HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2016 December 08.

Published in final edited form as:

Nat Med. 2015 October ; 21(10): 1190–1198. doi:10.1038/nm.3940.

Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis

Jiyuan Zhang[1], **David Dominguez-Sola**[1,2,3], **Shafinaz Hussein**[1,^], **Ji-Eun Lee**[4], **Antony B. Holmes**[1], **Mukesh Bansal**[5], **Sofija Vlasevska**[1], **Tongwei Mo**[1], **Hongyan Tang**[1], **Katia Basso**[1,6], **Kai Ge**[4], **Riccardo Dalla-Favera**[1,6,7,8], and **Laura Pasqualucci**[1,6,*]

[1]Institute for Cancer Genetics and Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA

[2]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[4]Laboratory of Endocrinology and Receptor Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

[5]Department of Systems Biology and Herbert Irving Comprehensive Cancer Center, Columbia University, New York, NY 10032, USA

[6]Department of Pathology and Cell Biology, Columbia University, New York, NY, USA

[7]Department of Genetics & Development, Columbia University, New York, NY, USA

[8]Department of Microbiology & Immunology, Columbia University, New York, NY, USA

Abstract

Mutations in the gene encoding the KMT2D (also called MLL2) methyltransferase are highly recurrent and occur early during tumorigenesis in diffuse large B cell lymphoma (DLBCL) and follicular lymphoma (FL). However, the functional consequences of *KMT2D* mutations and their role in lymphomagenesis are unknown. Here we show that FL/DLBCL-associated *KMT2D* mutations impair KMT2D enzymatic activity, leading to diminished global H3K4 methylation in

*Corresponding author: Laura Pasqualucci (lp171@columbia.edu).

^Current address: Department of Pathology and Laboratory Medicine, North Shore LIJ, Staten Island University Hospital, Staten Island, NY, USA

Author Contributions: L.P. designed and directed the study, analyzed data and wrote the manuscript, with contributions from R.D-F. and J.Z. J.Z. performed experiments, analyzed data and prepared figures. D.D.-S. and S.H. were responsible for histopathological analysis. J-E.L. and K.G. generated the *Kmt2d* conditional knock-out mouse model. K.B. performed the ChIP-Seq experiment and, together with A.B-H, contributed to the implementation of tools for the analysis of ChIP-Seq data. S.V. performed mutation analysis of the rearranged immunoglobulin genes from mouse tumors and provided technical support in various experiments. M.B. conducted supervised analysis and extended GSEA analysis of gene expression data. T.M. was responsible for animal husbandry. H.T. performed mouse autopsies and processed tissues for FACS and histologic analysis. R.D. F. contributed to the study design, data analysis and writing the manuscript. All authors reviewed the manuscript and provided final approval for submission.

Competing financial interests: The Authors declare no competing financial interests.

Accession codes. The gene expression profile data of mouse GC B cells and the ChIP-seq data of human GC B cells have been deposited in the GEO database under accession no. GSE67388 and GSE67494.

germinal-center (GC) B-cells and DLBCL cells. Conditional deletion of *Kmt2d* early during B cell development, but not after initiation of the GC reaction, results in an increase in GC B-cells and enhances B cell proliferation in mice. In mice overexpressing BCL2, which develop GC-derived lymphomas resembling human tumors, genetic ablation of *Kmt2d* leads to a further increase in tumor incidence. These findings suggest that *KMT2D* acts as a tumor suppressor gene whose early loss facilitates lymphomagenesis by remodeling the epigenetic landscape of the cancer precursor cells. Eradication of KMT2D-deficient cells may represent a rational therapeutic approach for targeting early tumorigenic events.

B cell non-Hodgkin lymphomas (B-NHL) represent a heterogeneous group of malignancies that originate mostly from B cells in the germinal center (GC) and are driven by distinct genetic lesions disrupting key oncogenic pathways[1,2]. Recent exome/transcriptome sequencing efforts have revealed recurrent mutations in epigenetic modifiers, including methyltransferases, acetyltransferases, and histone proteins themselves, suggesting that perturbations of epigenetic mechanisms play critical roles in lymphomagenesis[3-8]. Among these genes, *KMT2D* (also known as *MLL2/ALR/MLL4*) is the single most frequently mutated gene, being found in ~30% of *de novo* DLBCL (including both molecular subtypes, GCB- and ABC-DLBCL)[9] and ~90% of FL[3,5-8,10,11], which together account for over 70% of all B-NHL diagnoses. Moreover, recent studies investigating the history of clonal evolution during histologic transformation of FL to DLBCL (also called transformed FL, tFL) revealed that mutations in *KMT2D* represent early events introduced in a common ancestor before divergent evolution to FL or tFL through the acquisition of additional genetic lesions and final clonal expansion in the GC[7,8,10,11].

KMT2D encodes a highly conserved protein belonging to the SET1 family of histone lysine methyltransferases (KMT), a group of enzymes that catalyze the methylation of lysine 4 on histone H3 (H3K4) associated with transcriptionally active chromatin[12-14]. The enzymatic function of KMT2D depends on a cluster of C-terminal conserved domains, including a PHD domain, two FY-rich motifs (FYRC and FYRN) and a catalytic SET domain. While, in yeast, a single multi-subunit complex (also known as COMPASS) is responsible for all methylation of H3K4[15-18], six different KMTs have been identified in higher eukaryotes, which fall into three subgroups, based on homologies in protein sequence and subunit composition: SET1A/SET1B, MLL1/MLL4 (KMT2A/B), and MLL3/MLL2 (KMT2C/D)[12-14]. These findings suggest that the three KMT complexes exert non-overlapping, highly specialized functions by regulating the transcription of discrete subsets of genes. In particular, KMT2C/D function as major histone H3K4 mono- and di-methyltransferases at enhancers in *Drosophila*[19] and mammalian cells [20-23]. Consistent with the involvement of KMT2D in critical cellular functions, germline homozygous deletion of this gene in mice results in embryonic lethality[21].

In DLBCL and FL, *KMT2D* mutations are predominantly represented by premature stop codons, frameshift insertions/deletions and splice-site mutations that are predicted to generate truncated proteins lacking part or all of the C-terminal protein domains[3,5]. Additionally, multiple missense mutations have been found across the KMT2D protein, but their functional consequences remain unexplored. In 30–75% of the affected cases, *KMT2D*

genetic lesions are biallelically distributed, while the remaining ones retain one intact *KMT2D* allele, suggesting that this gene may function as a haploinsufficient tumor suppressor in at least a subset of cases. Indeed, monoallelic truncating mutations of *KMT2D* are considered the causative event in a rare congenital disease known as Kabuki syndrome, offering direct proof for the dose-dependent pathogenic effect of this enzyme in other tissues[24].

A few studies have investigated the biochemical function of KMT2D in mammals (during mouse adipogenesis and myogenesis, or in human colon cancer cell lines and haematopoietic cells, among others)[20-22,25,26]; however, little is known about the general role of this protein and its mutant alleles in B cells, and the mechanisms by which *KMT2D* mutations contribute to lymphoma development. Here we performed a comprehensive characterization of the mechanisms (genetic and epigenetic) that disrupt KMT2D function in B-NHL, and explored its role in normal B cell development and lymphomagenesis in mice.

Results

Genetic and epigenetic inactivation of *KMT2D* in DLBCL

We first characterized the mRNA expression pattern of KMT2D in healthy mouse and human mature B cell subpopulations. Consistent with the ubiquitous nature of other MLL family members, KMT2D transcripts were detected in purified naïve, GC and memory B cells (Supplementary Fig.1). Accordingly, co-immunofluorescence analysis of KMT2D and the GC-specific marker BCL6 in reactive human tonsils revealed positive KMT2D staining in the nuclei of all mature B cell compartments, including the GC (Fig. 1a).

We next used immunoblot analysis and immunohistochemistry (IHC) to measure the abundance of KMT2D in 123 DLBCL samples (26 cell lines and 97 primary biopsies) representative of various *KMT2D* genomic configurations (biallelic truncating mutations, heterozygous truncating or missense mutations, and wild-type alleles). We used an antibody that specifically recognizes an epitope in the C-terminal portion of the KMT2D protein, thus allowing for identification of the full-length polypeptide (see Supplementary Fig. 2 for antibody characterization). As expected, all samples carrying biallelic truncating mutations that eliminate this epitope lacked expression of the intact protein (Fig. 1b,c). Moreover, immunoblot analysis using antibodies directed against the N-terminal half of KMT2D failed to detect any signal corresponding to the truncated proteins in all cell lines tested, despite expression of the mutant cDNA (Supplementary Fig. 2e,f). Significantly lower to undetectable levels were also observed following treatment with the protease inhibitor MG132, suggesting that mechanisms other than proteosomal degradation are responsible for the lack of expression of the truncated proteins in these cells (data not shown).

We then investigated whether the residual wild-type allele is still expressed in samples carrying *KMT2D* monoallelic (truncating) mutations. While the KMT2D protein was barely detectable in 2/18 samples, the remaining cell line and all 15 biopsies analyzed stained positive for the full-length protein (Fig. 1b–d). Of note, significantly reduced or absent protein amounts were also observed in 9/77 (11.7%) samples lacking *KMT2D* genetic lesions (Fig. 1b–d), suggesting that additional epigenetic mechanisms may interfere with

KMT2D protein expression in a small subset of tumors. Overall, approximately 45% of DLBCLs lack KMT2D either completely, due to biallelic genetic lesions/loss of protein expression (20% of cases), or partially, due to monoallelic mutations (25% of cases) (Fig. 1e).

Missense mutations impair KMT2D methyl-transferase activity

Most *KMT2D* mutations are postulated to generate truncated proteins that are functionally defective due to the loss of the catalytic SET domain. However, ~7.3% of *de novo* DLBCLs and 23.0% of tFLs in our dataset harbor missense mutations affecting various residues along the KMT2D protein (18/89 mutational events, 20.2%; and 33/182 events from published DLBCL/tFL genomic studies, 18.1%) (Fig. 2a and Supplementary Fig. 3)[3,5-8].

To test the functional impact of *KMT2D* missense mutations, we generated HA-tagged versions of 16 DLBCL-derived mutant alleles and measured their basal expression upon transient transfection in 293T cells. With one exception (S5404F), all alleles tested produced similar amounts of both mRNA and protein as the wild-type allele or three germline variants included as controls, indicating that missense mutations generally do not interfere with the stability of the KMT2D protein (Supplementary Fig. 4).

We then measured the ability of these proteins to catalyze methylation of H3K4 *in vitro*, using an artificial H3 peptide substrate or purified nucleosomal histones[27]. In both assays, all nine mutations affecting the C-terminal domains significantly impaired enzymatic activity, compared to the wild-type protein (Fig. 2c,d). In particular, immunoblot analysis of nucleosomal histones with specific antibodies recognizing distinct methylation states revealed a significant reduction in H3K4 mono, di-, and trimethylation (Fig. 2d). In contrast, the enzymatic activity of seven polypeptides carrying mutations in the N-terminal half of the protein was only modestly affected or indistinguishable from the wild-type control, as were the germline mutants (Fig. 2b–d). These amino-acid changes are located outside known functional domains and, for five of them, no matched normal DNA was available to document their somatic origin, raising the possibility that they represent private germline variants not currently annotated in mutation databases. Alternatively, these changes may interfere with other modes of KMT2D regulation. While additional studies will be required to dissect the precise mechanisms underlying the observed reduced activity of the C-terminal mutants (e.g. substrate preference or kinetic characteristics), these data show that human KMT2D is capable of mono-, di-, and trimethylating H3K4 *in vitro*, and provide direct experimental evidence that the missense mutations commonly affecting its C-terminal enzymatic domains are loss-of-function events.

Loss of *KMT2D* leads to reduced H3K4 methylation

To investigate whether loss of KMT2D function is sufficient to alter global H3K4 methylation *in vivo*, we examined splenic $B220^+$ and GC B cells purified from mice (*KMT2D*$^{fl/fl}$*CD19-Cre* and *KMT2D*$^{fl/fl}$*Cγ1-Cre*) where the *Kmt2d* gene had been conditionally deleted via Cre-mediated recombination (see next paragraph). Compared to wild-type littermates, *Kmt2d*-deficient B cells exhibited significantly reduced levels of H3K4me1, H3K4me2 and H3K4me3 (Fig. 2e). Moreover, all three epigenetic marks were

reduced in four DLBCL cell lines that lack KMT2D protein expression due to the presence of biallelic truncating mutations; the degree of reduction was variable depending on culture conditions, but was consistently significant when compared to three wild-type cell lines ($P<0.001$) (Fig. 2f for representative results).

