

Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma

Jessica Okosun^{1,14}, Csaba Bödör^{1,2,14}, Jun Wang^{3,14}, Shamzah Araf¹, Cheng-Yuan Yang⁴, Chenyi Pan^{5,6}, Sören Boller⁴, Davide Cittaro⁷, Monika Bozek⁸, Sameena Iqbal¹, Janet Matthews¹, David Wrench¹, Jacek Marzec³, Kiran Tawana¹, Nikolay Popov¹, Ciaran O'Riain¹, Derville O'Shea¹, Emanuela Carlotti¹, Andrew Davies⁹, Charles H Lawrie¹⁰, András Matolcsy², Maria Calaminici¹, Andrew Norton¹¹, Richard J Byers¹², Charles Mein⁸, Elia Stupka⁷, T Andrew Lister¹, Georg Lenz¹³, Silvia Montoto¹, John G Gribben¹, Yuhong Fan^{5,6}, Rudolf Grosschedl⁴, Claude Chelala³ & Jude Fitzgibbon¹

Follicular lymphoma is an incurable malignancy¹, with transformation to an aggressive subtype representing a critical event during disease progression. Here we performed whole-genome or whole-exome sequencing on 10 follicular lymphoma-transformed follicular lymphoma pairs followed by deep sequencing of 28 genes in an extension cohort, and we report the key events and evolutionary processes governing tumor initiation and transformation. Tumor evolution occurred through either a 'rich' or 'sparse' ancestral common progenitor clone (CPC). We identified recurrent mutations in linker histone, JAK-STAT signaling, NF- κ B signaling and B cell developmental genes. Longitudinal analyses identified early driver mutations in chromatin regulator genes (*CREBBP*, *EZH2* and *KMT2D* (*MLL2*)), whereas mutations in *EBF1* and regulators of NF- κ B signaling (*MYD88* and *TNFAIP3*) were gained at transformation. Collectively, this study provides new insights into the genetic basis of follicular lymphoma and the clonal dynamics of transformation and suggests that personalizing therapies to target key genetic alterations in the CPC represents an attractive therapeutic strategy.

Follicular lymphoma, the most common indolent non-Hodgkin lymphoma, remains a significant clinical burden, as the majority of affected individuals undergo multiple relapses and eventually develop resistance to standard therapies. Furthermore, in a subset of individuals, follicular lymphoma transforms into the more aggressive diffuse

large B cell lymphoma (DLBCL), with this transformation associated with poor clinical outcomes^{2–4}. Recent genetic profiling and case studies of donor-derived follicular lymphoma following stem cell transplantation suggest the putative existence of a 'long-lived' tumor-initiating progenitor cell compartment from which successive disease events occur^{5–8}. An in-depth characterization and chronicling of the underlying genetic events from follicular lymphoma to transformed follicular lymphoma will serve as a guide in the development of effective targeted therapies. To address this need, we conducted whole-genome or whole-exome sequencing of sequential follicular lymphoma-transformed follicular lymphoma (FL-tFL) pairs and matched germline samples from 10 cases (Fig. 1a and Online Methods), with deep targeted sequencing of 28 genes in an extension cohort (Supplementary Fig. 1 and Supplementary Table 1).

For the ten index cases (six whole-genome sequenced and four whole-exome sequenced), the sequencing yielded mean genomic and exonic coverage of 37 $\times$ and 110 $\times$, respectively, with 96% of the targeted bases covered by >10-fold (Supplementary Table 2). In the 6 whole-genome sequenced cases, we detected approximately 10,000 somatic variants per tumor (Supplementary Table 3). G>A/C>T transitions were the most common nucleotide substitutions, consistent with data from other cancer types (Supplementary Fig. 2a). We observed a higher frequency of transversions in the transformation-specific variants ($P = 0.006$; Fig. 1b and Supplementary Fig. 2b), reflected in the lower transition/transversion (Ti/Tv) ratios for transformed samples ($P = 0.008$; Supplementary Fig. 2c). Focusing on protein-altering changes, we identified between 21 and 143 nonsynonymous somatic

¹Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK. ²1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary. ³Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London, UK. ⁴Department of Cellular and Molecular Immunology, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. ⁵School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA. ⁶Parker H. Petit Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia, USA. ⁷Centre for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Milan, Italy. ⁸Genome Centre, Barts and the London School of Medicine and Dentistry, London, UK. ⁹Cancer Sciences Division, University of Southampton, Southampton, UK. ¹⁰Oncology Department, Biodonostia Research Institute, San Sebastian, Spain. ¹¹Department of Histopathology, Christie National Health Service (NHS) Foundation Trust, Manchester, UK. ¹²Department of Histopathology, Manchester Royal Infirmary, Manchester, UK. ¹³Department of Hematology, Oncology and Tumor Immunology, Charité Universitätsmedizin, Berlin, Germany. ¹⁴These authors contributed equally to this work. Correspondence should be addressed to J.O. (j.e.okosun@qmul.ac.uk) or C.B. (bodor.csaba1@med.semmelweis-univ.hu).

