnick, A. Breaking bad in the germinal center: how deregulation of BCL6 contributes to lymphomagenesis. *Trends Mol Med* **20,** 343–352 (2014).

32. Kerfoot, S. M. *et al.* Germinal center B cell and T follicular helper cell development initiates in the interfollicular zone. *Immunity* **34,** 947–960 (2011).

33. Kitano, M. *et al.* Bcl6 protein expression shapes pre-germinal center B cell dynamics and follicular helper T cell heterogeneity. *Immunity* **34,** 961–972 (2011).

34. Lee, C. H. *et al.* Regulation of the germinal center gene program by interferon (IFN) regulatory factor 8/IFN consensus sequence-binding protein. *J. Exp. Med.* **203,** 63–72 (2006).

35. Ochiai, K. *et al.* Transcriptional regulation of germinal center B and plasma cell fates by dynamical control of IRF4. *Immunity* **38,** 918–929 (2013).

36. Ying, C. Y. *et al.* MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nat. Immunol.* **14,** 1084–1092 (2013).

37. Cattoretti, G. *et al.* BCL-6 protein is expressed in germinal-center B cells. *Blood* **86,** 45–53 (1995).

38. Allman, D. *et al.* BCL-6 expression during B-cell activation. *Blood* **87,** 5257–5268 (1996).

39. Fukuda, T. *et al.* Disruption of the Bcl6 gene results in an impaired

germinal center formation. *J. Exp. Med.* **186,** 439–448 (1997).

40. Toyama, H. *et al.* Memory B cells without somatic hypermutation are generated from Bcl6-deficient B cells. *Immunity* **17,** 329–339 (2002).

41. Hatzi, K., Jiang, Y., Huang, C. & Garrett-Bakelman, F. A Hybrid Mechanism of Action for BCL6 in B Cells Defined by Formation of Functionally Distinct Complexes at Enhancers and Promoters. *Cell Rep* (2013).

42. Huang, C. *et al.* The BCL6 RD2 domain governs commitment of activated B cells to form germinal centers. *Cell Rep* **8,** 1497–1508 (2014).

43. Huang, C., Hatzi, K. & Melnick, A. Lineage-specific functions of Bcl-6 in immunity and inflammation are mediated by distinct biochemical mechanisms : Nature Immunology : Nature Research. *Nat. Immunol.* (2013).

44. Pasqualucci, L. *et al.* Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood* **101,** 2914–2923 (2003).

45. Ci, W. *et al.* The BCL6 transcriptional program features repression of multiple oncogenes in primary B cells and is deregulated in DLBCL. *Blood* **113,** 5536–5548 (2009).

46. Basso, K. *et al.* Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. *Blood* **115,** 975–984 (2010).

47. Velichutina, I. *et al.* EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood* **116,** 5247–5255 (2010).

48. O'Carroll, D. *et al.* The Polycomb-Group GeneEzh2 Is Required for Early Mouse Development. *… and cellular biology* (2001).

49. Chen, Y.-H., Hung, M.-C. & Li, L.-Y. EZH2: a pivotal regulator in controlling cell differentiation. *Am J Transl Res* **4,** 364–375 (2012).

50. Su, I., Basavaraj, A., Krutchinsky, A. N. & Hobert, O. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement - Nature Immunology. *Nature* (2003).

51. van Galen, J. C. *et al.* Distinct expression patterns of polycomb oncoproteins and their binding partners during the germinal center reaction. *Eur. J. Immunol.* **34,** 1870–1881 (2004).

52. Raaphorst, F. M. *et al.* Cutting Edge: Polycomb Gene Expression Patterns Reflect Distinct B Cell Differentiation Stages in Human Germinal Centers. *The Journal of Immunology* **164,** 1–4 (2000).

53. Niu, H., Ye, B. H. & Dalla-Favera, R. Antigen receptor signaling induces MAP kinase-mediated phosphorylation and degradation of the BCL-6 transcription factor. *Genes & Development* **12,** 1953–1961 (1998).

54. Klein, U., Casola, S., Cattoretti, G., Shen, Q. & Lia, M. Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination - Nature Immunology. *… immunology* (2006).

55. Sciammas, R. *et al.* Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. *Immunity* **25,** 225–236 (2006).

56. Saito, M. *et al.* A signaling pathway mediating downregulation of BCL6 in germinal center B cells is blocked by BCL6 gene alterations in B cell lymphoma. *Cancer Cell* **12,** 280–292 (2007).

57. Tunyaplin, C. & Shaffer, A. L. Direct Repression of prdm1 by Bcl-6 Inhibits Plasmacytic Differentiation | The Journal of Immunology. *The Journal of …* (2004).

58. Lin, K.-I., Angelin-Duclos, C., Kuo, T. C. & Calame, K. Blimp-1-dependent repression of Pax-5 is required for differentiation of B cells to immunoglobulin M-secreting plasma cells. *Molecular and Cellular Biology* **22,** 4771–4780 (2002).

