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

Jiyuan Zhang¹, David Dominguez-Sola¹–³, Shafinaz Hussein¹,10, Ji-Eun Lee⁴, Antony B Holmes¹, Mukesh Bansal², Sofija Vlasevskaja¹, Tongwei Mo¹, Hongyan Tang¹, Katia Basso¹,6, Kai Ge⁴, Riccardo Dalla-Favera¹,6–⁹ & Laura Pasqualucci¹,6,⁹

Mutations in the gene encoding the KMT2D (or 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 these mutations and their role in lymphomagenesis are unknown. Here we show that FL- and DLBCL-associated KMT2D mutations impair KMT2D enzymatic activity, leading to diminished global H3K4 methylation in 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. Moreover, genetic ablation of Kmt2d in mice overexpressing Bcl2 increases the incidence of GC-derived lymphomas resembling human tumors. 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 thus 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 GC and are driven by distinct genetic lesions disrupting key oncogenic pathways¹–². Recent exome and 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 lymphomagenesis³–⁸. Among these genes, KMT2D (also known as MLL2, ALR or MLL4) is the one most frequently mutated, being found in ~30% of de novo DLBCLs (including both molecular subtypes: germinal center B cell–like (GCB) DLBCL and activated B cell–like (ABC) DLBCL)³ and ~90% of FLs³,⁵–⁸,¹⁰,¹¹, which together account for >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 that are 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 GC³,⁷,⁸,¹⁰,¹¹.

KMT2D encodes a conserved member of the SET1 family of histone lysine methyltransferases (KMTs), which catalyze the methylation of lysine 4 on histone H3 (H3K4)—a modification associated with transcriptionally active chromatin¹²–¹⁴. The enzymatic function of KMT2D depends on a cluster of conserved C-terminal domains, including a plant homeodomain (PHD), two phenylalanine and tyrosine (FY)-rich motifs (FY-rich, N-terminal (FYRN) and FY-rich, C-terminal (FYRC)) and a catalytic Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) domain. Whereas a single multisubunit complex known as ‘Complex of Proteins Associated with Set 1’ (COMPASS) is responsible for all methylation of H3K4 in yeast¹⁵–¹⁷, six different KMTs have been identified in higher eukaryotes, which fall into three subgroups on the basis of homologies in protein sequence and subunit composition: SET1A-SET1B, MLL1-MLL4 (KMT2A-KMT2B) and MLL3-MLL2 (KMT2C-KMT2D)¹²–¹⁴. These findings suggest that the three KMT complexes exert nonoverlapping, highly specialized functions by regulating the transcription of discrete subsets of genes. In particular, KMT2C-KMT2D localize and function as major histone H3K4 mono- and dimethyltransferases at enhancers in mammalian cells¹⁹–²³. Consistent with the involvement of KMT2D in critical cellular functions, germline homozygous deletion of this gene in mice results in embryonic lethality²¹.

In DLBCL and FL, KMT2D mutations are predominantly represented by premature stop codons, frameshift insertions or deletions and splice-site mutations that are predicted to generate truncated proteins lacking part or all of the C-terminal protein domain³,⁵–⁸. Additionally, multiple missense mutations have been found across the

¹Institute for Cancer Genetics, Columbia University, New York, New York, USA. ²Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ³Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ⁴Laboratory of Endocrinology and Receptor Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA. ⁵Department of Systems Biology, Columbia University, New York, New York, USA. ⁶Department of Pathology and Cell Biology, Columbia University, New York, New York, USA. ⁷Department of Genetics & Development, Columbia University, New York, New York, USA. ⁸Department of Microbiology & Immunology, Columbia University, New York, New York, USA. ⁹Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York, USA. ¹⁰Present address: Department of Pathology and Laboratory Medicine, North Shore–LIJ, Staten Island University Hospital, Staten Island, New York, USA. Correspondence should be addressed to L.P. (l.p171@columbia.edu).

Received 1 April; accepted 11 August; published online 14 September 2015; doi:10.1038/nm.3940
KMT2D protein, but their functional consequences remain unexplored. In 30–75% of the mutated tumor samples KMT2D genetic lesions are biallelically distributed, whereas in the remainder of the samples, one intact KMT2D allele is retained, 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 to be the causative event in a rare congenital disease known as Kabuki syndrome, an observation that offers direct proof for the dose-dependent pathogenic effect of this enzyme in other tissues24.

A few studies have investigated the biochemical function of KMT2D in mammals (in 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 human and mature B cell subpopulations. Consistent with the ubiquitous nature of other MLL family members, KMT2D transcripts were detected in purified naive, GC and memory B cells (Supplementary Fig. 1). Accordingly, coimmunofluorescence 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) that are representative of various KMT2D genomic configurations (wild-type alleles, biallelic truncating mutations and heterozygous truncating or missense mutations). 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, we did not detect expression of the intact protein in all of the samples carrying biallelic truncating mutations that eliminated this epitope (Fig. 1b,c). Moreover, immunoblot analysis using antibodies directed against the N-terminal half of KMT2D failed to reveal any signal corresponding to the truncated proteins in all seven cell lines tested (Supplementary Fig. 2f,g), even after treatment with the proteasomal inhibitor MG132 (data not shown), despite expression of the mutant cDNA, suggesting that mechanisms other than proteosomal degradation are responsible for the lack of expression of the truncated proteins in these cells.