Received 13 August; accepted 2 December; published online 22 December 2013; doi:10.1038/ng.2856

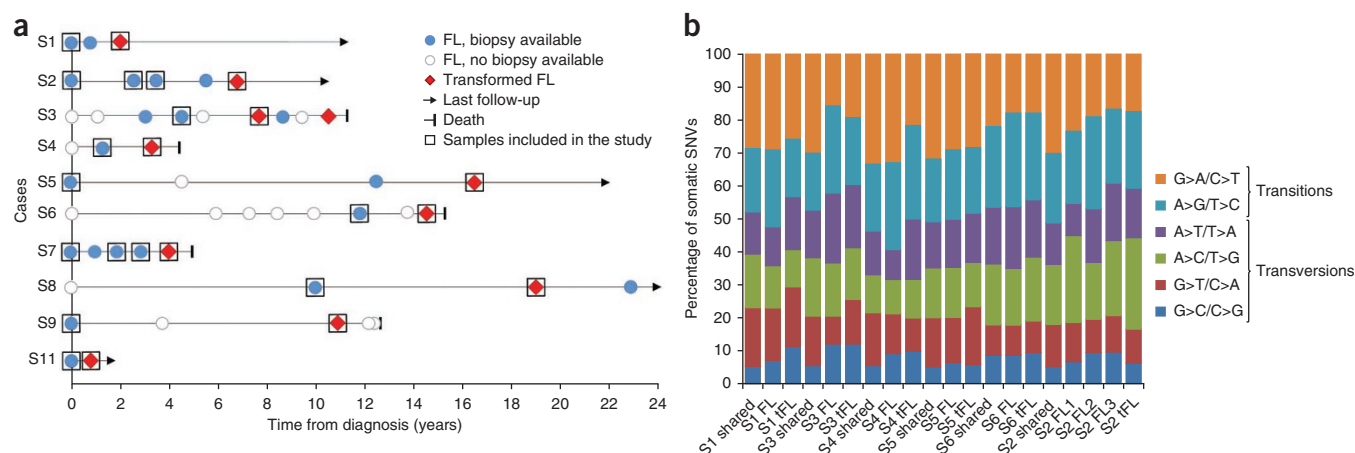


Figure 1 Clinical timelines and somatic mutation profiles for the discovery cases. **(a)** Biopsy information with disease event timelines for the ten whole-genome or whole-exome sequenced follicular lymphoma (FL) cases. **(b)** Distribution of base substitution patterns for all somatic SNVs identified in the six paired follicular lymphoma–transformed follicular lymphoma (FL–tFL) genomes. Somatic SNVs for each case were sorted into three categories: shared (variants present in both follicular lymphoma and transformed follicular lymphoma samples), follicular lymphoma specific and transformed follicular lymphoma specific. Aberrant somatic hypermutation (aSHM) did not contribute to the majority of variants detected, with the exception of previously described gene targets (**Supplementary Table 13**).

variants per sample, with an increased frequency of mutations present after transformation in the majority of cases (**Supplementary Fig. 3**). In total, 1,560 protein-altering variants affecting 908 genes were detected across the 10 cases, comprising missense changes (84.8%), short indels (8.9%) and nonsense mutations (6.3%) (**Supplementary Table 4**). We verified 235 single-nucleotide variants (SNVs) and 48 indels out of 293 variants, corresponding to 118 genes, thereby achieving concordance rates of 98% and 88%, respectively (**Supplementary Table 5**). Interrogation of the six whole-genome sequenced cases identified copy number alterations (CNAs), previously unreported structural variants, fusion transcripts and recurrent mutations in noncoding regions (5' UTRs, 3' UTRs and promoter regions) (**Supplementary Fig. 4** and **Supplementary Table 6**).