Consistent with these findings, over 95% of KMT2D-bound chromatin in GC B cells is decorated by H3K4me1 (1,517/4,153 peaks, 36%) and/or H3K4me3 (2,578/4,153 peaks, 62%) (Fig. 2g,h). Collectively, our data indicate that, despite being present in only one of the six mammalian COMPASS-like complexes, KMT2D functions as a non-redundant methyltransferase that controls, directly or indirectly, the methylation state of a large number of regions in the mature B cell compartment.

Early loss of *Kmt2d* affects follicular B cell development

Since the stage of hematopoietic differentiation at which *KMT2D* lesions occur is unknown, we generated mice where *Kmt2d* is conditionally deleted at distinct stages of B cell differentiation by crossing mice bearing a *Kmt2d* floxed allele[21] with mice expressing the Cre-recombinase specifically in GC B cells (*Kmt2d*$^{fl/fl}$*Cγ1-Cre*)[28] or in B cell progenitors (*Kmt2d*$^{fl/fl}$*CD19-Cre*)[29]. Animals were analyzed before and 10 days after immunization with sheep red blood cells (SRBC), a T-cell dependent antigen that elicits robust GC responses and, in the case of the *Cγ1-Cre* model, is required to induce the expression of Cre.

We did not find significant differences between unimmunized *Kmt2d*$^{fl/fl}$*CD19-Cre* and littermate *Kmt2d*$^{fl/+}$ and *Kmt2d*$^{+/+}$ controls in the distribution of bone marrow (BM) B cell subpopulations, nor in splenic follicular (FO) and marginal zone (MZ) B cells (Supplementary Fig. 5). The percentage and absolute number of pro-, pre-, and immature B cells in the BM were comparable even after SRBC immunization (Supplementary Fig. 6a,b). However, immunized *Kmt2d*$^{fl/fl}$*CD19-Cre* mice displayed a significant reduction in splenic and lymph node B220^{+} cells (Supplementary Fig. 6c,d), and in the fraction of mature B cells recirculating in the BM (Supplementary Fig. 6b). Since mature B lymphocytes derive sequentially from transitional type 1 (T1) and type 2 (T2) precursors[30], we then examined the distribution of various B cell subsets in the spleen of immunized mice. While no significant differences were noted in the absolute number of transitional[31] and MZ B cells, FO B cells were markedly diminished in the absence of *Kmt2d* (Supplementary Fig. 6c,d), suggesting that its function is more important in the development and maintenance of this population. Since these defects were absent or modest in unimmunized mice, we conclude that these effects are not due to a block in differentiation, but may reflect impaired homeostasis of FO cells.

Early deletion of *Kmt2d* enhances GC formation

In the *Cγ1-Cre* cohort, *Kmt2d*$^{fl/fl}$, *Kmt2d*$^{fl/+}$ and *Kmt2d*$^{+/+}$ mice did not show significant differences in the fraction of GC B cells (B220^{+}CD95^{+}PNA^{hi}), despite efficient ablation of both Kmt2d RNA and protein (Supplementary Fig. 7a–d). Accordingly, *Kmt2d*$^{fl/fl}$*Cγ1-Cre* and *Kmt2d*$^{fl/+}$*Cγ1-Cre* mice were indistinguishable from wild-type controls in the size and number of GCs, as well as in IgG1^{+} B cells and plasma cells, and in total IgG1 serum levels

(Supplementary Fig. 7e–g). Thus, GC development occurs normally when *Kmt2d* is deleted as early as four days after the initiation of the GC reaction[28].

In contrast, SRBC-immunized *Kmt2d*$^{fl/fl}$*CD19-Cre* mice displayed up to a three-fold expansion in the percentage and absolute number of GC B cells, which was paralleled by a significant increase in mean GC size, overall GC area, and, at 6-months of age, total number of GCs in the spleen (Fig. 3), along with augmented surface IgG1^{+} cells (Supplementary Fig. 8). Analogous results were obtained when animals were challenged with the clonally restricted T-cell dependent antigen NP-KLH (Supplementary Fig. 9a–d). Interestingly, mutant mice responded efficiently to NP-KLH, as documented by the increase of NP-binding serum IgM and IgG1 antibodies 12 and 28 days after immunization; however, the magnitude of the response was ~10 fold lower than in control mice (Supplementary Fig. 9e). Accordingly, a lower proportion of GC B cells harbored the W33L substitution in the immunoglobulin variable heavy-chain region 186.2, a well-established readout of high affinity for the NP hapten (Supplementary Fig. 9f)[32]. Together with the observation of otherwise normal somatic hypermutation (Supplementary Fig. 9g,h), these data suggest a modest impairment of *Kmt2d*-deficient cells in affinity maturation, which in turn could reflect inefficient cell selection.

The observation that loss of Kmt2d early during B cell development leads to enhanced GC responses was notable when considering that B220^{+} cells, and specifically the number of FO B cells –representing the population that gives rise to GC B cells–were markedly reduced in these animals (Supplementary Fig. 6d). An intermediate phenotype could also be appreciated in the GC responses of heterozygous mice (Fig. 3), suggesting that *Kmt2d* dosage reduction has subtle pathogenic effects in this population.

Collectively, these findings suggest a propensity of *Kmt2d*-deficient FO B cells to enter the GC reaction in response to antigenic stimuli, and support a model whereby the deletion of *Kmt2d* early before GC formation may influence T cell-dependent responses by remodeling the epigenetic landscape of the precursor B cells.

Early loss of *Kmt2d* induces transcriptional changes in GC B cells

To understand the mechanisms underlying the differential effect of *Kmt2d* loss in early vs late B cell development, we contrasted the gene expression profiles (GEP) of GC B cells isolated from immunized *Kmt2d*$^{fl/fl}$, *Kmt2d*$^{fl/+}$, and *Kmt2d*$^{+/+}$ animals in both the *Cγ1-Cre* and *CD19-Cre* cohorts. We found 221 genes differentially expressed between *Kmt2d*-deficient and *Kmt2d*-proficient cells in the *CD19-Cre* background, vs 60 in *Cγ1-Cre* compound mice (Student's *t*-test, FDR < 0.15, FC ≥ 1.5), suggesting that stage/timing of deletion influences the transcriptional changes occurring in the GC population (Fig. 4a and Supplementary Table 1). However, cross-comparison of the signatures obtained in the two genetic backgrounds by an extended GSEA approach[33] revealed a significant overlap between the top differentially expressed genes and their relative ranking in the two models (Fig. 4b). Thus, the nature of the induced transcriptional changes was highly similar, independent of whether ablation of *Kmt2d* occurred before or after GC initiation. Consistent with this observation, both *Cγ1-Cre* and *CD19-Cre Kmt2d*-deficient GC B cells showed enrichment in overlapping biological programs, the most notable of which were cell cycle

regulation (e.g. Ccnd3, Ccnf and Cdk6) and apoptosis (e.g. upregulation of the anti-apoptotic genes Bcl2l1 and Bcl2, and downregulation of the pro-apoptotic genes Bmf and Dedd2) (Fig. 4c,d, Supplementary Fig. 10a and Supplementary Table 2).

Intersection of GEP data with KMT2D ChIP-seq data from human GC centroblasts revealed that only a minority of the down-regulated genes (n = 142/1,028, 14%) were bound by KMT2D, even when using less stringent cutoffs (Supplementary Notes, Supplementary Fig. 10b–c and Supplementary Table 3). Thus, within the inherent limits of the cross-species comparison, the phenotypic changes imposed by *Kmt2d* deficiency are mostly indirect. Nonetheless, the set of predicted *bona fide* direct targets of KMT2D activity (i.e. genes occupied by KMT2D at active promoters/enhancers and down-regulated following *Kmt2d* deletion; see Methods) were significantly enriched in biological programs with key functions in B cell physiology and pathology, including B cell receptor activation/signaling, chemokine/cytokine signaling, and lymphocyte migration[34-36] (Fig. 4e,f and Supplementary Table 3). Taken together, our data indicate that the transcriptional signature imposed by *Kmt2d* loss is enriched in biological programs that favor proliferation and survival, and that loss of *Kmt2d* early before GC induction facilitates the establishment and influences the magnitude of these transcriptional changes.

KMT2D loss confers B cell proliferative advantage

Following on the above findings, we investigated whether loss of *Kmt2d* affects proliferation and cell cycle regulation. Because GC B cells cannot survive *ex vivo* due to the rapid inception of a pro-apoptotic program[37], we first measured the proliferative capacity of $B220^+$ cells isolated from the spleen of *CD19-Cre* and *Cγ1-Cre* mice and stimulated with anti-CD40+IL4, which mimics the signals delivered *in vivo* to induce GC responses. By day 3 post-stimulation, *CD19-Cre Kmt2d*-deficient cells (but not *Cγ1-Cre Kmt2d*-deficient cells; Supplementary Fig. 11) had undergone a significantly higher number of cell divisions and generated greater numbers of viable cells, compared to their heterozygous and wild-type littermates (Fig. 5a–c). Accordingly, the proportion of cells residing in the S phase of the cell cycle was significantly augmented in *Kmt2d*-deficient mice (on average, 45% vs 35% in *Kmt2d*-wild-type controls; $P < 0.001$)(Fig. 5d,e). The increased response to anti-CD40 and IL4 is unlikely due to an altered distribution of B cell subsets in the splenic $B220^+$ population of *Kmt2d*-deficient mice, since the proportion of FO and MZ B cells in unimmunized conditions was similar to wild-type controls (Supplementary Fig. 5). Analysis of DNA content in freshly purified GC B cells confirmed that a significantly larger number of cells were actively cycling in the *Kmt2d*-null mice (Fig. 5f,g). Conversely, we did not detect measurable changes in apoptosis under these conditions (not shown). Taken together, these findings indicate that loss of *Kmt2d* provides substantial proliferative advantage, and suggest one possible mechanism by which tumor-associated inactivating mutations might contribute to tumor development.

Loss of *Kmt2d* cooperates with Bcl2 deregulation in lymphomagenesis

To examine whether *KMT2D* is a *bona fide* tumor suppressor gene in B cells, we investigated the impact of *Kmt2d* deficiency on lymphoma development *in vivo*, either as a single lesion or in cooperation with BCL2 deregulation, a genetically-driven event that co-

occurs with *KMT2D* mutations in nearly all affected FL/tFL cases and in ~36% of *KMT2D*-mutated DLBCL cases[3,5,8]. Toward this end, we crossed $Kmt2d^{fl/fl}$*Cγ1-Cre* animals with *VavP-Bcl2* transgenic mice, which develop FL preceded by GC hyperplasia in 37–50% of cases[38], and monitored them for tumor incidence and survival over a period of 18 months.

We did not observe significant differences between $Kmt2d^{fl/fl}$, $Kmt2d^{fl/+}$ and $Kmt2d^{+/+}$ *Cγ1-Cre* littermates, indicating that loss of *Kmt2d* in the GC is not sufficient to drive malignant transformation (Fig. 6a and data not shown). *Kmt2d* deletion also had no impact on the event-free survival of compound *VavP-Bcl2* animals, which all died by 13 months of age, likely due to the severe glomerulonephritis and autoimmune diseases developing in this strain at nearly 100% penetrance, as reported[38,39].

However, histological analysis of the lymphoid organs revealed that, compared to *VavP-Bcl2* littermate controls, a significantly higher proportion of $Kmt2d^{fl/fl}$*Cγ1-Cre/VavP-Bcl2* mice had evolved to develop clonal B cell lymphoproliferative disorders recapitulating the spectrum of phenotypes associated with the progression of FL to DLBCL, i.e. early stages of FL (eFL), overt FL of various grades, and DLBCL (22/28, 78.6% in $Kmt2d^{fl/fl}$ *vs* 15/24, 62.5% in $Kmt2d^{fl/+}$ and 12/27, 44.4% in $Kmt2d^{+/+}$; $P < 0.05$)(Fig. 6a,b). These tumors were of GC B-cell origin, as they stained positive for the pan-B cell markers B220 and Pax5, as well as for Bcl6 (Fig. 6b and not shown). Furthermore, Southern blot analysis of the immunoglobulin locus documented a clonal origin in 92% DLBCL and 55% FL tested, as well as in 23.5% eFL (Fig. 6c, and 6d for representative cases). The inability to detect clonal rearrangements in many of the latter is likely explained by the limited representation of tumor cells in the biopsy (see Fig. 6b, eFL panel as an example); as such, the larger number of clonal lymphoproliferations observed in the *Kmt2d* knock-out mice might reflect a more advanced degree of tumor infiltration/aggressiveness. Most importantly, 7/7 clonal tumors sequenced harbor somatically hypermutated immunoglobulin genes, providing unequivocal evidence of their derivation from a GC-experienced cell (Supplementary Table 4). Together these data suggest that loss of *Kmt2d* cooperates with *BCL2* deregulation to facilitate the development of GC B cell malignancies recapitulating the features of the human FL and DLBCL.