We then investigated whether the residual wild-type allele is still expressed in samples carrying KMT2D monoallelic (truncating) mutations. Whereas the KMT2D protein was barely detectable in 2 of 18 samples analyzed, the remaining cell line and all 15 biopsies stained positive for the full-length protein (Fig. 1b–d). Of note, markedly reduced amounts of protein, or none, were also observed in 9 of the 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, ~45% of DLBCLs lack KMT2D either completely, due to biallelic genetic lesions and loss of protein expression (20% of cases), or partially, due to monoallelic mutations (25% of cases) (Fig. 1e).

Missense mutations impair KMT2D methyltransferase 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% of tHLS in our data set harbor missense mutations that affect various residues along the KMT2D protein (18 of 89 mutational events, 20.2%; and 33 of 182 events from published DLBCL and tHL genomic studies, 18.1%) (Fig. 2a and Supplementary Fig. 3)3,5–8.
To test the functional impact of KMT2D missense mutations, we generated hemagglutinin (HA)-tagged versions of 16 lymphoma-derived mutant alleles (Fig. 2b and Supplementary Fig. 4a) and measured their basal expression upon transient transfection in 293T cells. With one exception (S5404F), all of the alleles tested produced similar amounts of both mRNA and protein as those from the wild-type allele or from three germline variants that were included as controls, indicating that missense mutations generally do not interfere with the stability of the KMT2D protein (Supplementary Fig. 4b).

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\textsuperscript{27}. In both assays, each of the nine mutations that were present in the C-terminal domains significantly impaired enzymatic activity as compared to that of the wild-type protein (Fig. 2c,d). In particular, immunoblot analysis of nucleosomal histones using specific antibodies to distinct methylation states revealed a substantial reduction in H3K4 monomethylation (H3K4me1), dimethylation (H3K4me2) and trimethylation (H3K4me3) (Fig. 2d). In contrast, the enzymatic activities of seven polypeptides carrying mutations in the N-terminal half of the protein were only modestly affected or indistinguishable from those of either the wild-type control or the germline mutants (Fig. 2b–d). These amino acid changes

**Figure 2** KMT2D mutations are associated with defective methyltransferase activity and diminished H3K4 methylation. (a) KMT2D mutation types in DLBCL and IFL. (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). In red, missense mutations affecting C-terminal domains; in gray, germline variants. The truncated E4712\textsuperscript{*} protein served as a kinase-dead control for the experiments in Figure 2c,d. (c) H3K4 methyltransferase activity of semipurified HA-KMT2D proteins on a synthetic H3 peptide, as measured by ELISA. Data are shown as fold differences over the activity of wild-type HA-KMT2D, which was set as 1, and represent one of two independent experiments performed in triplicate (data are mean ± s.d.). *P < 0.05, **P < 0.01 (Student’s t-test). Bottom, immunoblot analysis for HA. (d) Lysine methyltransferase activity of mutated HA-KMT2D proteins on HeLa nucleosomes. Quantification of signal intensity after normalization to total H3 is shown below each lane. For normalization purposes, the ratio of (H3K4 methylation level)/(total H3 levels) in the wild-type HA-KMT2D sample was set to 1. (e) Immunoblot analysis of splenic B cells from Kmt2d\textsuperscript{+/−}, Kmt2d\textsuperscript{−/−} and Kmt2d\textsuperscript{wt/Kmt2d\textsuperscript{−/−}} CD19-Cre mice. Top, immunoblot image is representative of three independent experiments. Bottom, signal intensity quantification of each histone mark, normalized to H3, in a total of six mice per 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, quantification of histone marks. ***P < 0.001, one-way ANOVA. (g) Enrichment profile of KMT2D, H3K4me1 (associated with active promoters) and H3K4me3 (associated with active promoters) high-confidence peaks relative to the nearest transcription start site (TSS); windows of ±50 kb and ±5 kb are shown for the two epigenetic marks. Genes are ranked based on the relative distance between the KMT2D peak and the closest TSS. (h) Overview of 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,b) Representative flow cytometry analysis (a) and quantification (b) of splenic B220+ cells from Kmt2dfl/fl, Kmt2d+/+ and Kmt2d−/−CD19-Cre mice analyzed 10 d after SRBC immunization. (a) Numbers in each panel indicate the percentage of CD95+PNA+ GC B cells in the B220+ gate. (b) Percentage (left graph; n = 8 per genotype at 3 months of age; 5 Kmt2d+/+, 4 Kmt2d+/+ and 5 Kmt2d−/− at 6 months of age) and absolute number (right graph; mean ± s.d., n = 7 Kmt2d+/+, 6 Kmt2d+/+ and 6 Kmt2d−/−) of GC B cells in SRBC-immunized mice from the indicated genotypes, analyzed at 3 and 6 months of age.