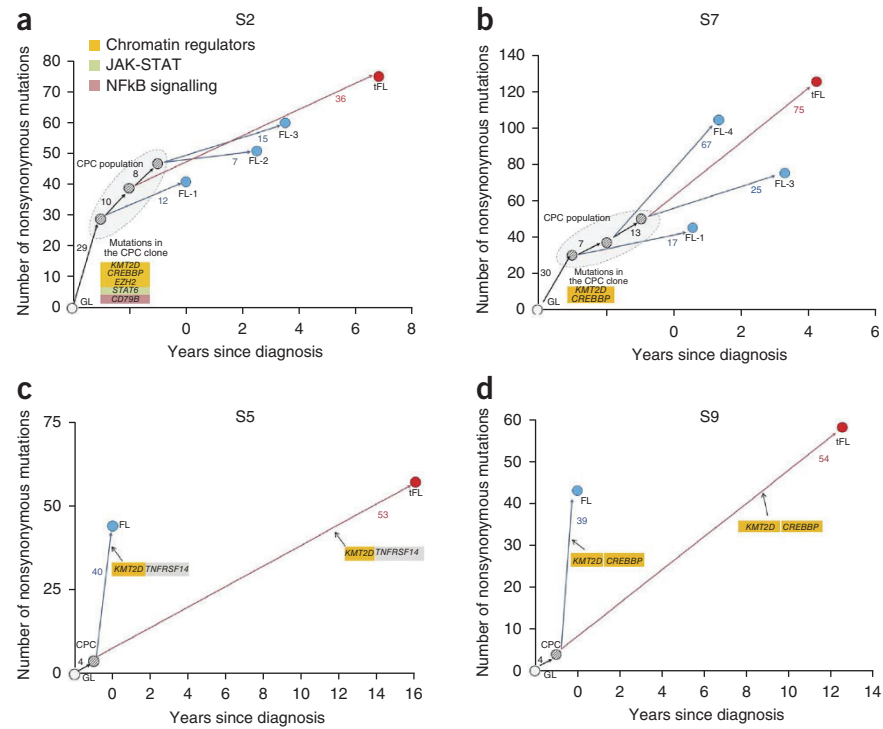
To determine the evolutionary relationship between the 10 paired follicular lymphoma–transformed follicular lymphoma samples, we constructed phylogenetic trees for each case using the somatic variants detected (only nonsynonymous variants and all variants). Each tree showed branching rather than linear evolution, with the 'trunk' representing genetic events shared by the follicular lymphoma tumor(s) and their paired transformed follicular lymphoma samples, thus supporting the presence of an ancestral CPC. Two distinct patterns of evolution from the CPC emerged. In the first pattern, all cases with the exception of S5 and S9 showed high clonal semblance between the paired follicular lymphoma and transformed follicular lymphoma tumors, and we refer to this pattern as evolution through a 'rich' ancestral CPC (**Fig. 2a,b** and **Supplementary Figs. 5a–f** and **6**). The mutations harbored in each CPC demonstrated an enrichment for genes involved in chromatin regulation, with concurrent mutations occurring in *KMT2D* (also known as *MLL2*), which encodes a histone methyltransferase responsible for histone H3 lysine 4 (H3K4) trimethylation, and other histone-modifying genes (*CREBBP*, *EP300*, *EZH2* and *MEF2B*)^{9–12}. Apart from frequent mutations in histone modifiers, other mutational targets in the eight rich CPCs included genes involved in immune modulation (*B2M*, *CD58* and *TNFRSF14*), JAK–STAT signaling (*SOCS1* and *STAT6*) and B cell receptor (BCR)–NF-κB signaling (*BCL10*, *CARD11* and *CD79B*), with many of these genes previously reported in DLBCL^{10–16}. In contrast, a second, unique pattern of evolution was seen in cases S5 and S9, in which

only four nonsynonymous mutations were shared by the follicular lymphoma and transformed follicular lymphoma samples (referred to as evolution through a 'sparse' CPC) (**Fig. 2c,d**). Intriguingly, in both cases, different *KMT2D* (S5 and S9), *TNFRSF14* (S5) and *CREBBP* (S9) mutations occurred in the diagnostic follicular lymphoma and transformed follicular lymphoma biopsies, appearing to have arisen from independent clones and reflecting a convergent pattern of evolution. These observations strongly support tumoral dependency on *CREBBP*, *KMT2D* and *TNFRSF14* alterations during lymphomagenesis and progression.

To evaluate the mutation prevalence for the genes identified in our discovery cohort, we performed deep targeted resequencing of 28 candidate genes (mainly prioritizing genes involved in functional processes not previously implicated in follicular lymphoma (**Supplementary Tables 7** and **8**)) in an extension cohort of 100 independent follicular lymphoma biopsies and 32 paired follicular lymphoma–transformed follicular lymphoma cases (including the 10 index cases), achieving a mean depth of coverage of 840×. A prominent feature of the genetic landscape was the presence of mutations targeting genes involved in chromatin regulation through histones and histone modifications (**Fig. 3** and **Supplementary Fig. 7**), confirming previous studies^{9–12}. Notably, over 70% of cases had concurrent mutations in at least two of the histone-modifying enzymes screened (*CREBBP*, *EZH2*, *MEF2B* and *KMT2D*). *KMT2D* was mutated in 82% of our cohort, with approximately half of the cases harboring multiple mutations. Mutation frequencies for the histone acetyltransferase *CREBBP* and the histone methyltransferase *EZH2* were higher than in earlier studies at 64% and 20%, respectively^{9–11}.

Notably, linker histone genes were recurrently mutated in follicular lymphoma. These genes encode proteins that facilitate the folding of higher order chromatin structures and regulate the access of histone-modifying enzymes and chromatin-remodeling complexes to their target genes^{17–19}. Overall, we observed mutations affecting at least one histone H1 gene in 28% of our series, with *HIST1H1C* and *HIST1H1E* being the most frequently mutated (**Fig. 3** and **Supplementary Table 9**). Most of the histone H1 alterations were missense and were clustered in the highly conserved globular domain (**Supplementary Fig. 8**), targeting residues directly involved in DNA binding^{20–23}. To assess

Figure 2 Patterns of evolution of follicular lymphoma to transformed follicular lymphoma for four cases. (a–d) In each case, a phylogenetic tree was constructed using nonsynonymous mutations and is shown against the corresponding event timeline. All trees have shared mutations (trunk), which represent a common ancestral origin from a CPC, and unique branches, depicting the mutations present in that biopsy alone. The length of individual branches denotes the number of mutations separating the two disease events. S2 (a) and S7 (b), both with three follicular lymphomas (blue) and a single transformed follicular lymphoma (red) biopsies, show divergent clonal evolution from a rich CPC pool (highlighted by the dashed lines) from which subsequent events arise. S5 (c) and S9 (d) show a sparse CPC (four shared mutations between follicular lymphoma and transformed follicular lymphoma samples in both cases) and convergent evolution, in which similar genetic alterations occur independently on separate branches of the tree. All paired tumors were confirmed to be clonally related through *BCL2-IGH* rearrangement analyses. The pattern of clonal evolution was identical with all types of somatic variants considered (Supplementary Fig. 6). Functional annotation of the genes was based on previously reported functions^{10–16}. GL, germ line.



the functional relevance of these mutations, we generated histone H1 (*Hist1h1c*, *Hist1h1d* and *Hist1h1e*) triple-knockout mouse embryonic stem cells (ESCs) overexpressing human *HIST1H1C* encoding wild-type or Ser102Phe histone H1 protein¹⁸ (Supplementary Fig. 9). The p.Ser102Phe alteration in histone H1 resulted in greatly impaired association with chromatin relative to wild-type protein (Fig. 4a,b), suggesting that the identified mutations most likely lead to a loss-of-function phenotype by reducing the binding affinities and residence times of the histone H1 subtypes in chromatin-compromising chromatin compaction and the regulation of specific genes.