Discussion

The first result of this study is the identification of two additional mechanisms by which KMT2D is inactivated in DLBCL and FL, i.e. missense mutations affecting its C-terminal cluster of SET domains and, in a small subset of cases, loss of protein expression, implicating epigenetic silencing. Collectively, various modes of inactivation are responsible for defective KMT2D function in nearly 50% of DLBCL and possibly all FL cases.

We document that, in samples with monoallelic *KMT2D* mutations, constituting up to two thirds of affected DLBCL cases and ~25% of FL/tFL cases, the residual wild-type allele is always expressed, and that conditional heterozygous deletion of *Kmt2d* leads to a GC phenotype intermediate between that of wild-type and knock-out littermates, suggesting that dose reduction of KMT2D may be pathogenic also in the context of B cells, analogous to the haploinsufficient role played in Kabuki syndrome[24]. The reason why a subset of tumors

retain one copy of *KMT2D* is not known, but may be related to the presence of unique combinations of genetic alterations in molecules with partly overlapping functions within the complex genomic landscape of DLBCL and tFL. One such candidate is the KMT2D paralogue KMT2C, that was shown to have partial functional redundancy[21] and is also mutated in these diseases, though at lower frequencies[6-8,40]. Although *KMT2C* mutations have not been functionally characterized and the number of affected cases is too low to assess statistical significance, these lesions seem more common in individuals with monoallelic vs biallelic *KMT2D* lesions (5/42 vs 1/38) in our data set.

Both *in vitro* and *in vivo*, loss of KMT2D activity resulted in significant reduction of bulk H3K4 methylation, a highly conserved histone modification that regulates transcriptionally active promoters and enhancers[12-14]. While recent studies in mammalian cells have emphasized the role of KMT2D as the major enzyme responsible for H3K4me1 at enhancer regions[20-22], *KMT2D*-deficient alleles in this study were also consistently associated with diminished tri-methylation, suggesting a non-redundant role in mature B cells. Accordingly, KMT2D-bound regions in GC B cells display widespread overlap with both H3K4me1 at putative GC enhancers[41,42] (including predicted super-enhancers)[43,44] and H3K4me3 at gene promoters (Supplementary Results and Supplementary Fig. 10b–c), consistent with the requirement of *Drosophila Trithorax-related* (*Trr*) in enhancer-promoter communication for faithful and tissue-specific gene expression[45-46].

Loss of *Kmt2d* perturbed the expression of genes that sustain proliferation and survival, and the KMT2D protein directly binds and associates with an active chromatin conformation in negative modulators of the BCR and lymphocyte migration pathways, which in turn could affect B cell responses to antigen. Moreover, KMT2D mapped to 349/593 putative super-enhancers identified in GC B cells, many of which were assigned to master regulators of the GC phenotype, based on proximity to the nearest transcriptionally active TSS (Supplementary Results and data not shown).

One key observation of this study is that stage-specific deletion of *Kmt2d* (before vs after initiation of the GC reaction) influences the magnitude of the transcriptional and phenotypic changes imposed on this population. We hypothesize that the more robust phenotype observed upon early KMT2D inactivation be explained by two not mutually exclusive scenarios. First, in *CD19-Cre* mice the longer window of time, and hence the higher number of divisions completed by the cell in the absence of Kmt2d activity before entering the GC reaction might enable the implementation of chromatin changes, as cell division facilitates the replacement of Kmt2d-dependent H3K4me3 and H3K4me1 modified histones by *de novo* synthesized non-modified histone H3[47]. In support of this hypothesis, the GEP of *Kmt2d*-null GC B cells was qualitatively similar irrespective of whether the deletion occurred before or after GC formation; consistently, the reduction in global histone methylation was less evident in GC B cells purified from *Cγ1-Cre* compared to *CD19-Cre* mice (Supplementary Fig. 7b vs Fig. 2e). On the other hand, the stage of differentiation/cellular context in which KMT2D is lost may have a specific impact on the target cell by remodeling the epigenetic landscape of the precursor cell, and potentially skewing cell fate decisions. Regardless of the mechanism, two observations lend support to the role of early stage *Kmt2d* deletion: i) immunized *CD19-Cre* mice exhibit an over 2-fold increase in GC B

cells, despite the considerably reduced number of FO B cells, suggesting that *Kmt2d*-null cells have a higher propensity to enter the GC reaction upon antigen stimulation; ii) in human tumors, *KMT2D* alterations represent early events acquired by a common mutated precursor before divergent evolution and clonal expansion into FL and tFL[7,8,11]. By enhancing GC responses, early *KMT2D* loss will create a larger pool of cells that could serve as targets for additional lesions in this genetically unstable environment[48-51]. Indeed, GC enlargement has been previously shown to favor lymphomagenesis in several mouse tumor models (e.g., Bcl6 knock-in and Blimp1 knock-out)[52,53].

In summary, we provide evidence from human genetics, functional studies and mouse models that *KMT2D* is a *bona fide* tumor suppressor gene in FL and DLBCL. The mechanism of activity of KMT2D is different from that of classical tumor suppressor genes, as it may allow the selection of cell phenotypes (i.e. the GC expansion) and transcriptional programs compatible with transformation driven by other oncogenic lesions, rather than having a direct oncogenic action. Finally, the fact that BCL2 deregulation and KMT2D loss are early events preceding FL or DLBCL transformation suggests that targeting these two lesions may represent a rational strategy to eradicate the putative pre-malignant precursor. The animal model reported here may be useful for the pre-clinical testing of therapeutic approaches combining BCL2- and histone methyltransferase/demethylase-inhibitors[54-56].

Online Methods

Cell lines and isolation of human B cell subsets

DLBCL cell lines were obtained as detailed in the Table below and were cultured in Iscove's Modified Dulbecco's Medium (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich) and antibiotics, except for OCI-Ly10 and OCI-Ly4, which were cultured in RPMI 1640 medium (Life Technologies) containing 20% human plasma and 55 μm β-mercaptoethanol (Life Technologies). Cell lines were verified for identity by FISH analysis or by targeted resequencing of previously reported mutated genes (when applicable), unless purchased from commercial sources. HT-DHL8 is a subclone of the published HT DLBCL cell line, which carries one additional *KMT2D* truncating mutation on the same allele as the reported missense mutation[5]. HEK 293T cells were cultured in Dulbecco's Modified Eagle Medium (Life Technologies) with 10% FBS and antibiotics. All cell lines were verified to be free from Mycoplasma contamination. Human GC centroblasts, naïve B cells and memory B cells were isolated from reactive tonsils as described[57].

Cell line	Source	Cell line	Source
OCI-Ly1	Ontario Cancer Institute*	SUDHL2	Stanford University^,*
OCI-Ly3	Ontario Cancer Institute*	SUDHL4	Stanford University^,*
OCI-Ly4	Ontario Cancer Institute*	SUDHL5	Stanford University^
OCI-Ly7	Ontario Cancer Institute*	SUDHL6	Stanford University^,*
OCI-Ly8	Ontario Cancer Institute*	SUDHL7	Stanford University^
OCI-Ly10	Ontario Cancer Institute*	SUDHL10	Stanford University^

Cell line	Source	Cell line	Source
OCI-Ly18	Ontario Cancer Institute *	FARAGE	ATCC
HBL1	Tohoku et al., J Exp Med 1988§*	HT-DHL8	Dr. M. Shipp (DFCI)
U2932	DSMZ	WSU-NHL	DSMZ
RIVA (RI-1)	Th'ng et al., Int J Cancer 1987§	BJAB	Dr. T. Gilmore (BU) *
RC-K8	DSMZ *	DB	ATCC *
TOLEDO	ATCC	Karpas 422	DSMZ
PFEIFFER	ATCC	VAL	DSMZ

^Epstein et al., Cancer 1978; kindly provided by Dr. M. Shipp (Dana Farber Cancer Institute);

§kind gift of Dr. L. Staudt. BU, Boston University;

*verified for identity.

Primary tumor samples

Primary biopsies from 97 newly diagnosed, previously untreated DLBCL patients with known *KMT2D* genetic status were obtained as part of independent studies and have been described[5,8]. Samples were studied according to a protocol approved by the Institutional Review Board of Columbia University (Exempt Human Subject Research of anonymized/de-identified existing pathological specimens, under regulatory guideline 45 CFR 46.101(b)(4)).

KMT2D genetic data

For a comprehensive assessment of the *KMT2D* mutation pattern in DLBCL and tFL (Supplementary Fig. 3), data generated in our previous studies[5,8] were integrated with published information from exome sequencing and/or transcriptome sequencing studies[3,6,7], resulting in a total of 250 DLBCL primary cases, 31 DLBCL cell lines, and 71 tFL.

Immunohistochemistry (IHC) and immunofluorescence analysis of human tissues

DLBCL tissue microarrays (TMA) were constructed according to standard procedures[58] and analyzed for KMT2D expression by IHC, using a rabbit polyclonal antibody directed against the C-terminus of the KMT2D protein (HPA035977, Sigma-Aldrich) (1:200 dilution). Cases were independently scored by two pathologists (D.D-S. and S.H.), and were considered positive if ≥30% tumor cells showed staining in the nucleus. Double-immunofluorescence analysis of KMT2D and BCL6 was performed on formalin-fixed paraffin embedded (FFPE) material from reactive human tonsils using the above-mentioned anti-KMT2D antibody (1:200 dilution) and a mouse monoclonal anti-BCL6 antibody (1:300 dilution)(clone GI191E/A8, Cell Marque). Detection of KMT2D was obtained using the EnVision System-HRP-Rabbit antibody (Dako) followed by Tyramide Signal Amplification system (PerkinElmer); the secondary antibody for BCL6 was Cy-3 AffiniPure Donkey Anti-Mouse IgG (H+L) (1:400 dilution) (Jackson ImmunoResearch Laboratories, cat# 715-167-003), as reported[59].

Generation of pCMV-HA-KMT2D wild-type and mutant expression constructs

The KMT2D full-length ORF (NCBI Reference Sequence: NM_003482.3) was assembled into the pCMV-HA vector (Clontech) by sequential subcloning of six contiguous segments, obtained in separate RT-PCR amplification reactions using the PfuTurbo Cx DNA Polymerase (Agilent) and, as template, cDNA from human GC B cells, followed by cloning in the Zero Blunt TOPO PCR vector (Life Technologies). The DLBCL/FL-derived missense mutants alleles and the R4712* truncation mutant were generated from the wild-type pCMV-HA-KMT2D construct using the QuikChange II Site-Directed Mutagenesis Kit (Agilent). All plasmids were verified for integrity by diagnostic restriction enzyme digestions using the SalI, NotI, BglII, SacII, XbaI and MfeI sites, and subjected to Sanger sequencing of the full-length KMT2D CDS.

Transient transfection assays

HEK 293T cells were transiently transfected with equimolar amounts (1 μg/ml) of wild-type and mutant pCMV-HA-KMT2D expression vectors using the polyethylenimine method, following published protocols[60]. Cells were harvested 48 hours after transfection and used for protein extraction or for the isolation of semipurified HA-KMT2D proteins, as described below.

Protein extraction

Whole cell extracts were obtained from cell lines in log phase of growth or from purified mouse B cells using NP-40 lysis buffer (150 mM NaCl, 1.0% NP-40, 50 mM Tris-HCl pH 8.0) supplemented with proteinase inhibitor cocktail (Sigma-Aldrich) and 0.5 mM phenylmethanesulfonyl fluoride (PMSF) (Roche), according to previously described protocols[61]. For the analysis of histone modifications, the chromatin pellet was subsequently resuspended in RIPA-coIP buffer (Tris-HCl pH 8.0 50mM, NaCl 200mM, Glycerol 5%, $MgCl_2$ 2.5mM, Sodium deoxycholate 0.05%, SDS 0.05%) and mobilized by sonication in a Bioruptor Standard sonication device (Diagenode). Extracts were cleared by centrifugation at 12,000 rpm for 10 minutes.