Data are representative of two independent experiments. (c) IHC analysis of BcI6 expression (brown) in representative spleen sections from the same animals. Scale bar, 500 μm. (d) Quantification of mean GC size and GC number or total GC area per spleen section in Kmt2d+/+CD19-Cre, Kmt2d−/−CD19-Cre and Kmt2d+/+CD19-Cre animals analyzed at 3 and 6 months of age, 10 d after SRBC immunization. Analysis was performed on BcI6-immunostained sections using the ImageJ software, and values represent the mean ± s.d. of 4 Kmt2d+/+, 5 Kmt2d+/+ and 5 Kmt2d+/+ CD19-Cre mice after normalization to the total spleen area, with average levels from controls set to 1. (e) Percentage of GC B cells in the mesenteric lymph nodes of the same animals (n = 7 Kmt2d+/+, 7 Kmt2d+/+ and 6 Kmt2d−/−). Data are representative of two experiments. *P < 0.05, **P < 0.01, ****P < 0.0001, one-way ANOVA.

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 could interfere with other modes of KMT2D regulation.

Although additional studies will be required to dissect the precise mechanisms underlying the observed reduction in activity of the C-terminal mutants (such as changes in 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+/+CD19-Cre and Kmt2d+/+Cγ1-Cre) in which the Kmt2d gene had been conditionally deleted via Cre-mediated recombination (see below). When compared to those from wild-type littermates, Kmt2d-deficient B cells exhibited significantly (P < 0.001) 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 as compared to that of 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 was decorated by H3K4me1 (1,517/4,153 peaks, 36%) and/or H3K4me2 (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 nonredundant methyltransferase that controls, either directly or indirectly, the methylation state of a large number of regions in the mature B cell compartmen.

Early loss of Kmt2d affects follicular B cell development Since the stage of hematopoietic differentiation in which KMT2D lesions occur is unknown, we generated mouse models in which Kmt2d could be conditionally deleted at distinct stages of B cell differentiation by crossing mice bearing a floxed Kmt2d allele21 with mice expressing the Cre recombinase specifically either in GC B cells (Kmt2dfl/+/Cγ1-Cre)28 or in B cell progenitors (Kmt2dfl/+/CD19-Cre)29. Animals were analyzed before and 10 d 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 in the distribution of bone marrow (BM) B cell subpopulations, splenic follicular (FO) B cells and marginal zone (MZ) B cells between unimmunized Kmt2dfl/+/CD19-Cre and littermate Kmt2d+/+ and Kmt2d+/+ controls (Supplementary Fig. 5). The percentages and absolute numbers of pro-, pre- and immature B cells in the BM were comparable even after SRBC immunization (Supplementary Fig. 6a,b). However, immunized Kmt2dfl/+/CD19-Cre mice displayed a significant (P < 0.05) reduction in splenic and lymph node B220+ cells (Supplementary Fig. 6c,d) as well as 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) precursors30, we then examined the distribution of various B cell subsets in the spleens of immunized mice. While no significant differences were noted in the absolute numbers of transitional31 and MZ B cells, FO B cells were markedly less numerous 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 B cells.

Early deletion of Kmt2d enhances GC formation In the Cγ1-Cre cohort, Kmt2dfl/+, Kmt2d+/+ and Kmt2d+/+ mice did not show significant differences in the fraction of GC B cells (B220+CD95+PNA (peanut agglutinin)hi), despite efficient ablation
of both Kmt2d RNA and protein (Supplementary Fig. 7a–d). Accordingly, Kmt2dfl/flCre1-Cre and Kmt2dfl/flCre1-Cre mice were indistinguishable from wild-type controls in the size and number of GCs, in the numbers of IgG1* B cells and plasma cells and in the total IgG1 serum levels (Supplementary Fig. 7e–g). Thus, GC development occurs normally when Kmt2d is deleted as early as 4 d after the initiation of the GC reaction28.

In contrast, SRBC-immunized Kmt2dfl/flCD19-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 numbers of surface IgG1–expressing cells (Supplementary Fig. 8). Analogous results were obtained when animals were challenged with the clonally restricted T cell–dependent antigen NP-KLH (4-hydroxy-3-nitrophenylacetyl hapten conjugated to keyhole limpet hemocyanin) (Supplementary Fig. 9a–d). Interestingly, mutant mice responded efficiently to NP-KLH, as documented by the increase of NP-binding serum IgM and