59. Shaffer, A. L. *et al.* Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17,** 51–62 (2002).

60. Shapiro-Shelef, M. *et al.* Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* **19,** 607–620 (2003).

61. Kallies, A. *et al.* Initiation of plasma-cell differentiation is independent of the transcription factor Blimp-1. *Immunity* **26,** 555–566 (2007).

62. Nutt, S. L., Hodgkin, P. D., Tarlinton, D. M. & Corcoran, L. M. The generation of antibody-secreting plasma cells. *Nature Publishing Group* **15,** 160–171 (2015).

63. MacLennan, I. C. Germinal centers. *Annu. Rev. Immunol.* **12,** 117–139 (1994).

64. Phan, R. T. & Dalla-Favera, R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. *Nature* **432,** 635–639 (2004).

65. Shaffer, A. L. *et al.* BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. *Immunity* **13,** 199–212 (2000).

66. Hu, B. T., Lee, S. C., Marin, E., Ryan, D. H. & Insel, R. A. Telomerase is upregulated in human germinal center B cells in vivo and can be re-expressed in memory B cells activated in vitro. *J. Immunol.* **159,** 1068–1071 (1997).

67. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403,** 503–511 (2000).

68. Klein, U. & Dalla-Favera, R. Germinal centres: role in B-cell physiology and malignancy. *Nature Publishing Group* **8,** 22–33 (2008).

69. Küppers, R., Klein, U., Hansmann, M. L. & Rajewsky, K. Cellular origin of human B-cell lymphomas. *N. Engl. J. Med.* **341,** 1520–1529 (1999).

70. Stevenson, F. K. *et al.* The occurrence and significance of V gene mutations in B cell-derived human malignancy. *Adv. Cancer Res.* **83,** 81–116 (2001).

71. Coiffier, B. Diffuse large cell lymphoma. *Curr Opin Oncol* **13,** 325–334 (2001).

72. Campo, E. *et al.* The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* **117,** 5019–5032 (2011).

73. Rosenwald, A. *et al.* The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* **346,** 1937–1947 (2002).

74. McKelvey, E. M., Gottlieb, J. A., Wilson, H. E. & Haut, A.

Hydroxyldaunomycin (adriamycin) combination chemotherapy in malignant lymphoma - McKelvey - 1976 - Cancer - Wiley Online Library. *Cancer* (1976).

75. Coiffier, B. *et al.* Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood* **116,** 2040–2045 (2010).

76. Friedberg, J. W. Relapsed/Refractory Diffuse Large B-Cell Lymphoma. *Hematology* **2011,** 498–505 (2011).

77. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412,** 341–346 (2001).

78. Liu, M. *et al.* Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451,** 841–845 (2008).

79. Müschen, M. *et al.* Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* **192,** 1833–1840 (2000).

80. Pasqualucci, L. *et al.* BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 11816–11821 (1998).

81. Shen, H. M., Peters, A., Baron, B., Zhu, X. & Storb, U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* **280,** 1750–1752 (1998).

82. Pasqualucci, L. *et al.* AID is required for germinal center-derived lymphomagenesis. *Nat. Genet.* **40,** 108–112 (2008).

83. Khodabakhshi, A. H. *et al.* Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget* **3,** 1308–1319 (2012).

84. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476,** 298–303 (2011).

85. Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* **43,** 830–837 (2011).

86. Lohr, J. G., Stojanov, P. & Lawrence, M. S. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. in (2012). doi:10.1073/pnas.1121343109/-

/DCSupplemental

87. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471,** 189–195 (2011).

88. Sneeringer, C. J. *et al.* Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 20980–20985 (2010).

89. Yap, D. B. *et al.* Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. *Blood* **117,** 2451–2459 (2011).

90. McCabe, M. T. *et al.* Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci. U.S.A.* **109,** 2989–2994 (2012).

91. van Kemenade, F. J. *et al.* Coexpression of BMI-1 and EZH2 polycomb-group proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood* **97,** 3896–3901 (2001).

92. Knutson, S. K., Wigle, T. J. & Warholic, N. M. A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells : Nature Chemical Biology : Nature Research. *Nature chemical …* (2012).

93. McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492,** 108–112 (2012).

94. Qi, W., Chan, H. M., Teng, L., Li, L. & Chuai, S. Selective inhibition of Ezh2 by a small molecule inhibitor blocks tumor cells proliferation. in (2012).

95. Newell-Price, J., Clark, A. & King, P. DNA Methylation and Silencing of Gene Expression - ScienceDirect. *Trends in Endocrinology & Metabolism* (2000).

96. Tsirigos, A., Ong, C. T., Low, H. M., Sung, K. & Rigoutsos, I. Dynamic changes in the human methylome during differentiation. (2010).

97. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian

development : Abstract : Nature Reviews Genetics. *Nat. Rev. Genet.* (2013).

98. Pollex, T. & Heard, E. Recent advances in X-chromosome inactivation research. *Curr. Opin. Cell Biol.* **24,** 825–832 (2012).

99. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 5237–5242 (2000).

100. Haines, T. R., Rodenhiser, D. I. & Ainsworth, P. J. Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. *Dev. Biol.* **240,** 585–598 (2001).

101. Dodge, J. E., Ramsahoye, B. H., Wo, Z. G., Okano, M. & Li, E. De novo methylation of MMLV provirus in embryonic stem cells: CpG versus non-CpG methylation. *Gene* **289,** 41–48 (2002).

102. Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. *Nucleic Acids Res.* **10,** 2709–2721 (1982).

103. Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40,** 91–99 (1985).

104. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 1412–1417 (2006).

105. Song, F. *et al.* Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 3336–3341 (2005).

106. Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **27,** 353–360 (2009).

107. Mohn, F. *et al.* Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30,** 755–766 (2008).

108. Li, E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3,** 662–673 (2002).

109. Bird, A. P. & Wolffe, A. P. Methylation-Induced Repression— Belts, Braces, and Chromatin - ScienceDirect. *Cell* (1999).

110. Reik, W. Epigenetic Reprogramming in Mammalian Development. *Science* **293,** 1089–1093 (2001).

111. Sanford, J. P., Clark, H. J., Chapman, V. M. & Rossant, J. Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes & Development* **1,** 1039–1046 (1987).

112. Monk, M., Boubelik, M. & Lehnert, S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* (1987).

113. Howlett, S. K. & Reik, W. Methylation levels of maternal and paternal genomes during preimplantation development. *Development* **113,** 119–127 (1991).

114. Kafri, T. *et al.* Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes & Development* **6,** 705–714 (1992).

115. Hattori, N. *et al.* Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J. Biol. Chem.* **279,** 17063–17069 (2004).

116. Hattori, N. *et al.* Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes Cells* **12,** 387–396 (2007).

117. Maatouk, D. M. *et al.* DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133,** 3411–3418 (2006).

118. Shen, L. *et al.* Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet* **3,** 2023–2036 (2007).

119. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39,** 457–466 (2007).

120. Straussman, R. *et al.* Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.* **16,** 564–571 (2009).

121. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation

during early mouse development. *Nat. Genet.* **42,** 1093–1100 (2010).

122. Arand, J. *et al.* In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet* **8,** e1002750 (2012).

123. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48,** 849–862 (2012).

124. Jaenisch, R. DNA methylation and imprinting: why bother? - ScienceDirect. *Trends in Genetics* (1997).

125. Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome : Abstract : Nature Reviews Genetics. *Nat. Rev. Genet.* (2001).

126. Stewart, C. L., Stuhlmann, H. & Jähner, D. De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. in (1982).

127. Lei, H. *et al.* De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122,** 3195–3205 (1996).

128. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo - ScienceDirect. *Dev. Biol.* (2002).

129. Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases - ScienceDirect. *J. Mol. Biol.* (1988).

130. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases - Nature Genetics. *Nat. Genet.* (1998).

131. Xie, S. *et al.* Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene* (1999).

132. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei - ScienceDirect. *Cell* (1992).

133. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA

methyltransferase gene results in embryonic lethality - ScienceDirect. *Cell* (1992).

134. Bestor, T. H. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *The EMBO Journal* **11,** 2611 (1992).

135. Chen, T., Ueda, Y., Xie, S. & Li, E. A Novel Dnmt3a Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Activede Novo Methylation. *Journal of Biological Chemistry* (2002).

136. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development - ScienceDirect. *Cell* (1999).

137. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129,** 1983–1993 (2002).

138. Liang, G. *et al.* Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Molecular and Cellular Biology* **22,** 480–491 (2002).

139. Sutter, D. & Doerfler, W. Methylation of integrated viral DNA sequences in hamster cells transformed by adenovirus 12. *Cold Spring Harb. Symp. Quant. Biol.* **44 Pt 1,** 565–568 (1980).

140. Sutter, D. & Doerfler, W. Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **77,** 253–256 (1980).

141. Vardimon, L., Neumann, R., Kuhlmann, I., Sutter, D. & Doerfler, W. DNA methylation and viral gene expression in adenovirus-transformed and -infected cells. *Nucleic Acids Res.* **8,** 2461–2473 (1980).

142. Kuhlmann, I. & Doerfler, W. Shifts in the extent and patterns of DNA methylation upon explanation and subcultivation of adenovirus type 12-induced hamster tumor cells. *Virology* **118,** 169–180 (1982).

143. Kruczek, I. & Doerfler, W. The unmethylated state of the promoter/leader and 5'-regions of integrated adenovirus genes correlates with gene expression. *The EMBO Journal* **1,** 409–414 (1982).

144. van der Ploeg, L. H. & Flavell, R. A. DNA methylation in the human gamma delta beta-globin locus in erythroid and nonerythroid tissues. *Cell* **19,** 947–958 (1980).