Immunoblot analysis

Protein extracts were resolved on NuPAGE Tris-acetate 3–8% gels (for KMT2D) or Tris-Glycine 4–20% gels (for histone H3) (Life Technologies) and transferred to nitrocellulose membranes (GE Healthcare) according to the manufacturer's instructions. Antibodies used were: rabbit polyclonal anti-KMT2D (kindly provided by E. McIntush, Bethyl Laboratories, and designated as KMT2D-1185 in Supplementary Fig. 2 and Supplementary Table 5), mouse monoclonal anti-α-tubulin (clone DM1A, Sigma-Aldrich), rabbit polyclonal anti-H3K4me1 (Abcam, cat#ab8895), anti-H3K4me2 (Active Motif, cat#39142), anti-H3K4me3 (Abcam, cat#ab8580), and rabbit monoclonal anti-Histone H3 (clone D1H2, Cell Signaling Technology). The specificity of the antibodies against H3K4 methylation marks has been documented in previous studies (ENCODE Project: http://genome.ucsc.edu/ENCODE/antibodies.html) (see also http://www.activemotif.com). For immunoblot analysis of Kmt2d expression in mouse tissues, we used a rabbit polyclonal antibody from Diagenode (C15310100). The detailed list of anti-KMT2D antibodies is provided in Supplementary

Table 5. Loading of chromatin extracts (typically, 0.5–1 μg) was adjusted to ensure comparable amounts of total histone H3, which can vary in different cell lines. Quantitation of signal intensity was obtained in the ImageJ software (http://imagej.nih.gov/ij/), by subtracting the background signal measured above each band from the signal measured in each band; areas of the same size (set on the image of the wild-type protein) were used for all measurements. Values are expressed as fold differences relative to the wild-type protein sample, set at 1, after normalization for the loading control.

In vitro histone methyltransferase (KMT) assay

Partially purified HA-KMT2D wild-type and mutant derivative proteins were obtained from transfected HEK 293T cells by lysis in co-IP buffer (50 mM Tris, pH 7.5, 250 mM NaCl, 1% TritonX-100, 1 mM EDTA), followed by overnight incubation with EZview™ Red Anti-HA Affinity Gel (Sigma-Aldrich) at 4 °C and final elution in BC100 buffer (20 mM Tris pH 7.5, 10% Glycerol, 0.2 mM EDTA, 1% TritonX-100,100 mM NaCl) containing HA peptide (Sigma-Aldrich). Products were concentrated in the Vivaspin Sample Concentrators (GE healthcare, 10 kDa cutoff) at 12,000 rpm for 10 minutes. KMT2D protein amounts were quantified by Coomassie staining and immunoblot analysis using rat monoclonal anti-HA antibodies (clone 3F10, Roche).

KMT activity against an artificial H3 peptide was assayed with the EpiQuik Histone Methyltransferase Activity/Inhibition Assay Kit (H3K4) (Epigentek), following the manufacturer's protocol. Relative activity was calculated as the fold change in OD_{450nm} over the mean reading of wild-type samples. Experiments were performed in triplicates and repeated independently at least twice.

Enzymatic activity against native nucleosomes was measured following a published method[27]. Briefly, equal amounts of wild-type and mutant HA-KMT2D proteins were incubated at 37 °C for 2 hours with Hela nucleosomes (Reaction Biology) in KMT buffer (50 mM Tris pH 8.5, 100 mM KCl, 5 mM MgCl2, 10% Glycerol, 4 mM DTT) supplemented with S-Adenosyl methionine (New England Biolabs). Reactions were stopped by adding equal volumes of 2× Laemmli buffer and heated at 100 °C for 5 minutes before loading into Tris-Glycine 4–20% gradient gels. All assays were performed twice independently.

RNA extraction, cDNA synthesis and quantitative real-time PCR

Total RNA was extracted from DLBCL cell lines and primary mouse lymphocytes by TRIzol (Life Technologies) and treated with DNAse prior to cDNA synthesis, which was performed using the SuperScript® First-Strand Synthesis System (Life Technologies), according to the manufacturer's instructions. The ABsolute QPCR SYBR green mix (Thermo Scientific) was then utilized to amplify specific cDNA fragments with the oligonucleotides listed below, in the 7300 Real Time PCR system (Applied Biosystems). Data were analyzed by the change-in-threshold ($2^{-\Delta\Delta CT}$) method[62], using *Actin* as housekeeping reference gene.

Oligonucleotide Name	Oligonucleotide Sequence	Organism
hKMT2D-Forward	5′-GTGCAGCAGAAGATGGTGAA-3′	*Homo sapiens*
hKMT2D-Reverse	5′-GCACAATGCTGTCAGGAGAA-3′	*Homo sapiens*
hACTB-Forward	5′-CATGTACGTTGCTATCCAGGC-3′	*Homo sapiens*
hACTB-Reverse	5′-CTCCTTAATGTCACGCACGAT-3′	*Homo sapiens*
mKmt2d-Forward	5′-GCTATCACCCGTACTGTGTCAACA-3′	*Mus musculus*
mKmt2d-Reverse	5′-CACACACGATACACTCCACACAA-3′	*Mus musculus*
mActb-Forward	5′-ATGGAGGGGAATACAGCCC-3′	*Mus musculus*
mActb-Reverse	5′-TTCTTTGCAGCTCCTTCGTT-3′	*Mus musculus*
mTnf-Forward	5′-GAATGGGCTTTCAATCTGGA-3′	*Mus musculus*
mTnf-Reverse	5′-GCACCTCAGGGAAGAGTCTG-3′	*Mus musculus*
mTnf-Forward-1	5′-GGCCTTCCTACCTTCAGACC-3′	*Mus musculus*
mTnf-Reverse-1	5′-AGCAAAAGAGGAGGCAACAA-3′	*Mus musculus*
mBcl2-Forward	5′-CATCTGCAGAACCTCCCTGT-3′	*Mus musculus*
mBcl2-Reverse	5′-GCACTACCTGCGTTCTCCTC-3′	*Mus musculus*
mBcl2l1-Forward	5′-CCCCAGAAGAAACTGAAGCA-3′	*Mus musculus*
mBcl2l1-Reverse	5′-CCTGGGGTTATGTGAAGCTG-3′	*Mus musculus*
mCdk6-Forward	5′-GGCGTACCCACAGAAACCATA-3′	*Mus musculus*
mCdk6-Reverse	5′-AGGTAAGGGCCATCTGAAAACT-3′	*Mus musculus*
mCcnd3-Forward	5′-TCCGTACTCCAGCTGCTCTT-3′	*Mus musculus*
mCcnd3-Reverse	5′-AGCTAAGCAGCGAAGCAAAG-3′	*Mus musculus*
mCcnf-Forward	5′-AAACCCATCCCTGTCTACCC-3′	*Mus musculus*
mCcnf-Reverse	5′-CCAGCAAGGGTTTGAACTGT-3′	*Mus musculus*
mDedd2-Forward	5′-ACTCTGGGCTGATGCTGTCT-3′	*Mus musculus*
mDedd2-Reverse	5′-GCCCAGGTTGGTTAGATGAA-3′	*Mus musculus*
mIfngr1-Forward	5′-GCCTGCTGGTGGTAAAGAAA-3′	*Mus musculus*
mIfngr1-Reverse	5′-AGGAGCCAGAAGCGACAATA-3′	*Mus musculus*
mLta-Forward	5′-CTCCATCCTGACCGTTGTTT-3′	*Mus musculus*
mLta-Reverse	5′-GTCCGACCTAGACCCACAAA-3′	*Mus musculus*
mMyb-Forward	5′-GCCTGAATGTTCATCCGTTT-3′	*Mus musculus*
mMyb-Reverse	5′-AAGGCAGAAACTGGCTGTTG-3′	*Mus musculus*

Mice

The conditional *Kmt2d* knock-out mouse model (mixed C57BL/6 × Sv129 background) has been published[21]. Deletion of *Kmt2d* in GC B cells or in precursor B cells was obtained by breeding $Kmt2d^{fl/+}$ mice with *CD19-Cre*[29] or *Cγ1-Cre*[28] mice (both in pure C57BL/6 background), followed by offspring intercrossing. Compound mice carrying the conditional *Kmt2d* allele (either *CD19-Cre* or *Cγ1-Cre*) and the *VavP-Bcl2* transgene were obtained by breeding with *VavP-Bcl2* mice (C57BL/6 background)[38]. A sample size of ≥ 23 animals/genotype was calculated to ensure over 80% power to detect an increase in tumor formation in the *Kmt2d* compound lines at a $P < 0.05$ significance. Both females and males were included in the experiment, and no randomization was used. Genotyping was performed by

PCR analysis using oligonucleotides 5′-ATTGCATCAGGCAAATCAGC-3′ (forward) and 5′-GCAGAAGCCTGCTATGTCCA-3′ (reverse). Animals were monitored for tumor incidence and survival twice/week over a period of 18 months, and sacrificed for analysis when visibly ill or at the end of the study, according to protocols approved by the National Cancer Institute and Columbia University Institutional Animal Care and Use Committee.

Mouse Immunizations

For the analysis of T cell-dependent responses, age-matched 8–12 week-old mice (males and females) were immunized by intraperitoneal injection of 500 million sheep red blood cells (SRBC) (Cocalico Biologicals) in 200 μl PBS, and analyzed after 10 days. Alternatively, mice were immunized with 100 μg 4-Hydroxy-3-nitrophenylacetyl hapten conjugated to Keyhole Limpet Hemocyanin (NP-KLH) (Biosearch Technologies) along with Complete Freund's Adjuvant (Sigma-Aldrich), and analyzed after 12 and 28 days. To achieve a higher yield of GC B cells for gene expression profile analysis, mice were immunized with two sequential SRBC injections (day 0: 1×10^8 cells; day 5: 1×10^9 cells) and sacrificed at day 12. Mice in the tumor-watch cohort were subjected to chronic immunization by repeated antigenic stimulation with 5×10^8 SRBC, delivered every two months until tumor development, death or end of the study.

Flow cytometric analysis

Multi-color flow cytometric analysis of B and T cell lymphoid compartments was performed at 3 and 6 months of age as previously reported[52]. The B cell compartment was also examined in all animals from the tumor-watch cohort, when sacrificed. Briefly, single cell suspensions prepared from various lymphoid organs and from the peritoneal cavity were stained using different combinations of fluorescent-labeled antibodies, listed in Supplementary Table 5. Samples were acquired on a FACSCanto II (BD Biosciences), and analyzed using the FlowJo software (Tree Star).

Ex vivo stimulation of splenic lymphocytes with anti-CD40 + IL4

Splenic mouse B-cells were isolated from age- and sex-matched 8–12 weeks old mice by negative selection using the mouse B-cell isolation kit from Miltenyi Biotec, according to the manufacturer's instructions. The purity of the B cell population was 96–98% in all experiments, as documented by flow cytometric analysis of B220 expression. Cells were cultured in RPMI 1640 medium supplemented with 15% FBS, 55 μm β-mercaptoethanol, 1 μg/ml Hamster monoclonal anti-CD40 (clone HM40-3, BD Pharmingen) and 50 ng/ml recombinant mouse IL4 (R&D systems).

Cell viability and proliferation assays

The number of viable cells in the *ex vivo* cultures was determined by measuring the presence of ATP, which signals the presence of metabolically active cells, using the CellTiter-Glo® Luminescent Cell Viability Assay (Promega) as per manufacturer's instructions. Cell proliferation was analyzed with the CellTrace™ Violet Cell Proliferation Kit (Life Technologies), which monitors distinct generations of proliferating cells by a fluorescent dye dilution. Data were acquired at day 2, 3 and 4 of stimulation on a

FACSCanto II (BD Biosciences) flow cytometer with 405 nm excitation and an emission filter in the 450 nm range. All experiments were performed in at least two independent groups of animals, in biological triplicate.

Cell cycle analysis

To assess the cell cycle profile of *Kmt2d*$^{fl/fl}$ vs *Kmt2d*$^{fl/+}$ and *Kmt2d*$^{+/+}$ B cells in the *CD19-Cre* and *Cγ1-Cre* cohorts, B220^{+} splenocytes were labeled with bromodeoxyuridine (BrdU) for 2 hours, and analyzed for BrdU incorporation and DNA content using the APC BrdU flow kit (BD Biosciences) on a FACSCanto II (BD Biosciences) flow cytometer. Experiments were performed in triplicate at day 3, 4, and 5 after stimulation with anti-CD40 + IL4.