Figure 4 Kmt2d deletion induces differentiation stage–dependent transcriptional changes in GC B cells. (a) Differentially expressed genes in Kmt2dfl/flCre1-Cre versus Kmt2dfl/flCre1-Cre GC B cells from CD19-Cre and Cy1-Cre cohorts (Kmt2dfl/flCre1-Cre mice are also shown; n = 3 mice per 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 Cy1-Cre cells (see Online Methods). The reciprocal analysis is shown in the right panel. (c) qRT-PCR of representative genes differentially expressed between Kmt2dfl/flCre1-Cre and Kmt2dfl/flCre1-Cre GC B cells. Data are displayed as fold change (mean ± s.d.) over the mean value of Kmt2dfl/flCre1-Cre samples that was set as 1 (dashed line) (n = 3 mice per genotype, in triplicate). *P < 0.05, **P < 0.01, ***P < 0.001, one-way ANOVA. (d) Representative significantly (P < 0.05) enriched pathways in Kmt2d-deficient versus Kmt2d wild-type cells, as identified by GSEA (see Supplementary Table 2 for a 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 scales denote upregulation and downregulation, respectively, in Kmt2d-deficient cells (FDR < 0.15). (e) DAVID pathway enrichment analysis of KMT2D core targets (genes bound by KMT2D in human GC B cells, enriched in active histone marks and downregulated in Kmt2dfl/flCre1-Cre GC B cells). Benjamini-corrected P value (see also Supplementary Table 3). (f) Genomic snapshots of KMT2D peaks at three identified targets involved in BCR signal transduction, showing co-localization with H3K4me3 at promoters or lysine acetylation at H3K27 (H3K27Ac) and H3K4me1 at enhancers. H3K27me3 is an epigenetic mark of poised or silenced promoters. RPM, reads per million mapped reads.
IgG1 antibodies 12 and 28 d after immunization; however, the magnitude of the response was ~10-fold lower than in control mice (Supplementary Fig. 9e). Accordingly, a smaller 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 (which represents 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 early deletion of Kmt2d (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 versus late B cell development, we contrasted the gene expression profiles (GEP) of GC B cells isolated from immunized Kmt2dfl/fl, Kmt2dfl/+/ and Kmt2d+/+ animals in both the Cyll-Cre and the CD19-Cre cohorts. We found that 221 genes were differentially expressed between Kmt2d-deficient and Kmt2d-proficient cells in the CD19-Cre background, whereas 60 genes were differentially expressed in the Cyll-Cre compound mice (Student’s t-test, false discovery rate (FDR) < 0.15, fold change (FC) ≥ 1.5), suggesting that the stage and/or timing of Kmt2d 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 approach33 revealed a significant (P < 0.001) 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, independently of whether the ablation of Kmt2d occurred before or after GC initiation. Consistent with this observation, both Cyll-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., Ccdn3, 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 downregulated 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., the genes occupied by KMT2D at active promoters and/or enhancers and downregulated following Kmt2d deletion; see Online Methods) were significantly (P < 0.05) enriched in biological programs with key functions in B cell physiology and pathology, including B cell receptor activation and signaling, chemokine and cytokine signaling and lymphocyte migration34–36 (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 the loss of Kmt2d early on, before GC induction, facilitates the establishment and influences the magnitude of these transcriptional changes.
Kmt2d loss confers a B cell–proliferative advantage

On the basis of 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 apoptotic program, we first measured the proliferative capacity of B220+ cells that were isolated from the spleens of CD19-Cre and Cyt1-Cre mice and stimulated with CD40-specific antibody and IL4, which mimics the signals delivered in vivo to induce GC responses. By day 3 poststimulation, CD19-Cre Kmt2d-deficient cells (but not Cyt1-Cre Kmt2d-deficient cells; Supplementary Fig. 11) had undergone a significantly (P < 0.01) higher number of cell divisions and had generated greater numbers of viable cells, as 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% versus 35% in Kmt2d wild-type controls; P < 0.001; Fig. 5d,e). The increased response to the stimulation with CD40-specific antibody and IL4 is unlikely to be due to an altered distribution of B cell subsets in the splenic B220+ population of Kmt2d-deficient mice, as the proportion of FO and MZ B cells in these mice under unimmunized conditions was similar to those in the 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 (P < 0.01; Fig. 5f,g). Conversely, we did not detect measurable changes in apoptosis under these conditions (data not shown). Taken together, these findings indicate that loss of Kmt2d provides a 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 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 and t(14;18) cases and in ~36% of KMT2D-mutated DLBCL cases. Toward this end, we crossed Kmt2dfl/fl/Cyt1-Cre animals with VavP-Bcl2 transgenic mice, which develop FL preceded by GC hyperplasia in 37–50% of cases, and monitored the resulting mice for tumor incidence and survival over a period of 18 months.

We did not observe significant differences between Kmt2dfl/fl/Cyt1-Cre and Kmt2dfl/fl/Cyt1-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, all of which died by 13 months of age, probably because of the severe glomerulonephritis and autoimmune diseases that develop in this strain at nearly 100% penetrance, as previously reported.