145. Smith, S. S., Yu, J. C. & Chen, C. W. Different levels of DNA modification at 5'CCGG in murine erythroleukemia cells and the tissues of normal mouse spleen. *Nucleic Acids Res.* **10,** 4305–4320 (1982).

146. Mandel, J. L. & Chambon, P. DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. *Nucleic Acids Res.* **7,** 2081–2103 (1979).

147. Weintraub, H., Larsen, A. & Groudine, M. Alpha-Globin-gene switching during the development of chicken embryos: expression and chromosome structure. *Cell* **24,** 333–344 (1981).

148. Desrosiers, R. C., Mulder, C. & Fleckenstein, B. Methylation of Herpesvirus saimiri DNA in lymphoid tumor cell lines. *Proc. Natl. Acad. Sci. U.S.A.* **76,** 3839–3843 (1979).

149. Cohen, J. C. Methylation of milk-borne and genetically transmitted mouse mammary tumor virus proviral DNA. *Cell* **19,** 653–662 (1980).

150. Guntaka, R. V., Rao, P. Y., Mitsialis, S. A. & Katz, R. Modification of avian sarcoma proviral DNA sequences in nonpermissive XC cells but not in permissive chicken cells. *J. Virol.* **34,** 569–572 (1980).

151. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).

152. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).

153. Zhang, Y. *et al.* QDMR: a quantitative method for identification of differentially methylated regions by entropy. *Nucleic Acids Res.* **39,** e58 (2011).

154. Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **13,** R83 (2012).

155. Varley, K. E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23,** 555–567 (2013).

156. Warnecke, P. M. & Clark, S. J. DNA methylation profile of the mouse skeletal alpha-actin promoter during development and differentiation. *Molecular and Cellular Biology* **19,** 164–172 (1999).

157. Walsh, C. P. & Bestor, T. H. Cytosine methylation and mammalian development. *Genes & Development* **13,** 26–34 (1999).

158. Keshet, I. *et al.* Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* **38,** 149–153 (2006).

159. Sproul, D. *et al.* Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 4364–4369 (2011).

160. Sproul, D. *et al.* Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. *Genome Biol.* **13,** R84 (2012).

161. Sproul, D. & Meehan, R. R. Genomic insights into cancer-associated aberrant CpG island hypermethylation. *Brief Funct Genomics* **12,** 174–190 (2013).

162. Ball, M. P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27,** 361–368 (2009).

163. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites - ScienceDirect. *Trends in Genetics* (1997).

164. Xu, G. L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402,** 187–191 (1999).

165. Gonzalo, S. *et al.* DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat. Cell Biol.* **8,** 416–424 (2006).

166. Guo-Liang, X., Bestor, T. H., Bourc'his, D. & Chih-Lin, H. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* (1999).

167. Guo, G., Wang, W. & Bradley, A. Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature* **429,** 891–895 (2004).

168. Kim, M., Trinh, B. N., Long, T. I., Oghamian, S. & Laird, P. W. Dnmt1 deficiency leads to enhanced microsatellite instability in mouse embryonic stem cells. *Nucleic Acids Res.* **32,** 5742–5749 (2004).

169. Wang, K.-Y. & James Shen, C.-K. DNA methyltransferase Dnmt1 and mismatch repair. *Oncogene* **23,** 7898–7902 (2004).

170. Dion, V., Lin, Y., Hubert, L., Waterland, R. A. & Wilson, J. H. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Human Molecular Genetics* **17,** 1306–1317 (2008).

171. Loughery, J. E. P. *et al.* DNMT1 deficiency triggers mismatch repair defects in human cells through depletion of repair protein levels in a process involving the DNA damage response. *Human Molecular Genetics* **20,** 3241–3255 (2011).

172. Karpf, A. R. & Matsui, S.-I. Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Research* **65,** 8635–8639 (2005).

173. Ji, H. *et al.* Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467,** 338–342 (2010).

174. Bock, C. *et al.* DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol. Cell* **47,** 633–647 (2012).

175. Bröske, A.-M. *et al.* DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat. Genet.* **41,** 1207–1215 (2009).

176. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44,** 23–31 (2011).

177. Challen, G. A. *et al.* Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell Stem Cell* **15,** 350–364 (2014).

178. Deaton, A. M. *et al.* Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Res.* **21,** 1074–1086 (2011).

179. Shaknovich, R. *et al.* DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* **118,** 3559–3569 (2011).

180. Berdasco, M. & Esteller, M. Aberrant Epigenetic Landscape in Cancer: How Cellular Identity Goes Awry. *Developmental Cell* **19,** 698–711 (2010).

181. Amara, K. *et al.* DNA methyltransferase DNMT3b protein overexpression as a prognostic factor in patients with diffuse large B-cell lymphomas. *Cancer Sci.* **101,** 1722–1730 (2010).

182. Bennett, L. B. *et al.* DNA hypermethylation accompanied by transcriptional repression in follicular lymphoma. *Genes Chromosomes Cancer* **48,** 828–841 (2009).