GC B cell DNA content analysis

For DNA content analysis, splenic B cell suspensions were stained for surface markers, fixed with Fixation/Permeabilization Solution Kit (BD Biosciences) for 30 minutes on ice, and then incubated with the DNA-binding dye 4′,6-Diamidino-2-phenylindole, dihydrochloride (DAPI, Sigma) at the final concentration of 300 ng/mL for 5 minutes, as described[59]. Samples were acquired at a maximum of 1,500 events per second on a LSR II Flow Cytometer System (BD Biosciences). Cell cycle analysis was performed in the Flow-Jo software (Tree star).

Calculating absolute numbers of cells within B cell subsets

For the analysis of splenic B cell subsets, we first calculated the total number of B220^{+} cells/spleen by multiplying the fraction of B220^{+} cells in the FACS profiles by the total number of erythrocyte-depleted cells in the spleen. Absolute numbers of transitional (T1, T2 and T3), FO, MZ and GC B cells were then calculated by multiplying the fraction of each FACS B cell subset analyzed by the total number of B220^{+} cells. The total number of plasma cells in BM and spleen was obtained by multiplying the fraction of B220^{-} CD138^{hi} cells in the FACS profiles by the total number of erythrocyte-depleted cells in these organs.

Histological and immunohistochemical analysis of mouse tissues

Histological analysis of mouse organs was performed on 4 μm thick FFPE tissue sections, stained with Hematoxylin & Eosin (H&E) (Thermo Scientific)[53]. The following primary antibodies and dilutions were used for immunohistochemical analysis: rabbit polyclonal anti-Bcl6 (1:300) (N3, Santa Cruz Biotechnology) and anti-Pax5 (1:400) (Neomarker); rabbit monoclonal anti-CD3 (1:800) (clone SP7, Neomarker) and anti-Ki67 (1:200) (clone SP6, Thermo Scientific); rat monoclonal anti-B220 (1:400) (clone RA3-6B2, BD Biosciences) and anti-CD138 (1:200) (clone 281-2, BD Biosciences); mouse monoclonal anti-BCL2 (1:200) (clone Bcl-2/100, BD Biosciences).

Diagnosis of mouse lymphoid tumors

Lymphoproliferative diseases developing in the *VavP-Bcl2/Kmt2d/Cγ1-Cre* cohort were diagnosed independently by two pathologists (D.D.-S. and S.H.) based on morphology and phenotype using the following criteria, in accordance with the classification of human

lymphoma: i) early stages of FL (eFL), defined by the presence of one or few oversized, often coalescent follicles with partial or absent mantle zone and loss of confinement in the context of a yet preserved tissue architecture; ii) overt FL of various grades, characterized by effacement of the nodal architecture by a proliferation of follicle center B cells with a follicular growth pattern occupying the medullary and/or paracortical areas; iii) DLBCL, defined by the effacement of the lymphoid organ architecture due to the expansion of large cells, with occasional infiltration beyond the capsule into surrounding soft tissues. The genotype of the animal was not disclosed to the pathologists. Immunohistochemical analysis of B220/CD3, Bcl6, Pax5 and CD138 was performed as described in the previous section.

Southern blot analysis

High molecular weight genomic DNA was obtained from frozen tissues by phenol-chloroform extraction, according to standard protocols[63]. Four μg of DNA were digested with EcoRI overnight at 37 °C. The digestion reaction was resolved on 0.8% agarose gel followed by transfer to nitrocellulose membrane, according to standard procedures. Hybridization was performed at 37 °C overnight using a Jh4 specific probe, as reported[49].

Mutation analysis of the rearranged immunoglobulin genes in mouse tumors

The rearranged immunoglobulin VH sequences were amplified from mouse tumor DNAs by PCR, using forward primers that anneal to the framework region III of the most abundantly used Igh-VJ558 (5′-CAGCCTGACATCTGAGGACTCTGC-3′), Igh-V7183 (5′-GAASAMCCTGTWC CTGCAAATGASC-3′) and Igh-VQ52 (5′-CAGGTGCAGCTGAARCAGTCA-3′) family, and reverse primers positioned in the Jh4 intron (5′-ACTATCCCTCCAGCCATAGG-3′), in separate reactions[64]. PCR conditions were 94 °C for 30 s, 63 °C for 30 s and 72 °C for 2 min, for 35 cycles. Products indicative of clonal rearrangements were gel-purified using the QIAquick PCR purification kit (QIAGEN) and sequenced directly using the same primers as in the PCR reaction, or additional internal primers annealing to the Jh2 (5′-AGATGG AGGCCAGTGAGGGAC-3′) and Jh3 (5′-GTCACTGACCCTGAAATTGTC-3′) segments (GENEWIZ, Inc.). Sequences were aligned to the NCBI databases and the IMGT database, and analyzed as previously described[52].

ELISA assay

Total IgM and IgG1 serum titers were measured by ELISA as previously described[65], using the following antibodies: anti-Mouse Ig H+L (1010-01, Southern Biotech); anti-Mouse IgM-AP (1020-04, Southern Biotech); and anti-Mouse IgG1-AP (1070-04, Southern Biotech). To quantify NP-specific antibodies, serum was collected at day 0, 12 and 28 after a single intraperitoneal injection of NP-KLH (100 μg) in complete Freund's adjuvant (Sigma). NP-specific high affinity and low affinity antibodies were captured on plates coated with 2 μg/ml NP_{25}-BSA or NP_5-BSA, respectively (SouthernBiotech), and bound antibodies were detected with alkaline phosphatase-conjugated anti-IgM or anti-IgG1 antibodies (SouthernBiotech). Pan-mouse immunoglobulins were captured on plates coated with 2.5 μg/ml goat anti-mouse Ig (H+L) (SouthernBiotech). Each assay was performed in duplicate using twelve serial dilutions.

Analysis of affinity maturation in mouse GC B cells

Genomic DNA was extracted from splenic GC B cell subpopulations sorted at day 12 after NP-KLH immunization, using the QIAamp DNA Micro Kit (Qiagen). The protocol and primers used for PCR amplification and sequencing of the rearranged immunoglobulin variable heavy-chain region 186.2 (V186.2) genes, that are predominant in GC responses to NP immunization, are described in Ref.[59]. Briefly, PCR products generated with high-fidelity DNA polymerase (AccuPrime Taq, Life Technologies) were subcloned into the pGEM®-T Easy Vector Systems (Promega) and individual clones (n = 48/mouse) were sequenced directly after colony-PCR. *Igh-V* sequences were analyzed with the HighV-QUEST tool (www.imgt.org) and compared to the most homologus mouse germline counterparts to assess the overall mutation frequency and to identify clones harboring the W33L substitution. Differences between groups were determined by a two-tail Fisher's Exact Test in the GraphPad Prism software.

Gene expression profiling

$B220^+CD95^+PNA^+$ GC B-cells were sorted from $Kmt2d^{fl/fl}$, $Kmt2d^{fl/+}$ and $Kmt2d^{+/+}$ splenocytes (*CD19-Cre* or *Cγ1-Cre*; n = 3 mice/genotype) in a BD FacsAria III cell sorter (BD Biosciences). Total RNA was extracted using the NucleoSpin RNA II kit according to the manufacturer's instructions (MACHEREY-NAGEL) and, from each sample, 20 ng were reverse transcribed and amplified using the Ovation RNA Amplification System (NuGEN), followed by labeling with the Encore Biotin Module (NuGEN). Labeled samples were hybridized to Affymetrix MG430 2.0 arrays as recommended by the manufacturer. Raw expression values were normalized using the Robust Multiarray Averaging (RMA) algorithm in GenePattern (http://www.broadinstitute.org/cancer/software/genepattern/)[66], and multiple probes corresponding to the same gene were collapsed to a single probe based on maximum t-statistic. Differentially expressed genes were determined by the Student's t-test using a FDR ≤ 0.15 (after Benjamini-Hochberg correction)[67] and absolute fold change ≥ 1.5. The GEO accession number for these data is GSE67388. For visualization of gene-expression intensity, all gene expression data were normalized by gene (z-score transformation).

Gene Set Enrichment Analysis (GSEA) and extended GSEA

Gene expression data generated from *CD19-Cre* or *Cγ1-Cre* compound mice were analyzed for enrichment in pre-defined sets of genes with GSEA-P[68] using the Canonical Pathway (CP) Molecular Signature Database gene sets collection and 1,000 gene set permutations. Enrichments were considered significant if FDR ≤ 0.15 after correction for multiple hypothesis.

The cross-comparison of the signatures obtained in *Kmt2d*-deficient vs *Kmt2d*-proficient GC B cells from the two genetic backgrounds was performed using an extended GSEA approach [33]. While GSEA uses the Kolmogorov-Smirnov statistical test to assess whether a predefined gene set is statistically enriched in genes that are the two extremes of a list ranked by differential expression between two biological states[68], extended GSEA also takes into account the direction of regulation of each gene in the gene set while calculating the enrichment. As described in[33], the algorithm first separates the gene set into two independent sets, each containing upregulated and downregulated genes, and then calculates

the enrichment such that the overlap of genes from both gene sets has to be in the opposite extremes of a list ranked by differential expression for the enrichment score to be significant. A MATLAB function implementing the extended GSEA described above is available upon request.

DAVID

To determine whether genes bound by KMT2D at active promoters/enhancers and down-regulated in *Kmt2d*-deficient GC B cells were enriched in annotated functional categories, we overlapped the list of down-regulated genes obtained by supervised analysis of mouse GC B cells ($P < 0.05$ and fold change ≥ 1.2) with genes identified by the integrated KMT2D and histone marks ChIP-Seq analysis of human GC B cells. The obtained unique list of overlapping genes was then analyzed in the DAVID software (http://david.abcc.ncifcrf.gov)[69] and interrogated for enrichment in Pathways annotated in BIOCARTA, KEGG and PANTHER; clusters with a significant *P* value (< 0.05 after Benjamini-Hochberg correction) were retained.

Chromatin Immunoprecipitation (ChIP)

GC B cells, purified from reactive tonsils as described [57], were cross-linked with 1% formaldehyde for 10 min at room temperature, quenched by the addition of glycine to a final concentration of 0.125 M, and frozen at –80 °C. ChIP was performed on two independent pools of GC cells, each comprising 3–5 donors. Cell lysis and nuclei isolation was performed using the TruChIP Chromatin Shearing Kit High Cell SDS (Covaris). Nuclei were sonicated using the S220 Ultrasonicator (Covaris) in order to obtain chromatin fragments of 200–500 bp. Sheared chromatin was incubated overnight with 4 μg of anti-KMT2D (HPA035977, Sigma-Aldrich), anti-H3K27Ac (Active Motif, cat#39133) or anti-H3K4me3 (Abcam, cat#ab8580), or with 2 μg of anti-H3K4me1 (Abcam, cat#ab8895) (ENCODE Project: http://genome.ucsc.edu/ENCODE/antibodies.html). Protein A magnetic beads were then added for 4 hr incubation followed by sequential washes at increasing stringency and reverse cross-linking. Upon RNAse and proteinase K treatment, ChIP DNA was purified using the MiniElute Reaction Clean Up Kit (Qiagen) and quantified by Quant-iT PicoGreen dsDNA Reagent (Life Technologies).

ChIP-seq library preparation and Illumina sequencing

ChIP-seq libraries were constructed starting from 4 ng of ChIP or input DNA as reported in Ref. [70]. Libraries were quantified using the KAPA SYBR FAST Universal qPCR Kit (KAPA Biosystems), normalized to 10 nM, pooled and sequenced in an Illumina HiSeq 2000 instrument as single-end 100 bp reads, obtaining on average 25×10^6 reads/sample.

ChIP-seq analysis

Sequencing data were acquired through the default Illumina pipeline using Casava V1.8. Reads were aligned to the human genome hg19 assembly (https://genome.ucsc.edu/) using the Bowtie2 aligner v2.1.0[71], allowing up to two mismatches to cope with human variation. Duplicate reads (i.e., reads of identical length mapping to exactly the same genomic locations) were removed with SAM tools v0.1.19 using the rmdup option[72]. ChIP-

sequencing produced a total of ≈24 million (KMT2D), ≈20 million (H3K4me3), ≈24 million (H3K4me1) and ≈24 million (H3K27C) of uniquely mapped reads, respectively. Reads were normalized to total reads aligned (counts per million). Peak detection was done using ChIPseeqer v2.0[73] enforcing a minimum fold change of 2 between ChIP and input reads, a minimum peak width of 100 bp, and a minimum distance of 100 bp between peaks. The threshold for statistical significance of peaks was set at 10^{-5} for KMT2D and 10^{-15} for H3K4me1, H3K4me3 and H3K27Ac, and only peaks identified as significant in both GC B cell pools were considered in the study. Peaks were considered overlapping if they shared at least 1 bp; H3K4me1 (or H3K27Ac) peaks were stitched together into regions if located within ±2 kb (or ±12.5 kb) of distance from each other, unless mapping around the transcription start site (TSS) (±2kb). H3K27me3 peaks were called using RSEG with 100 bp bin size; only peaks above 5 kb in size were considered. The significance of the overlap between KMT2D occupied regions and H3K4 methylated regions (or H3K27 acetylated regions) was assessed by binomial test considering the total size (in bp) of the significant peaks identified in the two ChIP-seq experiments (e.g., KMT2D vs H3K4me1) and the size of the overlapping region between the two experiments, relative to total genome size of the UCSC hg19 assembly.