However, histological analysis of the lymphoid organs revealed that, compared to VavP-Bcl2 littermate controls, a significantly higher proportion of Kmt2dfl/fl/Cyt1-Cre × VavP-Bcl2 mice had evolved to develop clonal B cell lymphoproliferative disorders that recapitulated the spectrum of phenotypes associated with the progression of FL to DLBCL: that is, early stages of FL (eFL), overt FL of various grades and DLBCL (overall, 22/27 (78.6%) Kmt2dfl/fl mice versus 15/24 (62.5%) Kmt2dfl/fl mice and 12/27 (44.4%) Kmt2dfl/fl mice; 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 data not shown). Furthermore, Southern blot analysis of the immunoglobulin locus documented a clonal origin in 92% of the DLBCLs, 55% of the FLs and 23.5% of the eFLs tested (Fig. 6c,d). The inability to detect clonal rearrangements in many of the latter can probably be explained by the limited representation of tumor cells in the biopsy (Fig. 6b; see the eFL panel), as such, the larger number of clonal lymphoproliferations observed in the Kmt2d-knockout mice might reflect a more advanced degree of tumor infiltration or aggressiveness. Most notably, 7 of 7 clonal tumors that were 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 that recapitulate the features of 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—missense mutations affecting its C-terminal cluster of SET domains and, in a small subset of cases, loss of protein expression, implicating epigenetic silencing. Thus, multiple modes of inactivation are responsible for defective KMT2D function in nearly 50% of cases of DLBCL and possibly in all cases of FL.

We documented that in samples with monoallelic KMT2D mutations, which constitute up to two-thirds of affected DLBCL cases and ~45% of FL and 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 knockout littersmates. This suggests that dose reduction of KMT2D may also be pathogenic in the context of B cells, in a manner analogous to the haploinsufficient role of KMT2D in Kabuki syndrome. The reason why a subset of tumors retain one copy of KMT2D is not known, but it 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 paralog KMT2C—which was shown to have partial functional redundancy and is also mutated in these diseases, albeit at lower frequencies. Although KMT2C mutations have not been functionally characterized and the number of individuals carrying these genetic aberrations is too low to assess statistical significance, these lesions seem more common in cases with monoallelic versus biallelic KMT2D lesions (5/42 versus 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. Whereas recent studies in mammalian cells have emphasized the role of KMT2D as the major enzyme responsible for H3K4me1 at enhancer regions and KMT2D-deficient alleles in this study were also consistently associated with diminished trimethylation, suggesting a nonredundant role in mature B cells. Accordingly, KMT2D-bound regions in GC B cells display widespread overlap with both H3K4me1 at putative GC enhancers (including predicted super-enhancers) and H3K4me3 at gene promoters (Supplementary Notes and Supplementary Fig. 10b,c), consistent with the requirement of Drosophila Trithorax-related (Trr) enhancer-promoter communication for faithful and tissue-specific gene expression.

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 out of 593 putative super-enhancers identified in GC B cells, many of which were assigned to master regulators of the GC phenotype on the basis of proximity to the nearest transcriptionally active TSS (Supplementary Notes and data not shown).

One key observation of this study is that stage-specific deletion of Kmt2d (before or 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 can be explained by two non–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 unmodified histone H3 (ref. 47). In support of this hypothesis, the GEPS of Kmt2d-null GC B cells in which the deletion occurred before or after GC formation were qualitatively similar; moreover, the reduction in global histone methylation was less evident in GC B cells purified from Cre mice as compared to that from CD19-Cre mice (compare Supplementary Fig. 7b and Fig. 2e). On the other hand, the stage of differentiation or cellular context in which KMT2D is lost may have a specific impact on the target cell by remodeling its epigenetic landscape 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 a greater than twofold 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, and (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. 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. Indeed, GC enlargement has been previously shown to favor lymphomagenesis in several mouse tumor models (such as the Bcl6 knock-in and Blimp1 knockout mouse models).

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., 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 premalignant precursor. The animal model reported here may thus be useful for the preclinical testing of therapeutic approaches that combine inhibitors to BCL2 and histone methyltransferases or demethylases.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Gene Expression Omnibus database: the gene expression profile data of mouse GC B cells and the ChIP-seq data of human GC B cells have been deposited under accession numbers GSE67388 and GSE67494, respectively.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

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. McIntosh (Bethyl Laboratories, Inc.) for help in generating several antibodies to KMT2D, 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 and the Genomic Technologies Shared 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

© 2015 Nature America, Inc. All rights reserved.
and Cyt-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 R01-CA174292 (L.P.) and R01-CA37295 (R.D.-F.), and J.Z. is supported by a Lymphoma Research Foundation postdoctoral fellowship. L.P. is on leave from the University of Perugia.