183. O'Riain, C. *et al.* Array-based DNA methylation profiling in follicular lymphoma. *Leukemia* **23,** 1858–1866 (2009).

184. Martín-Subero, J. I. *et al.* New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood* **113,** 2488–2497 (2009).

185. Shaknovich, R. *et al.* DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood* **116,** e81–9 (2010).

186. Choi, J.-H. *et al.* Genome-Wide DNA Methylation Maps in Follicular Lymphoma Cells Determined by Methylation-Enriched Bisulfite Sequencing. *PLoS ONE* **5,** e13020 (2010).

187. Leshchenko, V. V. *et al.* Genomewide DNA methylation analysis reveals novel targets for drug development in mantle cell lymphoma. *Blood* **116,** 1025–1034 (2010).

188. Eberle, F. C. *et al.* Methylation profiling of mediastinal gray zone lymphoma reveals a distinctive signature with elements shared by classical Hodgkin's lymphoma and primary mediastinal large B-cell lymphoma. *Haematologica* **96,** 558–566 (2011).

189. De, S. *et al.* Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet* **9,** e1003137 (2013).

190. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135,** 1201–1212 (2008).

191. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463,** 1042–1047 (2010).

192. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463,** 1101–1105 (2010).

193. Kumar, R. *et al.* AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature* **500,** 89–92 (2013).

194. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251–260 (1997).

195. Whitlock, J. P. & Simpson, R. T. Localization of the sites along nucleosome DNA which interact with NH2-terminal histone regions. *J. Biol. Chem.* **252,** 6516–6520 (1977).

196. Ausio, J., Dong, F. & van Holde, K. E. Use of selectively trypsinized nucleosome core particles to analyze the role of the histone 'tails' in the stabilization of the nucleosome. *J. Mol. Biol.* **206,** 451–463 (1989).

197. Wolffe, A. P. & Pruss, D. Deviant nucleosomes: the functional specialization of chromatin. *Trends Genet.* **12,** 58–62 (1996).

198. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14,** 1025–1040 (2007).

199. Voigt, P. & Reinberg, D. Histone tails: ideal motifs for probing epigenetics through chemical biology approaches. *Chembiochem* **12,** 236–252 (2011).

200. Probst, A. V., Dunleavy, E. & Almouzni, G. Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.* **10,** 192–206 (2009).

201. Margueron, R. & Reinberg, D. Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* **11,** 285–296 (2010).

202. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447,** 407–412 (2007).

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

204. Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the

Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38,** 413–443 (2004).

205. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128,** 735–745 (2007).

206. Shilatifard, A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* **81,** 65–95 (2012).

207. Santos-Rosa, H. *et al.* Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol. Cell* **12,** 1325–1332 (2003).

208. Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433,** 434–438 (2005).

209. Sims, R. J. *et al.* Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280,** 41789–41792 (2005).

210. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121,** 859–872 (2005).

211. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469,** 343–349 (2011).

212. Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell* **49,** 808–824 (2013).

213. Cao, R. *et al.* Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298,** 1039–1043 (2002).

214. Czermin, B., Melfi, R., McCabe, D., Seitz, V. & Imhof, A. Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites - ScienceDirect. *Cell* (2002).

215. Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L. & Sengupta, A. Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. *Cell* (2002).

216. Gao, Z. *et al.* PCGF Homologs, CBX Proteins, and RYBP Define Functionally Distinct PRC1 Family Complexes. *Mol. Cell* (2012).

217. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148,** 664–678 (2012).

218. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4,** e1000242 (2008).

219. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349–353 (2006).

220. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & Development* **24,** 265–276 (2010).

221. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).

222. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125,** 301–313 (2006).

223. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1,** 299–312 (2007).

224. Zhao, X. D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1,** 286–298 (2007).

225. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells - ScienceDirect. *Cell* **125,** 315–326 (2006).

226. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8,** 532–538 (2006).

227. Voigt, P., Tee, W.-W. & Reinberg, D. A double take on bivalent promoters. *Genes & Development* **27,** 1318–1338 (2013).

228. Chase, A. & Cross, N. C. P. Aberrations of EZH2 in cancer. *Clin.*

Cancer Res. **17,** 2613–2618 (2011).

229. Mahmoudi, T. & Verrijzer, C. P. Chromatin silencing and activation by Polycomb and trithorax group proteins. *Oncogene* **20,** 3055–3066 (2001).

230. Akalin, A. *et al.* Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet* **8,** e1002781 (2012).

231. Lai, A. Y. *et al.* 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 (2013).

232. Fritz, E. L. *et al.* A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat. Immunol.* **14,** 749–755 (2013).

233. McKean, D. *et al.* Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. U.S.A.* **81,** 3180–3184 (1984).

234. Robbiani, D. F. *et al.* AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell* **135,** 1028–1038 (2008).

235. Y Chang, C. J. P. G. E. W. Enumeration and characterization of DJH structures in mouse fetal liver. *The EMBO Journal* **11,** 1891 (1992).

236. Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors - ProQuest. *Nature* (2007).

237. Schlissel, M. S. & Corcoran, L. M. Virus-transformed pre-B cells show ordered activation but not inactivation of immunoglobulin gene rearrangement and transcription. | JEM. *Journal of Experimental …* (1991).

238. Meng, F.-L. *et al.* Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* **159,** 1538–1548 (2014).

239. Barlow, J. H. *et al.* Identification of early replicating fragile sites that contribute to genome instability. *Cell* **152,** 620–632 (2013).

240. Goodier, J. L. & Kazazian, H. H. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135,** 23–35 (2008).

241. Ryan, F. P. Human endogenous retroviruses in health and disease: a symbiotic perspective. *J R Soc Med* **97,** 560–565 (2004).

242. Stocking, C. & Kozak, C. A. Endogenous retroviruses. *Cell. Mol. Life Sci.* **65,** 3383–3398 (2008).

243. He, X. *et al.* Ki-67 is a valuable prognostic predictor of lymphoma but its utility varies in lymphoma subtypes: evidence from a systematic meta-analysis. *BMC Cancer* **14,** 153 (2014).

244. Pasqualucci, L. Expression of the AID protein in normal and neoplastic B cells. *Blood* **104,** 3318–3325 (2004).

245. Lossos, I. S., Levy, R. & Alizadeh, A. A. AID is expressed in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas and is not correlated with intraclonal heterogeneity. *Leukemia* **18,** 1775–1779 (2004).

246. Bird, A. DNA methylation patterns and epigenetic memory. *Genes & Development* **16,** 6–21 (2002).

247. Jones, P. A. The Role of DNA Methylation in Mammalian Epigenetics. *Science* **293,** 1068–1070 (2001).

248. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* **46,** 17–23 (2013).

249. Kulis, M. *et al.* Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nat. Genet.* **44,** 1236–1242 (2012).

250. Mayle, A., Yang, L., Rodriguez, B., Zhou, T. & Chang, E. Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation | Blood Journal. *Blood* (2015).

251. Zhang, J. *et al.* The genomic landscape of mantle cell lymphoma is related to the epigenetically determined chromatin state of normal B cells | Blood Journal. *Blood* (2014).

252. Ha, K. *et al.* Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Human Molecular Genetics*

20, 126–140 (2010).

253. Zan, H. & Casali, P. AID- and Ung-dependent generation of staggered double-strand DNA breaks in immunoglobulin class switch DNA recombination: A post-cleavage role for AID. *Molecular immunology* (2008).

254. Maul, R. W., Cao, Z. & Venkataraman, L. Spt5 accumulation at variable genes distinguishes somatic hypermutation in germinal center B cells from ex vivo–activated cells | JEM. *Journal of …* (2014).

255. Hogenbirk, M. A. *et al.* Differential programming of B cells in AID deficient mice. *PLoS ONE* **8,** e69815 (2013).

256. Rodriguez, B. A. T. *et al.* Methods for high-throughput MethylCap-Seq data analysis. *BMC Genomics* **13 Suppl 6,** S14 (2012).

257. Franchini, D.-M. *et al.* Processive DNA demethylation via DNA deaminase-induced lesion resolution. *PLoS ONE* **9,** e97754 (2014).

258. Gaudet, F., Hodgson, J. G. & Eden, A. Induction of Tumors in Mice by Genomic Hypomethylation | Science. *…* (2003).

259. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* (2013).

260. Chambwe, N. *et al.* Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes. *Blood* **123,** 1699–1708 (2014).

261. Pan, H. *et al.* Epigenomic evolution in diffuse large B-celllymphomas. *Nature Communications* **6,** 1–12 (2015).

262. Nonaka, T. *et al.* Involvement of activation-induced cytidine deaminase in skin cancer development. *J. Clin. Invest.* **126,** 1367–1382 (2016).

263. Sawai, Y. *et al.* Activation-Induced Cytidine Deaminase Contributes to Pancreatic Tumorigenesis by Inducing Tumor-Related Gene Mutations. *Cancer Research* **75,** 3292–3301 (2015).

264. Jiang, Y., Dominguez, P. M. & Melnick, A. M. The many layers of epigenetic dysfunction in B-cell lymphomas. *Current Opinion in Hematology* **23,** 377–384 (2016).

265. Landau, D. A., Carter, S. L., Getz, G. & Wu, C. J. Clonal evolution in

hematological malignancies and therapeutic implications. *Leukemia* **28,** 34–43 (2014).

266. Mazor, T., Pankov, A., Song, J. S. & Costello, J. F. Intratumoral Heterogeneity of the Epigenome. *Cancer Cell* (2016).

267. Landau, D. A. *et al.* Locally Disordered Methylation Forms the Basis of Intratumor Methylome Variation in Chronic Lymphocytic Leukemia. *Cancer Cell* (2014).