We observed frequent mutations in components of the JAK-STAT signaling pathway, including in *STAT6* (12%) and *SOCS1* (8%) (Supplementary Fig. 10). Similar *SOCS1* and *STAT6* mutations have been described in other germinal center lymphomas^{24,25} and have been shown to contribute to constitutive *STAT6* activation and the promotion of tumor cell survival in lymphoma cell lines^{26–28}. Constitutive activation of the antiapoptotic NF-κB signaling pathway, caused by mutations affecting positive and negative regulators of signaling, is an established feature of the activated B cell-like subtype of DLBCL (ABC DLBCL). In contrast, these mutations occur infrequently in germinal center B cell-like DLBCL

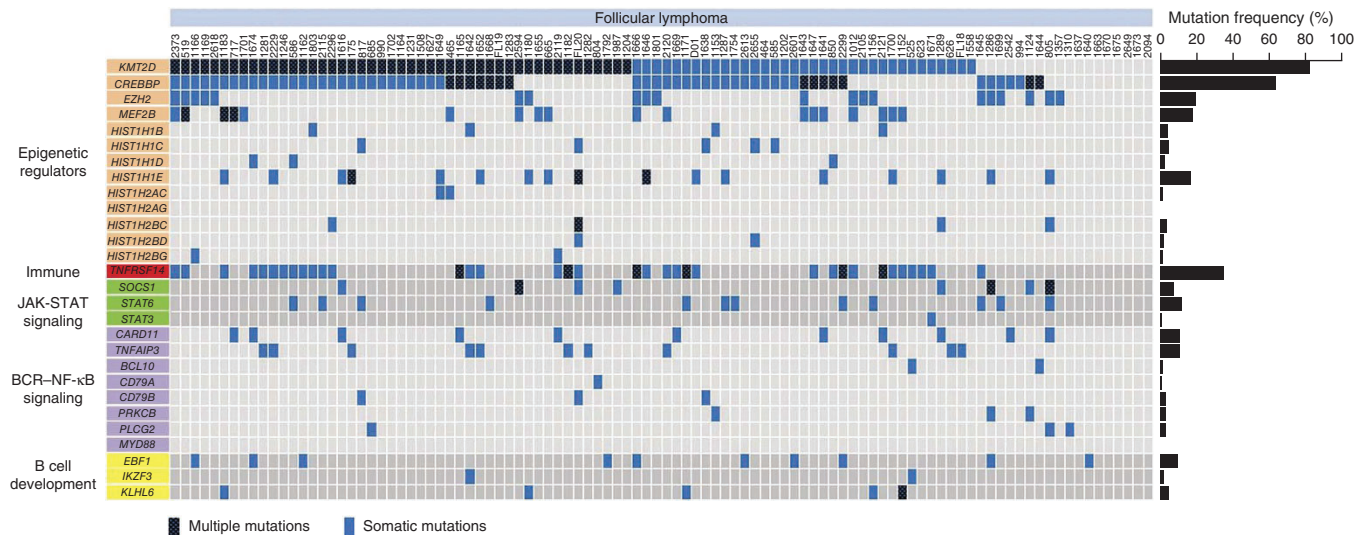
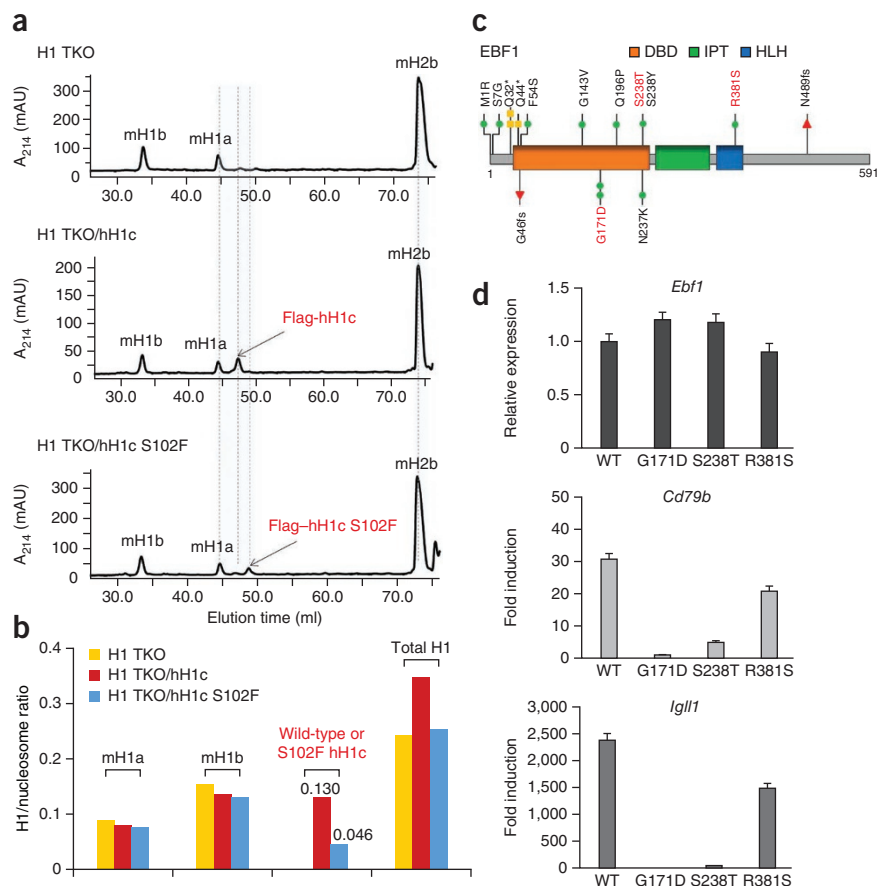