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 mouse autopsies and processed tissues for FACS and histologic analysis. R.D.-F. contributed to the study design and data analysis. All authors reviewed the manuscript and provided final approval for submission.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.
Online Methods

Cell lines and isolation of human B cell subsets. DLBCL cell lines were obtained as detailed in Supplementary Table 5 and were cultured in Iscove's Modified Dulbecco's Medium (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Sigma-Aldrich), penicillin (100 U/ml) and streptomycin (100 µg/ml), 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 fluorescent in situ hybridization (FISH) analysis or by targeted resequencing of previously reported mutated genes (when applicable), unless purchased from commercial sources. HEK 293T cells were cultured in Dulbecco's Modified Eagle Medium (Life Technologies) with 10% FBS, penicillin (100 U/ml) and streptomycin (100 µg/ml). All cell lines were verified to be free from Mycoplasma contamination. Human GC centroblasts, naive B cells and memory B cells were isolated from reactive tonsils as described[7].

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 or 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[5,6,7], resulting in a total of 250 DLBCL primary cases, 31 DLBCL cell lines and 71 tFL.

Immunohistochemistry 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% of tumor cells showed KMT2D 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 KMT2D-specific antibody (1:200 dilution) and a mouse monoclonal antibody to BCL6 (clone GI91E/A8, Cell Marque). Detection of KMT2D was obtained using the EnVision System–HRP–Rabbit antibody (Dako) followed by Tyramide Signal Amplification system (PerkinElmer); the second-ary 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 PhuTurbo Cx DNA Polymerase (Agilent) and, as template, cDNA from human GC B cells, followed by cloning into the Zero Blunt TOPO PCR vector (Life Technologies). The DLBCL and FL-derived missense mutant alleles and the R4712 truncation mutant were generated from the wild-type pCMV-HA-KMT2D construct using the QuickChange II Site-Directed Mutagenesis Kit (Agilent). All plasmids were verified for integrity by diagnostic restriction enzyme digestions using the Sall, Notl, BglIII, SacII, XbaI and MfI sites, and subjected to Sanger sequencing of the full-length KMT2D coding sequence (CD5).

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 h after transfection and used for protein extraction or for isolation of semipurified HA-KMT2D proteins, as described below.

Protein extraction. Whole cell extracts were obtained from cell lines in the 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[59]. For the analysis of histone modifications, the chromatin pellet was subsequently resuspended in RIPA-coIP buffer (Tris-HCl pH 8.0 50 mM NaCl 200 mM Glycerol 5%, MgCl2 2.5 mM, 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 r.p.m. for 10 min.

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 before 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 used to amplify specific cDNA fragments with the oligonucleotides listed in Supplementary Table 6, in the 7300 Real Time PCR system (Applied Biosystems). Data were analyzed by the change-in-threshold (2–ΔΔCt) method[53], using Actin as a housekeeping reference gene.

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 (designated as KMT2D-1185 in Supplementary Fig. 2 and Supplementary Table 7), mouse monoclonal antibody to 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 7. 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 mM EDTA), followed by overnight incubation with E2view™ 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 r.p.m. for 10 min. KMT2D protein amounts were quantified by Coomassie staining and immunoblot analysis using rat monoclonal antibodies to HA (clone 3F10, Roche). KMT activity against an artificial H3 peptide was assayed using the EpiQuik Histone Methyltransferase Activity-Inhibition Assay Kit (H3K14 (Epigentek), following the manufacturer’s protocol. Relative activity was calculated as the fold change in OD500nm over the mean reading of wild-type samples. Experiments were performed in triplicate and repeated independently at least twice.
Enzymatic activity against native nucleosomes was measured following a published method. Briefly, equal amounts of wild-type or mutant HA-KMT2D proteins were incubated at 37 °C for 2 h 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 5-adenosyl methionine (New England BioLabs). Reactions were stopped by adding equal volumes of 2× Laemmli buffer and heated at 100 °C for 5 min before loading onto Tris-glycine 4–20% gradient gels. All assays were performed twice independently.

Mice. The conditional Kmt2d knockout mouse model (mixed C57BL/6 × Sv129 background) has been published. Deletion of Kmt2d in GC B cells or in precursor B cells was obtained by breeding Kmt2dfl/fl mice with CD19-Cre or Cd19-Cre mice (both in pure C57BL/6 background), followed by offspring intercrossing. Compound mice carrying the conditional Kmt2d allele (either CD19-Cre or Cd19-Cre) and the VavP-Bcl2 transgene were obtained by breeding with VavP-Bcl2 mice (C57BL/6 background). A sample size of ≥23 animals per 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′-ATTGCATCATGCCAATGCAGC-3′ (forward) and 5′-GGACAAGGCTGCTATGTGCAC-3′ (reverse). Animals were monitored for tumor incidence and survival twice a week over a period of 18 months and were killed 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- to 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 per mouse. The mixture was left to stand at room temperature for 30 min before injection to allow erythrocytes to settle. The number of viable cells in the spleen by multiplying the fraction of B220+ cells in the FACS profiles by the total number of erythrocyte-depleted cells (60% of the total number of erythrocyte-depleted cells in these organs).