268. Li, S. *et al.* Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat. Med.* **22,** 792–799 (2016).

269. Hansen, K. D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* **43,** 768–775 (2011).

270. Timp, W. & Feinberg, A. P. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat. Rev. Cancer* **13,** 497–510 (2013).

271. Easwaran, H., Tsai, H.-C. & Baylin, S. B. Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. *Mol. Cell* **54,** 716–727 (2014).

272. Feinberg, A. P. & Irizarry, R. A. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. in (2010).

273. Van Parijs, L., Refaeli, Y., Lord, J. D., Nelson, B. H. & Abbas, A. K. RETRACTED: Uncoupling IL-2 Signals that Regulate T Cell Proliferation, Survival, and Fas-Mediated Activation-Induced Cell Death - ScienceDirect. *Immunity* (1999).

274. Vuong, B. Q., Lee, M., Kabir, S., Irimia, C. & Macchiarulo, S. Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination : Article : Nature Immunology. *Nature* (2009).

275. Ogilvy, S., Metcalf, D. & Bath, M. L. Constitutive Bcl-2 expression throughout the hematopoietic compartment affects multiple lineages and enhances progenitor cell survival. in (1999).

276. Ortega-Molina, A. *et al.* The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma

development. *Nat. Med.* **21,** 1199–1208 (2015).

277. Giannopoulou, E. G. & Elemento, O. An integrated ChIP-seq analysis platform with customizable workflows. *BMC Bioinformatics* **12,** 277 (2011).

278. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

279. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108 (2013).

280. Cardenas, M. G. *et al.* Rationally designed BCL6 inhibitors target activated B cell diffuse large B cell lymphoma. *J. Clin. Invest.* **126,** 3351–3362 (2016).

281. Jiang, Y. *et al.* Deep sequencing reveals clonal evolution patterns and mutation events associated with relapse in B-cell lymphomas. *Genome Biol.* **15,** 432 (2014).

282. Casola, S. *et al.* Tracking germinal center B cells expressing germ-line immunoglobulin gamma1 transcripts by conditional gene targeting. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 7396–7401 (2006).

283. Rui, L., Schmitz, R., Ceribelli, M. & Staudt, L. M. Malignant pirates of the immune system : Nature Immunology : Nature Research. *Nat. Immunol.* (2011).

284. Blackman, M. A., Tigges, M. A., Minie, M. E. & Koshland, M. E. A model system for peptide hormone action in differentiation: Interleukin 2 induces a B lymphoma to transcribe the J chain gene - ScienceDirect. *Cell* (1986).

285. Schwartz, Y. B. & Pirrotta, V. Polycomb silencing mechanisms and the management of genomic programmes : Abstract : Nature Reviews Genetics. *Nat. Rev. Genet.* (2007).

286. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12,** 7–18 (2011).

287. Zhang, Y., Liu, T. & Meyer, C. A. Model-based Analysis of ChIP-Seq (MACS) | Genome Biology | Full Text. *Genome …* (2008).

288. Morin, R. D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* **42,** 181–185 (2010).

289. Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10,** 697–708 (2009).

290. Cattoretti, G. *et al.* Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice. *Cancer Cell* **7,** 445–455 (2005).

291. Ci, W., Polo, J. M. & Melnick, A. B-cell lymphoma 6 and the molecular pathogenesis of diffuse large B-cell lymphoma. *Current Opinion in Hematology* (2008).

292. Ye, B. H., Chaganti, S., Chang, C. C. & Niu, H. Chromosomal translocations cause deregulated BCL6 expression by promoter substitution in B cell lymphoma. *The EMBO …* (1995).

293. Gearhart, M. D., Corcoran, C. M., Wamstad, J. A. & Bardwell, V. J. Polycomb Group and SCF Ubiquitin Ligases Are Found in a Novel BCOR Complex That Is Recruited to BCL6 Targets. *Molecular and Cellular Biology* **26,** 6880–6889 (2006).

294. Ghetu, A. F., Corcoran, C. M., Cerchietti, L. & Bardwell, V. J. Structure of a BCOR Corepressor Peptide in Complex with the BCL6 BTB Domain Dimer - ScienceDirect. *Mol. Cell* (2008).

295. Huynh, K. D., Fischle, W., Verdin, E. & Bardwell, V. J. BCoR, a novel corepressor involved in BCL-6 repression. *Genes & …* (2000).

296. Stogios, P. J., Downs, G. S., Jauhal, J. J. S., Nandra, S. K. & Privé, G. G. Sequence and structural analysis of BTB domain proteins. *Genome Biol.* **6,** R82 (2005).

297. Vandamme, J., Völkel, P., Rosnoblet, C., Le Faou, P. & Angrand, P.-O. Interaction Proteomics Analysis of Polycomb Proteins Defines Distinct PRC1 Complexes in Mammalian Cells. *Mol Cell Proteomics* **10,** M110.002642 (2011).

298. Creppe, C., Palau, A., Malinverni, R. & Valero, V. A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. *PLoS …* (2014).