Definition of active promoters and enhancers

KMT2D-bound regions (peaks) were classified as active promoters if located in close proximity to the TSS of an annotated gene (–2/+1 kb) and enriched in H3K4me3, but not H3K27me3. Intragenic (exonic or intronic) and intergenic regions were considered as active enhancers if located distal to TSS (greater than ± 2 kb) and enriched in H3K4me1 and H3K27Ac, but not in H3K4me3[41,42,74]. For the identification of super-enhancers in GC B cells, we applied the ROSE algorithm (https://bitbucket.org/young_computation/rose)[43,75] to our H3K27Ac ChIP-Seq dataset. Occupancy of KMT2D at super-enhancers was then determined based on the overlap between KMT2D peaks and genomic regions identified by ROSE.

Assignment of active enhancers to genes

KMT2D-bound putative active enhancers, identified as described above, were assigned to the nearest expressed and transcriptionally active gene (distance from enhancer center to TSS) as the most likely candidate target gene[44].

Statistical analyses

P values were calculated by one-way ANOVA with Tukey's multiple comparisons or by the Student's *t*-test (two-tailed, equal variance) in the GraphPad Prism 5 software. The Fisher's exact test (two-tailed) was used to assess whether differences in the incidence of lymphoproliferative disorders and in the fraction of W33L+ V186.2 rearrangements were significant in *Kmt2d*-null mice compared to heterozygous and control wild-type littermates. The specific test adopted for parametric and non–parametric analyses is described in each figure legend. Kaplan-Meier event-free survival curves were generated using the GraphPad Prism 5 software (GraphPad Sofware), and statistical significance was calculated using the log-rank (Mantel-Cox) test. No randomization was used in the study; for the

histopathological analysis of human and mouse tissue biopsies, investigators were blinded to sample identity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank U. Klein and S. Zha for discussions, S. Nataraj, L. Belver, C. Scuoppo, and N. De Silva for help and suggestions on various experimental procedures, E. McIntush (Bethyl Laboratories, Inc.) for help in generating several anti-KMT2D antibodies, the Flow Cytometry Shared Resource of the Herbert Irving Comprehensive Cancer Center at Columbia University for assistance with cell sorting procedures, the Molecular Pathology lab for the preparation of mouse FFPE tissue samples, the Genomic Resource for sequencing the ChIP-DNA libraries. We also thank S. Cory (Walter and Eliza Hall Institute of Medical Research, Parkville, Australia) for the VavP-Bcl2 mice and K. Rajewsky (Max Delbrück Center for Molecular Medicine, Berlin, Germany) for the CD19-Cre and Cγ1-Cre mice. The indicated DLBCL cell lines in Supplementary Table 5 were a kind gift of L. Staudt, M.A. Shipp and T. Gilmore. This work was supported by US National Institutes of Health Grants RO1-CA172492 (to L.P.) and RO1-CA37295 (to R.D.-F.); JZ is supported by a Lymphoma Research Foundation post-doctoral fellowship. L.P. is on leave from the University of Perugia.

References

1. Shaffer AL 3rd, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. Annual review of immunology. 2012; 30:565–610.
2. Basso K, Dalla-Favera R. Germinal centres and B cell lymphomagenesis. Nature reviews Immunology. 2015; 15:172–184.
3. Morin RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011; 476:298–303. [PubMed: 21796119]
4. Pasqualucci L, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature. 2011; 471:189–195. [PubMed: 21390126]
5. Pasqualucci L, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nature genetics. 2011; 43:830–837. [PubMed: 21804550]
6. Lohr JG, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A. 2012; 109:3879–3884. [PubMed: 22343534]
7. Okosun J, et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. Nat Genet. 2014; 46:176–181. [PubMed: 24362818]
8. Pasqualucci L, et al. Genetics of follicular lymphoma transformation. Cell reports. 2014; 6:130–140. [PubMed: 24388756]
9. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403:503–511. [PubMed: 10676951]
10. Green MR, et al. Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. Blood. 2013; 121:1604–1611. [PubMed: 23297126]
11. Green MR, et al. Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation. Proc Natl Acad Sci U S A. 2015; 112:E1116–1125. [PubMed: 25713363]
12. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem. 2012; 81:65–95. [PubMed: 22663077]
13. Kouzarides T. Chromatin modifications and their function. Cell. 2007; 128:693–705. [PubMed: 17320507]
14. Li B, Carey M, Workman JL. The role of chromatin during transcription. Cell. 2007; 128:707–719. [PubMed: 17320508]

15. Briggs SD, et al. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes & development. 2001; 15:3286–3295. [PubMed: 11751634]
16. Roguev A, et al. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. The EMBO journal. 2001; 20:7137–7148. [PubMed: 11742990]
17. Miller T, et al. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A. 2001; 98:12902–12907. [PubMed: 11687631]
18. Krogan NJ, et al. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J Biol Chem. 2002; 277:10753–10755. [PubMed: 11805083]
19. Herz HM, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes & development. 2012; 26:2604–2620. [PubMed: 23166019]
20. Hu D, et al. The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. Mol Cell Biol. 2013; 33:4745–4754. [PubMed: 24081332]
21. Lee JE, et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. eLife. 2013; 2:e01503. [PubMed: 24368734]
22. Guo C, et al. KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. Oncotarget. 2013; 4:2144–2153. [PubMed: 24240169]
23. Kaikkonen MU, et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Molecular cell. 2013; 51:310–325. [PubMed: 23932714]
24. Ng SB, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature genetics. 2010; 42:790–793. [PubMed: 20711175]
25. Guo C, et al. Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. Proc Natl Acad Sci U S A. 2012; 109:17603–17608. [PubMed: 23045699]
26. Santos MA, et al. DNA damage-induced differentiation of leukemic cells as an anti-cancer barrier. Nature. 2014; 514:107–111. [PubMed: 25079327]
27. Dhar SS, et al. Trans-tail regulation of MLL4-catalyzed H3K4 methylation by H4R3 symmetric dimethylation is mediated by a tandem PHD of MLL4. Genes & development. 2012; 26:2749–2762. [PubMed: 23249737]
28. 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. 2006; 103:7396–7401. [PubMed: 16651521]
29. Rickert RC, Roes J, Rajewsky K. B lymphocyte-specific, Cre-mediated mutagenesis in mice. Nucleic acids research. 1997; 25:1317–1318. [PubMed: 9092650]
30. Carsetti R, Kohler G, Lamers MC. Transitional B cells are the target of negative selection in the B cell compartment. The Journal of experimental medicine. 1995; 181:2129–2140. [PubMed: 7760002]
31. Allman D, et al. Resolution of three nonproliferative immature splenic B cell subsets reveals multiple selection points during peripheral B cell maturation. Journal of immunology. 2001; 167:6834–6840.
32. Jacob J, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. II. A common clonal origin for periarteriolar lymphoid sheath-associated foci and germinal centers. The Journal of experimental medicine. 1992; 176:679–687. [PubMed: 1512536]
33. Lim WK, Lyashenko E, Califano A. Master regulators used as breast cancer metastasis classifier. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing. 2009:504–515. [PubMed: 19209726]
34. Pao LI, et al. B cell-specific deletion of protein-tyrosine phosphatase Shp1 promotes B-1a cell development and causes systemic autoimmunity. Immunity. 2007; 27:35–48. [PubMed: 17600736]
35. Rudd CE. Lnk adaptor: novel negative regulator of B cell lymphopoiesis. Science's STKE: signal transduction knowledge environment. 2001; 2001:pe1.
36. Sauer K, Cooke MP. Regulation of immune cell development through soluble inositol-1,3,4,5-tetrakisphosphate. Nature reviews Immunology. 2010; 10:257–271.

37. Liu YJ, et al. Mechanism of antigen-driven selection in germinal centres. Nature. 1989; 342:929–931. [PubMed: 2594086]
38. Egle A, Harris AW, Bath ML, O'Reilly L, Cory S. VavP-Bcl2 transgenic mice develop follicular lymphoma preceded by germinal center hyperplasia. Blood. 2004; 103:2276–2283. [PubMed: 14630790]
39. Beguelin W, et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. Cancer Cell. 2013; 23:677–692. [PubMed: 23680150]
40. Zhang J, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A. 2013; 110:1398–1403. [PubMed: 23292937]
41. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010; 107:21931–21936. [PubMed: 21106759]
42. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]
43. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]
44. Chapuy B, et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell. 2013; 24:777–790. [PubMed: 24332044]
45. Ardehali MB, et al. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. The EMBO journal. 2011; 30:2817–2828. [PubMed: 21694722]
46. Calo E, Wysocka J. Modification of enhancer chromatin: what, how and why? Mol Cell. 2013; 49:825–837. [PubMed: 23473601]
47. Petruk S, et al. TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell. 2012; 150:922–933. [PubMed: 22921915]
48. Pasqualucci L. The genetic basis of diffuse large B-cell lymphoma. Curr Opin Hematol. 2013; 20:336–344. [PubMed: 23673341]
49. Pasqualucci L, et al. AID is required for germinal center-derived lymphomagenesis. Nat Genet. 2008; 40:108–112. [PubMed: 18066064]
50. Pasqualucci L, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature. 2001; 412:341–346. [PubMed: 11460166]
51. Hatzi K, Melnick A. Breaking bad in the germinal center: how deregulation of BCL6 contributes to lymphomagenesis. Trends in molecular medicine. 2014; 20:343–352. [PubMed: 24698494]
52. Cattoretti G, et al. Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice. Cancer Cell. 2005; 7:445–455. [PubMed: 15894265]
53. Mandelbaum J, et al. BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. Cancer Cell. 2010; 18:568–579. [PubMed: 21156281]
54. Souers AJ, et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nature medicine. 2013; 19:202–208.
55. Knutson SK, et al. A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. Nature chemical biology. 2012; 8:890–896. [PubMed: 23023262]
56. McCabe MT, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature. 2012; 492:108–112. [PubMed: 23051747]
57. Klein U, et al. Transcriptional analysis of the B cell germinal center reaction. Proc Natl Acad Sci U S A. 2003; 100:2639–2644. [PubMed: 12604779]
58. Compagno M, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. Nature. 2009; 459:717–721. [PubMed: 19412164]
59. Dominguez-Sola D, et al. The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry. Nat Immunol. 2012; 13:1083–1091. [PubMed: 23001145]
60. Thomas M, et al. Full deacylation of polyethylenimine dramatically boosts its gene delivery efficiency and specificity to mouse lung. Proc Natl Acad Sci U S A. 2005; 102:5679–5684. [PubMed: 15824322]
61. Bereshchenko OR, Gu W, Dalla-Favera R. Acetylation inactivates the transcriptional repressor BCL6. Nat Genet. 2002; 32:606–613. [PubMed: 12402037]

62. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25:402–408. [PubMed: 11846609]

63. Green, MR.; Sambrook, J.; Sambrook, J. Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, N.Y.: 2012.

64. Jolly CJ, Klix N, Neuberger MS. Rapid methods for the analysis of immunoglobulin gene hypermutation: application to transgenic and gene targeted mice. Nucleic acids research. 1997; 25:1913–1919. [PubMed: 9115357]

65. Klein U, et al. Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. Nat Immunol. 2006; 7:773–782. [PubMed: 16767092]

66. Reich M, et al. GenePattern 2.0. Nat Genet. 2006; 38:500–501. [PubMed: 16642009]

67. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995; 57:289–300.

68. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007; 23:3251–3253. [PubMed: 17644558]

69. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols. 2009; 4:44–57. [PubMed: 19131956]

70. Blecher-Gonen R, et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nature protocols. 2013; 8:539–554. [PubMed: 23429716]

71. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9:357–359. [PubMed: 22388286]

72. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

73. Giannopoulou EG, Elemento O. An integrated ChIP-seq analysis platform with customizable workflows. BMC bioinformatics. 2011; 12:277. [PubMed: 21736739]

74. Lin CY, et al. Transcriptional amplification in tumor cells with elevated c-Myc. Cell. 2012; 151:56–67. [PubMed: 23021215]

75. Loven J, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013; 153:320–334. [PubMed: 23582323]

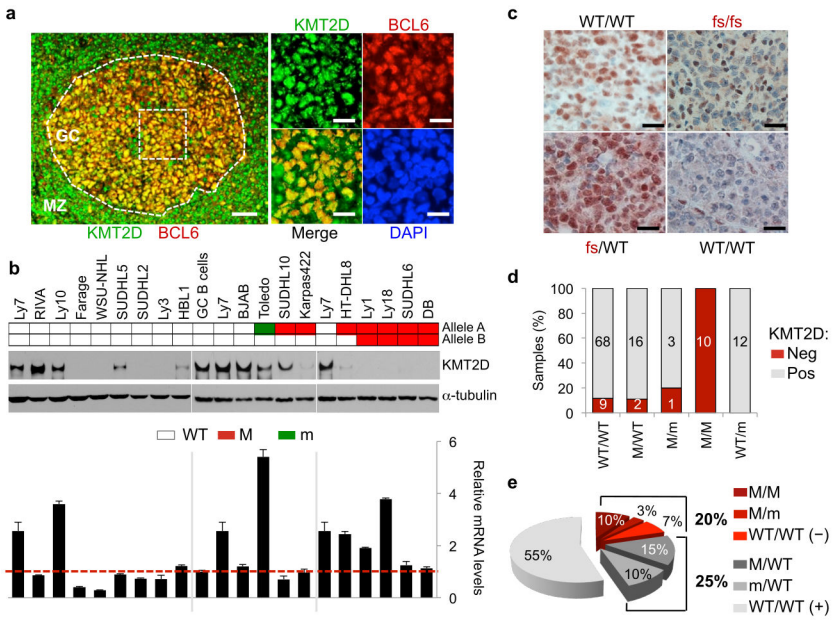

Figure 1. Genetic and epigenetic inactivation of KMT2D in DLBCL

(**a**) Co-immunofluorescence analysis of KMT2D (green) and BCL6 (red) in reactive human tonsils (GC, germinal center; MZ, mantle zone); DAPI is used to visualize the nuclei. Scale bar, 100 μm. Inset, 50 μm (data representative of two experiments). (**b**) Immunoblot (top) and Q-RT-PCR (bottom) analysis of KMT2D expression in DLBCL cell lines; WT, wild-type; M, truncating mutation (nonsense or frameshift); m, missense mutation; α-tubulin, loading control. The red dotted line in the histogram indicates the mRNA levels of normal GC B cells, arbitrarily set at 1 (n = 3, mean ± SD; data representative of two experiments). (**c**) Immunohistochemistry analysis of KMT2D expression (brown) in representative DLBCL biopsies harboring WT or truncated *KMT2D* alleles (fs, frameshift mutation). Hematoxylin stains the nuclei. Scale bar, 50 μm. (**d**) Percentage of KMT2D-protein positive (Pos) and negative (Neg) DLBCL samples, stratified based on genetic status as indicated in panel **b**. The number of samples in each category is provided inside the bar. (**e**) Overall proportion of DLBCL samples with defective KMT2D function, either complete (owing to biallelic genetic lesions or lack of protein expression [−]; red scale) or partial (due to monoallelic genetic lesions; grey scale) (total n = 123).

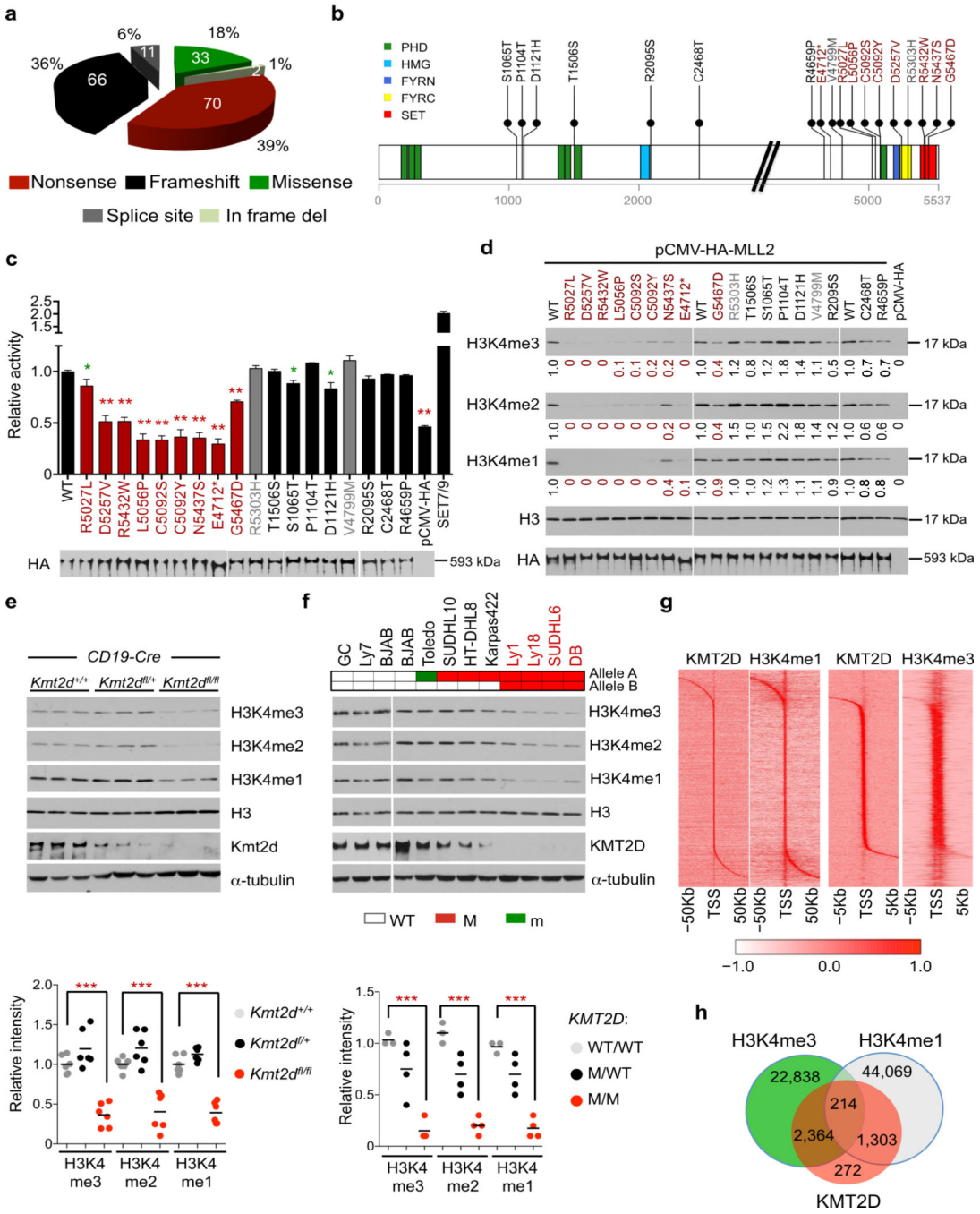

Figure 2. *KMT2D* mutations are associated with defective methyltransferase activity and diminished H3K4 methylation

(**a**) *KMT2D* mutation types in DLBCL and tFL. (**b**) Schematic representation of the KMT2D protein (PHD, plant homeodomain; HMG, High Mobility Group; FYRN, FY-rich, N-terminal; FYRC, FY-rich, C-terminal; SET, Su(var)3-9, Enhancer-of-zeste, Trithorax). Red, missense mutations affecting C-terminal domains; grey, germline variants; the truncated E4712* protein served as control. (**c**) H3K4 methyltransferase activity of semipurified HA-KMT2D proteins against a synthetic H3 peptide, measured by ELISA.

Data are shown as fold differences over wild-type, set as 1, and represent one of two independent experiments performed in triplicate (mean ± SD). *: $P < 0.05$; **: $P < 0.01$ (Student's *t*-test). Bottom, immunoblot analysis of HA. (**d**) KMT activity of mutated HA-KMT2D proteins against Hela nucleosomes. Quantification of signal intensity after normalization to total H3 is shown below each lane, with wild-type set as 1. (**e**) Immunoblot analysis of splenic B cells from *Kmt2d*$^{+/+}$, *Kmt2d*$^{fl/+}$ and *Kmt2d*$^{fl/fl}$ *CD19-Cre* mice. Image is representative of 3 independent experiments, and the bottom panel shows signal intensity quantification of each histone mark, normalized for H3, in a total of 6 mice/genotype. ***, $P < 0.001$, one way ANOVA. (**f**) Immunoblot analysis of GC B cells and DLBCL cell lines representative of various *KMT2D* genetic configurations. Bottom panel, quantification of histone marks. ***, $P < 0.001$, one way ANOVA. (**g**) Enrichment profile of KMT2D, H3K4me1 and H3K4me3 high-confidence peaks relative to the nearest TSS; windows of ± 50 kb and ± 5 kb are shown for the two epigenetic marks, associated with active enhancers and promoters, respectively. Genes are ranked based on relative distance between the KMT2D peak and the closest TSS. (**h**) Overlap between KMT2D, H3K4me3 and H3K4me1 ChIP-seq peaks in human GC B-cells.

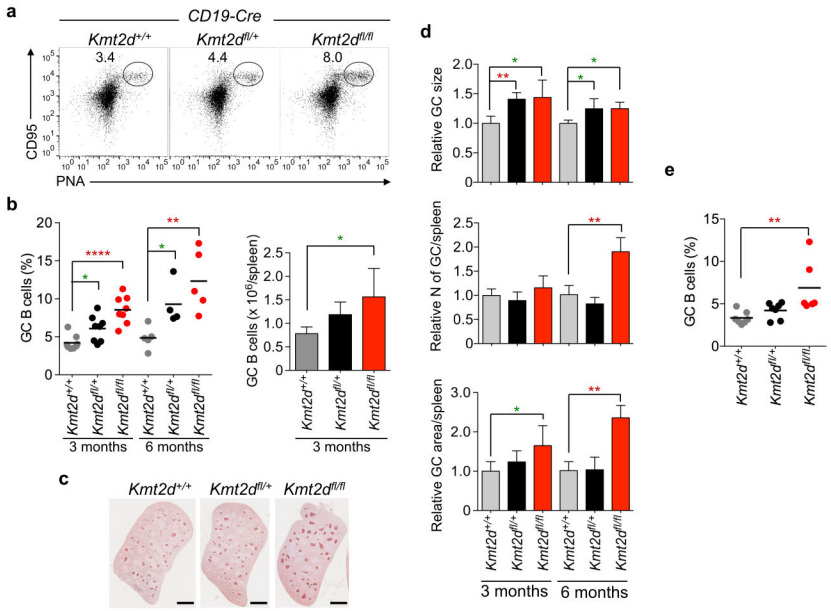

Figure 3. Deletion of *Kmt2d* early in B cell development leads to increased GC formation

(**a**) Representative flow cytometric analysis of splenic B220$^+$ cells from *Kmt2d*$^{fl/fl}$, *Kmt2d*$^{fl/+}$ and *Kmt2d*$^{+/+}$ *CD19-Cre* mice analyzed 10 days after SRBC immunization. Numbers in each panel indicate the percentage of CD95$^+$PNA$^+$ GC B cells in the B220$^+$ gate. (**b**) Percentage (left panel; $n = 8$ each) and absolute number (right panel; mean ± SD, $n = 7$ each) of GC B cells in SRBC-immunized mice from the indicated genotypes, analyzed at 3 and 6 months of age. Data representative of two independent experiments. (**c**) IHC analysis of BCL6 expression (brown) in representative FFPE spleen sections from the same animals. Scale bar, 500 μm. (**d**) Quantification of mean GC size, GC number and total GC area/spleen in *Kmt2d*$^{fl/fl}$ *CD19-Cre*, *Kmt2d*$^{fl/+}$ *CD19-Cre* and *Kmt2d*$^{+/+}$ *CD19-Cre* animals analyzed at 3 and 6 months of age, 10 days after SRBC immunization. Analysis was performed on Bcl6-immunostained sections using the ImageJ software, and values represent the mean ± SD of 4–5 mice in each group, after normalization to the total spleen area, with average levels from controls set at 1. (**e**) Percentage of GC B cells in the mesenteric lymph nodes of the same animals ($n = 7$ *Kmt2d*$^{+/+}$, 7 *Kmt2d*$^{fl/+}$ and 6 *Kmt2d*$^{fl/fl}$). Data representative of two experiments. *: $P < 0.05$, **: $P < 0.01$, ****: $P < 0.0001$, one way ANOVA.