Flow cytometric analysis. Multicolor flow cytometric analysis of B and T cell lymphoid compartments was performed at 3 and 6 months of age as previously reported. The T cell compartment was also examined in all animals from the tumor–watch cohort after they were killed. 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 7). Samples were acquired on a FACSCanto II (BD Biosciences) flow cytometer. Experiments were performed in triplicate at day 3, 4 and 5 after stimulation with CD40-specific antibody and IL4.

Histological and immunohistochemical analysis of mouse tissues. Histological analysis of mouse organs was performed on 4-µm thick FFPE tissue sections stained with H&E (Thermo Scientific). 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 Cdt1−Cre cohort were diagnosed independently by two pathologists (D.D.-S. and S.H.) on the basis of morphology and phenotype using the following criteria, in accordance with the classification of human lymphoma: (i) early stages of FL (eFL), as defined by the presence of one or a 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, as 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 and 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. Four micrograms of DNA were digested with EcoRI overnight at 37 °C. The digestion reaction was resolved on a 0.8% agarose gel followed by transfer to nitrocellulose membrane, according to standard procedures. Hybridization was performed at 37 °C overnight using a [γ-32P]-labeled probe, as reported.

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 dilution of a fluorescent dye. Data were acquired at day 2, 3 and 4 of stimulation on a FACS Canto II (BD Biosciences) flow cytometer with 488-nm excitation and an emission filter in the 480-500 nm range. All experiments were performed in at least two independent groups of animals, in biological triplicates.

Cell cycle analysis. To assess the cell cycle profiles of Kmt2dfl/fl versus Kmt2dfl/+ and Kmt2dfl/+ B cells in the CD19−Cre and Cdt1−Cre cohorts, B220+ splenocytes were labeled with bromodeoxyuridine (BrdU) for 2 h and analyzed for BrdU incorporation and DNA content using the APC BrdU flow kit (BD Biosciences) on a FACS Canto II (BD Biosciences) flow cytometer. Experiments were performed in triplicate at day 3, 4 and 5 after stimulation with CD40-specific antibody and IL4.

G C 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 min 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 min, as described. 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 FlowJo 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 per 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 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+ cells in the FACS profiles by the total number of erythrocyte-depleted cells in these organs.
Mutation analysis of the rearranged immunoglobulin genes in mouse tumors. The rearranged immunoglobulin variable heavy-chain region (Igh-V) genes were amplified from mouse tumor DNA by PCR, using forward primers that anneal to the framework region III of the most abundantly used Igh-V558 (5′-CAGCCTGTACATGAGGACTCTG-3′), Igh-V7183 (5′-GAASAMCCGTGTCCCTGAAATGASC-3′) and Igh-VQ52 (5′-CAGGTCGAGCTGAAARCTGCA-3′) family, in separate reactions, and a reverse primer positioned in the jh4 intron (5′-ACTATCCCTCAGCCATAGG-3′). 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 either the same primers as in the PCR reaction or additional internal primers annealing to the jh2 (5′-AGATGGAGGCCAGTGAGGAC-3′) and jh3 (5′-GTCACTGACCTGAAATTGTC-3′) segments (GENEWIZ, Inc.). Sequences were aligned to the NCBI databases and the international ImMunoGeneTics information system (IMGT) database and were analyzed as previously described.

ELISA assay. Total IgM and IgG1 serum titers were measured by ELISA as previously described, 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 NP25-BSA or NP, respectively (SouthernBiotech), and bound antibodies were detected with alkaline phosphatase–conjugated IgM- or IgG1-specific antibodies (SouthernBiotech). Pan–mouse immunoglobulins were captured on plates coated with 2.5 µg/ml of goat anti-mouse Ig (H+L) (SouthernBiotech). Each assay was performed in duplicate using 12 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 Igh-V66.2 genes, which 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 the individual clones (n = 48/mouse) were sequenced directly after colony PCR. Igh-V gene sequences were analyzed with the HighV-QUEST tool (http://www.imgt.org) and compared to the most homologous mouse germline counterparts to assess the overall mutation frequency and to identify clones harboring the W3L substitution. Differences between groups were determined by a two-tailed Fisher’s exact test in the GraphPad Prism software.

Gene expression profiling. B220+CD95+PNA+ GC B cells were sorted from Kont2d[+]/+ and Kont2d[−/−] splenocytes (CD19-Cre or Cyl-Cre; n = 3 mice per 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 20 ng of RNA from each sample was reverse transcribed and amplified using the Ovation RNA Amplification System (NuGEN) followed by labeling with the Encore Biotin Module (NuGEN). The cross-comparison of the signatures obtained in Kont2d-deficient versus Kont2d-proficient GC B cells from the two genetic backgrounds was performed using an extended GSEA approach. At variance with GSEA, which 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, extended GSEA also takes into account the direction of regulation of each gene in the gene set while calculating the enrichment. As described in ref. 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.