299. Sánchez, C. *et al.* Proteomics Analysis of Ring1B/Rnf2 Interactors Identifies a Novel Complex with the Fbxl10/Jhdm1B Histone Demethylase and the Bcl6 Interacting Corepressor. *… & cellular proteomics* (2007).

300. Wang, R., Taylor, A. B., Leal, B. Z., Chadwell, L. V. & Ilangovan, U. Polycomb Group Targeting through Different Binding Partners of RING1B C-Terminal Domain - ScienceDirect. *Structure* (2010).

301. Tan, J., Jones, M., Koseki, H., Nakayama, M. & Muntean, A. G. CBX8, a Polycomb Group Protein, Is Essential for MLL-AF9-Induced Leukemogenesis. *Cancer Cell* (2011).

302. Shin, D.-M. *et al.* Global Gene Expression Analysis of Very Small Embryonic-Like Stem Cells Reveals that the Ezh2-Dependent Bivalent Domain Mechanism Contributes to Their Pluripotent State. *Stem Cells and Development* **21,** 1639–1652 (2012).

303. Taniguchi, H., Jacinto, F. V., Villanueva, A. & Fernandez, A. F. Silencing of Kruppel-like factor 2 by the histone methyltransferase EZH2 in human cancer. *Oncogene* (2012).

304. Yu, J. *et al.* Integrative Genomics Analysis Reveals Silencing of β-Adrenergic Signaling by Polycomb in Prostate Cancer. *Cancer Cell* (2007).

305. Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nat. Rev. Cancer* **12,** 599–612 (2012).

306. Egle, A., Harris, A. W., Bath, M. L., O'Reilly, L. & Cory, S. VavP-Bcl2 transgenic mice develop follicular lymphoma preceded by germinal center hyperplasia | Blood Journal. *Blood* (2004).

307. Kaustov, L. *et al.* Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* **286,** 521–529 (2011).

308. Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P. & Brockdorff, N. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet* **9,** e1003717 (2013).

309. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* **21,** 569–571 (2014).

310. Blackledge, N. P., Farcas, A. M., Kondo, T. & King, H. W. Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation. *Cell* (2014).

311. Cooper, S., Dienstbier, M., Hassan, R. & Schermelleh, L. Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment - ScienceDirect. *Cell Rep* (2014).

312. Yuan, W. *et al.* H3K36 Methylation Antagonizes PRC2-mediated H3K27 Methylation. *J. Biol. Chem.* **286,** 7983–7989 (2011).

313. Farcas, A. M. *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1,** e00205 (2012).

314. Wu, X., Johansen, J. V. & Helin, K. Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol. Cell* **49,** 1134–1146 (2013).

315. Geng, H., Hurtz, C., Lenz, K. B., Chen, Z. & Baumjohann, D. Self-Enforcing Feedback Activation between BCL6 and Pre-B Cell Receptor Signaling Defines a Distinct Subtype of Acute Lymphoblastic Leukemia. *Cancer Cell* (2015).

316. Morse, H. C. Bethesda proposals for classification of lymphoid neoplasms in mice. *Blood* **100,** 246–258 (2002).

317. Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33,** e175 (2005).

318. Zheng, Y. *et al.* Total kinetic analysis reveals how combinatorial methylation patterns are established on lysines 27 and 36 of histone H3. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 13549–13554 (2012).

319. Trinkle-Mulcahy, L., Boulon, S. & Lam, Y. W. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes | JCB. *The Journal of cell …* (2008).

320. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data | Bioinformatics | Oxford Academic. *Bioinformatics* (2010).

321. Subramanian, A. & Tamayo, P. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression

profiles. in (2005).

322. Goodarzi, H., Elemento, O. & Tavazoie, S. Revealing Global Regulatory Perturbations across Human Cancers. *Mol. Cell* (2009).

323. Shaffer, A. L., Wright, G., Yang, L. & Powell, J. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunological …* (2006).

324. Elemento, O., Slonim, N. & Tavazoie, S. A Universal Framework for Regulatory Element Discovery across All Genomes and Data Types. *Mol. Cell* (2007).

325. Pikman, Y., Lee, B. H., Mercher, T. & McDowell, E. MPLW515L Is a Novel Somatic Activating Mutation in Myelofibrosis with Myeloid Metaplasia. *PLoS …* (2006).

326. Silverman, J., Suckow, M. A. & Murthy, S. *The IACUC Handbook, Third Edition - Google Books*. (2014).

327. Hanna, J., Markoulaki, S., Schorderet, P. & Carey, B. W. Direct Reprogramming of Terminally Differentiated Mature B Lymphocytes to Pluripotency - ScienceDirect. *Cell* (2008).

328. Verma, S. K. *et al.* Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Med Chem Lett* **3,** 1091–1096 (2012).

329. Cerchietti, L. C. *et al.* A small-molecule inhibitor of BCL6 kills DLBCL cells in vitro and in vivo. *Cancer Cell* **17,** 400–411 (2010).