Figure 3 Distribution of mutations in 28 genes in follicular lymphoma. The heat map shows the pattern of case-specific, concurrent and mutually exclusive mutations in 100 follicular lymphoma cases. Each column represents an individual affected case, and each row denotes a specific gene assigned to one of five functional categories (labeled on the left). The bar graph on the right shows the frequency of mutations found for each gene across all samples. A heat map corresponding to the paired follicular lymphoma-transformed follicular lymphoma series is included in Supplementary Figure 7b.

Figure 4 Functional analyses of mutations in linker histone *HIST1H1C* and *EBF1*. (a) Reversed-phase (RP)-HPLC profiles of histones extracted from chromatin isolated from histone H1 triple-knockout (TKO) mouse ESCs expressing wild-type or Ser102Phe human histone H1. The Ser102Phe mutant demonstrated higher hydrophobicity than the wild-type protein. mH1a, mouse histone H1a; mH1b, mouse histone H1b; mH2b, mouse histone H2b; hH1c, human histone H1c. (b) Ratio of individual histone H1 variants (and total histone H1) to the nucleosome of the indicated ESCs. The ratio is calculated^{38,39} from the HPLC analysis in a and demonstrates that the total histone H1 levels in histone H1 triple-knockout ESCs expressing human histone H1c Ser102Phe were reduced compared to cells expressing wild-type human histone H1c, as a result of the weaker association of the mutant histone with chromatin (only 35% of wild-type association). (c) Schematics of the domains and locations of alterations in *EBF1* detected by targeted resequencing (NCBI protein reference sequence [NP_076870.1](#)). DBD, DNA-binding domain; IPT, immunoglobulin domain; HLH, helix-loop-helix domain. Coding mutations are colored green (missense), yellow (nonsense) or red (frameshift). Alterations at the same amino acid in different cases are depicted. Highlighted variants were selected for functional analyses. (d) Quantitative RT-PCR (qRT-PCR)-based gene expression in *Ebf1*-deficient primary mouse pre-pro-B cells transduced with virus expressing wild-type or mutant *Ebf1*. All three *Ebf1* mutants had similar expression levels. Both DBD mutants (Gly171Asp and Ser238Thr) exhibit abrogation of target gene activation. *Igll1* and *Cd79b* transcript levels were normalized to the levels observed with control (pMYs) vector, and *Ebf1* transcript levels of mutant *Ebf1* constructs were normalized to the levels of wild-type *Ebf1* (WT). Error bars, s.d. of duplicate experiments.



(GCB DLBCL)^{13,14,29,30}. Interestingly, we detected mutually exclusive mutations in this pathway in one-third of follicular lymphomas, with a predilection for *CARD11* (11%) and *TNFAIP3* (11%) (Fig. 3). The type and location of the detected aberrations were reminiscent of the mutations found in ABC DLBCL, with the exception of variants in *CD79B* (Supplementary Fig. 10), which included frameshift mutations affecting the immunoreceptor tyrosine-based activation motif (ITAM) domain in follicular lymphoma in contrast to the predominant missense mutations affecting the Tyr196 residue reported in ABC DLBCL¹⁴. A number of emerging therapeutic agents target the NF- κ B signaling cascade, and such treatments may be beneficial to a preselected subset of patients with follicular lymphoma harboring mutations in this pathway.

We identified mutations in genes important in B cell development in 17% of our cohort (Fig. 3 and Supplementary Fig. 10). We focused on early B cell factor 1 (*EBF1*), a transcription factor frequently deleted in high-risk acute lymphoblastic leukemia³¹. The *EBF1* protein functions as a homodimer and binds the consensus sequence 5'-TCCCNNGGGA-3' (refs. 32,33), and more than half of the mutations in our series affected conserved residues of the DNA-binding domain (DBD) (Fig. 4c). Binding of *EBF1* to target genes can elicit different responses, including transcriptional activation or repression or H3K4 dimethylation³⁴. In particular, transcriptional activation of *Igll1*, encoding the $\lambda 5$ surrogate light chain of the pre-B cell receptor, has been used as an assay to assess the binding and function of *EBF1* (ref. 33). Therefore, we tested the effects of three *Ebf1* mutations, encoding p.Gly171Asp and p.Ser238Thr (located in the DBD) and p.Arg381Ser (located in the helix-loop-helix (HLH) domain),

on the ability to activate expression of the *Ebf1* target genes *Igll1* and *Cd79b* in *Ebf1*-deficient primary mouse pre-pro-B cells (Fig. 4d). Gly171Asp and Ser238Thr mutant lines demonstrated markedly impaired upregulation of both *Igll1* and *Cd79b* expression, most likely reflecting reduced binding to regulatory elements in their respective promoters. The HLH mutant, Arg381Ser, resulted in a more modest reduction in target gene expression. Taken together, the data suggest that the *EBF1* mutations detected in follicular lymphoma lead to loss of function and a reduction in *EBF1* target gene expression.