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, and we obtained 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 (ref. 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. ChIP-seq produced a total of >24 million (KMT2D), >20 million (H3K4me3), >24 million (H3K4me1) and >24 million (H3K27Ac) uniquely mapped reads per sample, respectively. Reads were normalized to total reads aligned (counts per million). Peak detection was done using ChiPseeker v2.0 (ref. 75) 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 enrichment in predefined sets of genes with GSEA-P68 using the Canonical Pathway (CP) Molecular Signature Database gene sets collection and 1,000 gene set permutations. Enrichments were considered significant if FDR 50.15 after correction for multiple hypotheses.

DAVID. To determine whether genes bound by KMT2D at active promoters and enhancers, and downregulated in Kont2d-deficient GC B cells, were enriched in annotated functional categories, we overlapped the list of downregulated 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 (https://david.ncifcrf.gov/) and interrogated for enrichment in Pathways as annotated in BIOCARTE, 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 reacting tonsils as described, were crosslinked with 1% formaldehyde for 10 min at room temperature, quenched by the addition of glycerine 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 nuclear isolation was performed using the TruChIP High Cell Chromatin Shearing Kit with SDS shearing buffer (Covaris). Nuclei were sonicated using the S220 Ultrasonicator (Covaris) to obtain chromatin fragments of 200–500 bp. Sheared chromatin was incubated overnight with 4 µg of antibodies to KMT2D (HPA035977, Sigma-Aldrich), H3K27Ac (Active Motif, cat#67153) or H3K4me1 (Abcam, cat#ab8380), or with 2 µg of antibody to H3K4me1 (Abcam, cat#ab8895)(ENCODE Project: http://; genome.ucsc.edu/ENCODE/antibodies.html). Protein A magnetic beads were then added for 4 h, followed by sequential washes at increasing stringency and reverse cross-linking. After RNase and proteinase K treatment, ChIP DNA was purified using the MinitElute Reaction Cleanup Kit (Qiagen) and quantified by Quant-iT Picogreen dsDNA Reagent (Life Technologies).
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) (±2 kb). 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 binominal tests considering the total size (in bp) of the significant peaks identified in the two ChIP-seq experiments (such as KMT2D versus 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 kb to +1 kb) and enriched in H3K4me1 but not H3K27Ac. Intragenic (exonic or intronic) and intergenic regions were considered to be active enhancers if they were located distal to a TSS (greater than ±2 kb) and enriched in H3K4me1 and H3K27Ac but not in H3K4me3 (refs. 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 data set. Occupancy of KMT2D at super-enhancers was then determined on the basis of overlap between the KMT2D peaks and the 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 gene44.

**Statistical analyses.** \( P \) values were calculated using 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 the GraphPad Prism 5 software. The graphical representation of the data was generated using GraphPad Software, 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.

References

57. Klein, U. et al. Transcriptional analysis of the B cell germinal center reaction. Proc. Natl. Acad. Sci. USA 100, 2639–2644 (2003).
58. Compagno, M. et al. Mutations of multiple genes cause deregulation of NF-κB in diffuse large B cell lymphoma. Nature 459, 717–721 (2009).
59. Dominguez-Sola, D. et al. The proto-oncogene is required for selection in the germinal center and cyclic reentry. Nat. Immunol. 13, 1083–1091 (2012).
60. Thomas, M. et al. Full deacetylation of polyethyleneimine dramatically boosts its gene delivery efficiency and specificity to mouse lung. Proc. Natl. Acad. Sci. USA 102, 5679–5684 (2005).
61. Bereshchenko, O.R., Gu, W. & Dalla-Favera, R. Acetylation inactivates the transcriptional repressor BCL6. Nat. Genet. 32, 606–613 (2002).
62. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25, 402–408 (2001).
63. Green, M.R. & Sambrook, J. Molecular Cloning: A Laboratory Manual 4th edn. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2012).
64. Jin, C.J., Klix, N. & Neuberger, M.S. Rapid methods for the analysis of immunoglobulin gene hypermutation: application to transgenic and gene-targeted mice. Nucleic Acids Res. 25, 1913–1919 (1997).
65. Klein, U. et al. Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. Nat. Immunol. 7, 773–782 (2006).
66. Reich, M. et al. GenePattern 2.0. Nat. Genet. 38, 500–501 (2006).
67. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57, 289–300 (1995).
68. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J.P. GSEA-P: a desktop application for gene set enrichment analysis. Bioinformatics 23, 3251–3253 (2007).
69. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).
70. Blecher-Gonen, R. et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat. Protoc. 8, 539–554 (2013).
71. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
72. Li, H. et al. The sequence alignment–map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
73. Giannopoulou, E.G. & Elemento, O. An integrated ChIP-seq analysis platform with customizable workflows. BMC Bioinformatics 12, 277 (2011).
74. Lin, C.Y. et al. Transcriptional amplification in tumor cells with elevated c-Myc. Cell 151, 56–67 (2012).
75. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 153, 320–334 (2013).