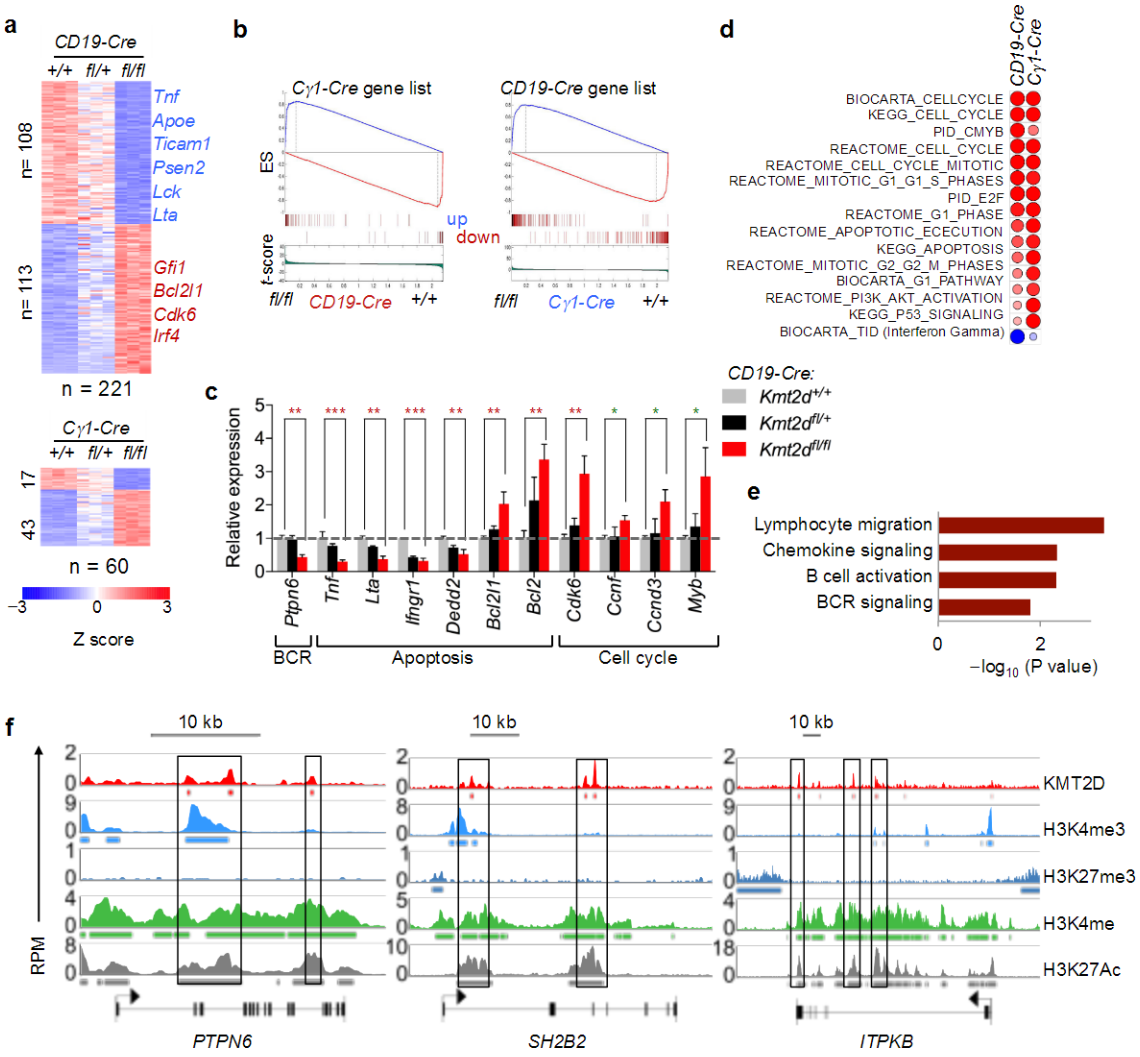

Figure 4. *Kmt2d* deletion induces differentiation stage-dependent transcriptional changes in GC B cells

(**a**) Differentially expressed genes in *Kmt2d*$^{fl/fl}$ vs *Kmt2d*$^{+/+}$ GC B cells from *CD19-Cre* and *Cγ1-Cre* cohorts (*Kmt2d*$^{fl/+}$ mice shown by default; $n = 3$ mice/genotype; FDR < 0.15; FC > 1.5). (**b**) Left: extended GSEA analysis of genes differentially expressed in *CD19-Cre* GC B cells, along the T-score rank of transcripts differentially expressed in *Cγ1-Cre* cells (see methods). The reciprocal analysis is shown in the right panel. (**c**) Q-RT-PCR of representative genes differentially expressed in *Kmt2d*$^{fl/fl}$ *CD19-Cre* vs *Kmt2d*$^{+/+}$ *CD19-Cre* GC B cells. Data are displayed as fold change (mean ± SD) over the mean value of *Kmt2d*$^{+/+}$ *CD19-Cre* samples, set as 1 (dashed line) ($n = 3$ mice/genotype, in triplicate). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one way ANOVA. (**d**) Representative significantly enriched pathways in *Kmt2d*-deficient vs *Kmt2d*-wild-type cells, identified by GSEA (see Supplementary Table 2 for complete list); colored circles denote the enrichment *P* value in the two models (the smaller the *P* value, the larger the size and the color intensity). Red and blue scale denotes up-regulation and down-regulation in *Kmt2d*-deficient cells, respectively (FDR < 0.15). (**e**) DAVID pathway enrichment analysis of KMT2D core targets (i.e. genes bound by KMT2D in human GC B cells, enriched in active histone marks, and down-

regulated in *Kmt2d*$^{fl/fl}$ *CD19-Cre* GC B cells). (**f**) Genomic snapshots of KMT2D peaks at 3 identified targets involved in BCR signal transduction, showing co-localization with H3K4me3 at promoters or H3K27Ac and H3K4me1 at enhancers.

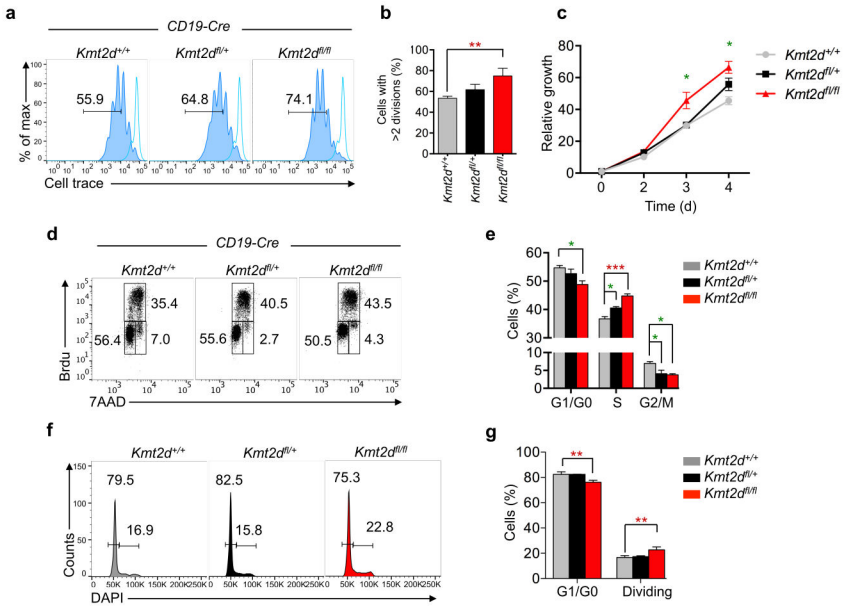

Figure 5. *Kmt2d*$^{fl/fl}$ *CD19-Cre* B cells exhibit proliferative advantage

(**a**) Representative proliferation profiles of B220^{+} splenocytes from *Kmt2d*$^{fl/fl}$, *Kmt2d*$^{fl/+}$ *and Kmt2d*$^{+/+}$ *CD19-Cre* mice, labeled with the CellTrace Violet dye and cultured *ex vivo* in the presence of anti-CD40 and IL4 for 3 days. Individual peaks in the plot correspond to different numbers of cell divisions, with the unstimulated parent population shown as an empty line. Numbers indicate the percentage of cells with >2 divisions, and the quantification of 3 samples/genotype (mean ± SD) is shown in (**b**). (**c**) Relative growth of B220^{+} splenocytes in the same experiment. Data are indicated as fold changes relative to the mean value of control mice, measured at day 0 (mean ± SD; n = 3 mice/genotype, plated in triplicate). (**d**) Representative cell cycle profiles of stimulated B220^{+} cells, labeled with BrdU and 7AAD and analyzed by flow cytometry at day 3 after stimulation. Numbers indicate the percentage of cells in the gated population. (**e**) Quantification of panel **d** (mean ± SD; n = 3 mice/genotype). (**f**) Representative cell cycle profiles (DNA content) of GC B cells from the indicated mice, analyzed 10 days after SRBC immunization. Gates identify dividing (S/G2/M) and non-dividing (G0/G1) cells, and the mean ± SD obtained from 3 mice/genotype is plotted in panel (**g**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, one way ANOVA.

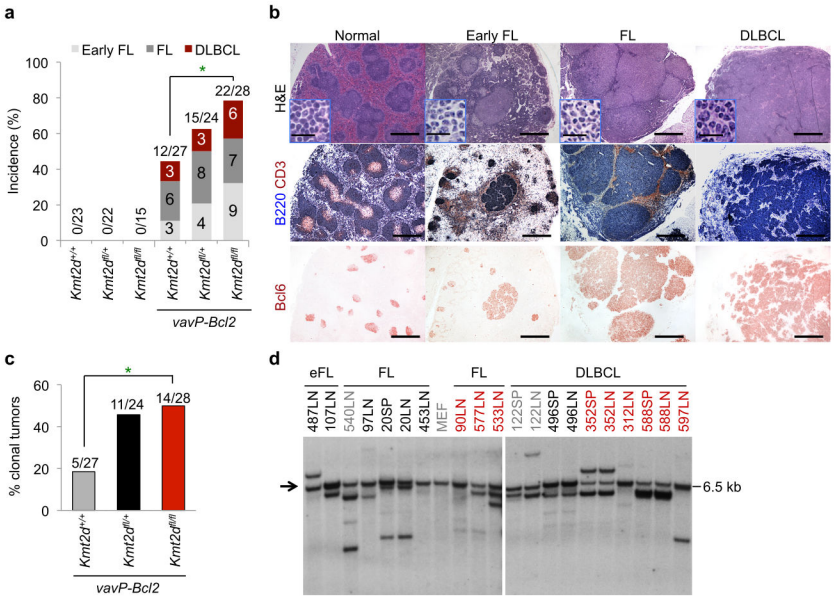

Figure 6. Deletion of *Kmt2d* in the GC cooperates with BCL2 deregulation to promote lymphomagenesis

(**a**) Incidence of B cell lymphoproliferative diseases in mice of the indicated genotypes. Color-coded segments correspond to distinct lymphoma entities diagnosed according to criteria that are described in the Methods section. The number of animals in each category is given inside the bar, and the overall number of animals diagnosed with lymphoma, out of the total analyzed, is on top. (**b**) Histologic and immunohistochemical analysis of representative spleens from control animals and *VavP-Bcl2/Kmt2d*$^{fl/fl}$*/Cγ1-Cre* mice diagnosed with eFL, FL and DLBCL.Top: H&E staining (scale bar, 1,250 μm; inset, 60 μm). Double immunostaining for B220 and CD3 (middle) identifies the B and T cell zones, while Bcl6 (bottom) serves as GC-specific marker (scale bar, 1,250 μm). (**c**) Overall percentage of clonal lymphoproliferations in mice of the indicated genotypes. (**d**) Southern blot analysis of EcoRI digested DNA from representative tumors. Mouse embryonic fibroblasts (MEF) served as control for the germline immunoglobulin heavy chain EcoRI fragment (arrow; expected size, ~6.5 kb). Color-coded text denotes the genotype of the corresponding mouse (red, *Kmt2d*$^{fl/fl}$; black, *Kmt2d*$^{fl/+}$; grey, *Kmt2d*$^{+/+}$, all in *VavP-Bcl2/Cγ1-Cre* background). Note the presence of the same clone in both lymph node (LN) and spleen (SP) biopsies of representative animals. * $P < 0.05$, Fisher's exact test.