To discriminate early from late genetic events, we integrated variant allele frequencies (VAFs) generated from our high-depth sequencing data with SNP array copy number profiles for 29 cases (26 follicular lymphoma-transformed follicular lymphoma pairs and 3 diagnostic follicular lymphomas), using the ABSOLUTE algorithm³⁵. This scheme allowed us to correct for variations in tumor content and ploidy as well as locus-specific CNAs, thereby establishing the accurate clonal or subclonal status of mutations in our 28 selected genes. Mutations present in nearly all tumor cells (clonal) would suggest early events and therefore represent initiating 'driver' genes. Our results showed that mutations in histone-modifying genes (*KMT2D*, *CREBBP* and *EZH2*) as well as mutations in *STAT6* and *TNFRSF14* were predominantly clonal events (Fig. 5a). These results expand on the findings of a recent study implicating *CREBBP* mutation as an early event in follicular lymphoma³⁶. Interestingly, many of these driver genes (*CREBBP*, *KMT2D* and *TNFRSF14*) harbored compound heterozygous mutations that were frequently accompanied by either deletions or acquired uniparental disomy (aUPD), serving as pathogenic mechanisms to render tumor cells homozygous for a

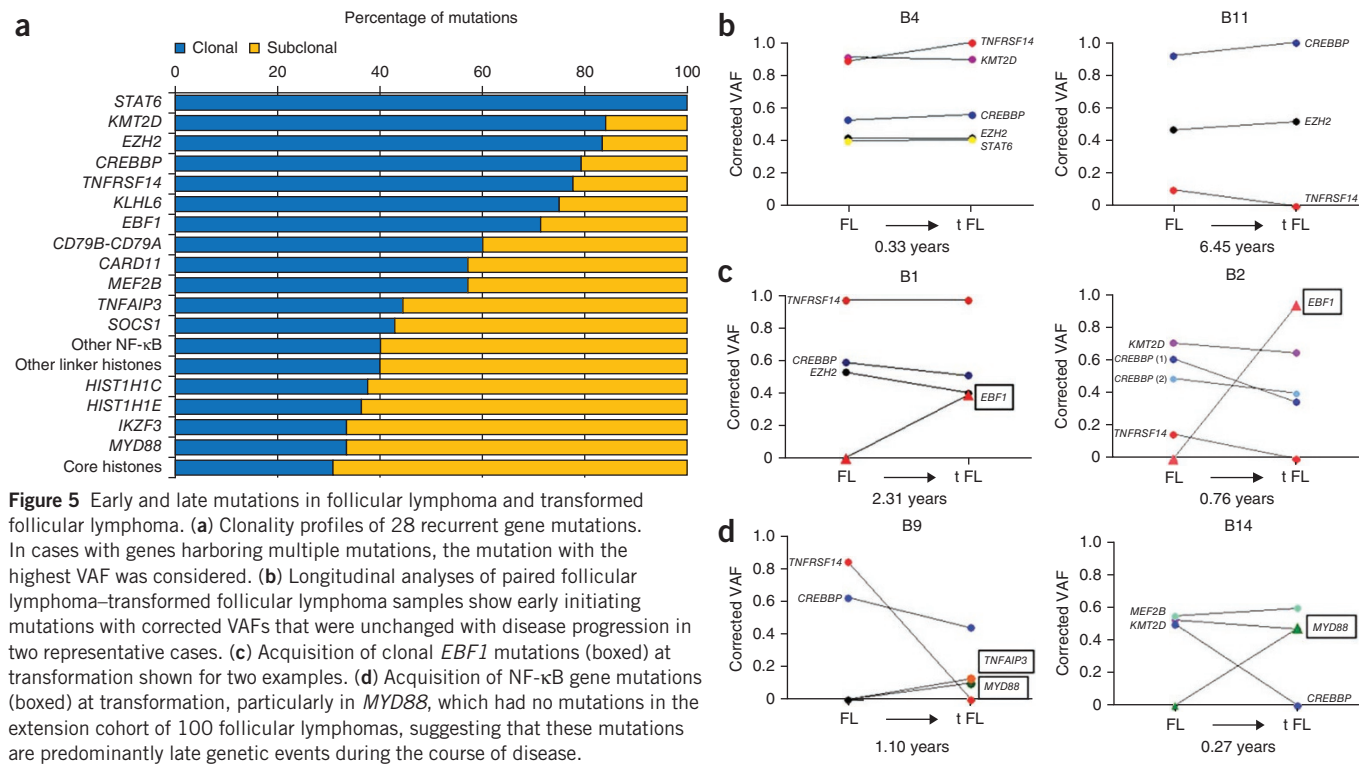


Figure 5 Early and late mutations in follicular lymphoma and transformed follicular lymphoma. **(a)** Clonality profiles of 28 recurrent gene mutations. In cases with genes harboring multiple mutations, the mutation with the highest VAF was considered. **(b)** Longitudinal analyses of paired follicular lymphoma-transformed follicular lymphoma samples show early initiating mutations with corrected VAFs that were unchanged with disease progression in two representative cases. **(c)** Acquisition of clonal *EBF1* mutations (boxed) at transformation shown for two examples. **(d)** Acquisition of NF-κB gene mutations (boxed) at transformation, particularly in *MYD88*, which had no mutations in the extension cohort of 100 follicular lymphomas, suggesting that these mutations are predominantly late genetic events during the course of disease.

pre-existing abnormality (Supplementary Fig. 11). Moreover, our longitudinal analyses of paired follicular lymphoma-transformed follicular lymphoma samples demonstrated that mutations in these genes were stable and remained clonally dominant, irrespective of therapy, in the majority of cases as the disease progressed (Fig. 5b and Supplementary Fig. 12), reaffirming that these mutations represent key events in the founding CPC that repopulates the tumor at transformation. Targeted resequencing in additional follicular lymphoma-transformed follicular lymphoma pairs confirmed the findings in our discovery cohort, demonstrating evolution predominantly occurring via rich CPCs (Supplementary Figs. 7 and 12).

We next evaluated the follicular lymphoma-transformed follicular lymphoma pairs for previously unreported transformation-associated events. Using copy number profiling, we found an increase in the number of regions affected by CNAs and, notably, an enrichment of particular CNAs at transformation, such as amplifications of *EZH2*, *MDM2*, *MYC* and *REL* (Supplementary Fig. 13 and Supplementary Table 10). *EBF1* mutations were acquired and were restricted to the transformation biopsies in 5 of 32 cases, with VAFs indicating clonal events (Fig. 5c). Remarkably, we also found the acquisition of alterations affecting regulators of NF-κB signaling at transformation, which specifically included mutations in *MYD88* and *TNFAIP3* (Fig. 5d and Supplementary Fig. 12). Four of the five *MYD88* mutations identified were restricted to the transformation biopsy, with none harboring the p.Leu265Pro alteration common to other B cell lymphomas^{30,37} (Supplementary Fig. 10). In all cases, these 'progressor' variants did not exist in the antecedent follicular lymphoma biopsies, suggesting that they were gained during the most recent clonal expansion leading to the transformation event. The acquisition of these late genetic events underscores the need for temporal mutational profiling as we move toward an era of precision medicine. It was particularly striking that the time interval from the biopsy of the antecedent follicular lymphoma to transformation in these cases was brief (range of 0.27–2.41 years),

raising the possibility that the acquisition of particular genetic events is selectively advantageous through promoting higher fitness for the clone that drives transformation.

In summary, we report the most comprehensive sequencing effort so far in follicular lymphoma. The mutational landscape of follicular lymphoma predominantly indicates epigenetic 'addiction' but highlights co-occurring aberrations in genes involved in B cell development as well as JAK-STAT and NF-κB signaling. By longitudinal profiling, we provide unequivocal evidence for a reservoir ancestral population, enriched in early mutations in driver genes, that propagates successive disease events. We did not identify a single compelling genetic event responsible for transformation but instead demonstrate that the acquisition of distinct genetic alterations may prompt the onset of aggressive disease. The high frequency of relapse suggests that CPCs are resistant to standard therapies and that adopting a stratified treatment approach targeting the specific early genetic lesions identified in the putative CPC may ultimately offer the best chance of eradicating these cells and curing follicular lymphoma.

URLs. CASAVA, http://support.illumina.com/sequencing/sequencing_software/casava.ilmn; PHYLIP package, <http://evolution.genetics.washington.edu/phylic.html>; Affymetrix Genome-Wide Human SNP Array 6.0, http://www.affymetrix.com/support/technical/byproduct.affx?product=genomewidesnp_6.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Sequencing data have been deposited in the European Genome-phenome Archive (EGA) under accession [EGAS00001000399](#). Affymetrix SNP 6.0 data have been deposited in the Gene Expression Omnibus (GEO) under accession [GSE42525](#).

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

ACKNOWLEDGMENTS

We are indebted to the patients for donating tumor specimens as part of this study. We would like to thank T. Chaplin-Perkins and B. Young for performing the SNP 6.0 array experiments as part of the Affymetrix Array Core Facility and G. Clark at the London Research Institute for automated DNA sequencing. This study was predominantly funded by Cancer Research UK through the Genomic Initiative and Programme grant (15968) to J.F. and was also supported by Leukemia and Lymphoma Research (grant to J.F.) and Hungarian Scientific Research Fund (Országos Tudományos Kutatási Alapprogramok, OTKA) grant K-76204 (to A.M.). Y.F. is a recipient of the Georgia Cancer Coalition Distinguished Scholar Award, and C.P. and Y.F. are, in part, supported by US National Institutes of Health grant GM085261 (to Y.F.). C.B. is a recipient of the European Hematology Association (EHA) Partner fellowship (2009/1) and was supported by the European Union and the State of Hungary, cofinanced by the European Social Fund in the framework of TÁMOP 4.2.4. A/1-11-1-2012-0001 National Excellence Program. J.O. is a recipient of the Kay Kendall Leukaemia Fund (KKLF) Junior Clinical Research Fellowship (KKL 557).

AUTHOR CONTRIBUTIONS

J.F., J.O. and C.B. designed and directed the study. J.O. and J.F. wrote the manuscript. T.A.L., S.M., J. Matthews and J.G.G. identified patients for the study and collected clinical information. S.I. prepared DNA samples. M.C., A.N., A.M. and R.J.B. conducted pathological review of specimens. J.W., D.C., J. Marzec, E.S., G.H.L., G.M. and C.C. performed the bioinformatic analysis. J.O., C.B., S.A., C.-Y.Y., S.B., C.P., M.B., D.W., K.T., N.P., C.O., D.O., E.C. and A.D. performed experiments. J.O., J.W., C.B., S.A., C.-Y.Y., S.B., G.L., Y.F. and R.G. analyzed the data. All authors read, critically reviewed and approved the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Swenson, W.T. *et al.* Improved survival of follicular lymphoma patients in the United States. *J. Clin. Oncol.* **23**, 5019–5026 (2005).
- Al-Tourah, A.J. *et al.* Population-based analysis of incidence and outcome of transformed non-Hodgkin's lymphoma. *J. Clin. Oncol.* **26**, 5165–5169 (2008).
- Montoto, S. & Fitzgibbon, J. Transformation of indolent B-cell lymphomas. *J. Clin. Oncol.* **29**, 1827–1834 (2011).
- Montoto, S. *et al.* Risk and clinical implications of transformation of follicular lymphoma to diffuse large B-cell lymphoma. *J. Clin. Oncol.* **25**, 2426–2433 (2007).
- Carlotti, E. *et al.* Transformation of follicular lymphoma to diffuse large B-cell lymphoma may occur by divergent evolution from a common progenitor cell or by direct evolution from the follicular lymphoma clone. *Blood* **113**, 3553–3557 (2009).
- Eide, M.B. *et al.* Genomic alterations reveal potential for higher grade transformation in follicular lymphoma and confirm parallel evolution of tumor cell clones. *Blood* **116**, 1489–1497 (2010).
- Ruminy, P. *et al.* Sp mutation patterns suggest different progression pathways in follicular lymphoma: early direct or late from FL progenitor cells. *Blood* **112**, 1951–1959 (2008).
- Weigert, O. *et al.* Molecular ontogeny of donor-derived follicular lymphomas occurring after hematopoietic cell transplantation. *Cancer Discov.* **2**, 47–55 (2012).
- Bödör, C. *et al.* EZH2 Y641 mutations in follicular lymphoma. *Leukemia* **25**, 726–729 (2011).
- Morin, R.D. *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat. Genet.* **42**, 181–185 (2010).
- Morin, R.D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).

- Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189–195 (2011).
- Lenz, G. *et al.* Oncogenic *CARD11* mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676–1679 (2008).
- Davis, R.E. *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88–92 (2010).
- Pasqualucci, L. *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat. Genet.* **43**, 830–837 (2011).
- Zhang, J. *et al.* Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. USA* **110**, 1398–1403 (2013).
- Bednar, J. *et al.* Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci. USA* **95**, 14173–14178 (1998).
- Croston, G.E., Kerrigan, L.A., Lira, L.M., Marshak, D.R. & Kadonaga, J.T. Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. *Science* **251**, 643–649 (1991).
- Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**, 1199–1212 (2005).
- Brown, D.T., Izard, T. & Misteli, T. Mapping the interaction surface of linker histone H1⁰ with the nucleosome of native chromatin *in vivo*. *Nat. Struct. Mol. Biol.* **13**, 250–255 (2006).
- Goytisolo, F.A. *et al.* Identification of two DNA-binding sites on the globular domain of histone H5. *EMBO J.* **15**, 3421–3429 (1996).
- Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. & Sweet, R.M. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* **362**, 219–223 (1993).
- Vyas, P. & Brown, D.T. N- and C-terminal domains determine differential nucleosomal binding geometry and affinity of linker histone isotypes H1⁰ and H1c. *J. Biol. Chem.* **287**, 11778–11787 (2012).
- Mottok, A. *et al.* Inactivating *SOCS1* mutations are caused by aberrant somatic hypermutation and restricted to a subset of B-cell lymphoma entities. *Blood* **114**, 4503–4506 (2009).
- Ritz, O. *et al.* Recurrent mutations of the STAT6 DNA binding domain in primary mediastinal B-cell lymphoma. *Blood* **114**, 1236–1242 (2009).
- Baus, D. *et al.* STAT6 and STAT1 are essential antagonistic regulators of cell survival in classical Hodgkin lymphoma cell line. *Leukemia* **23**, 1885–1893 (2009).
- Mottok, A., Renné, C., Willenbrock, K., Hansmann, M.L. & Bräuninger, A. Somatic hypermutation of *SOCS1* in lymphocyte-predominant Hodgkin lymphoma is accompanied by high JAK2 expression and activation of STAT6. *Blood* **110**, 3387–3390 (2007).
- Ritz, O. *et al.* STAT6 activity is regulated by SOCS-1 and modulates BCL-XL expression in primary mediastinal B-cell lymphoma. *Leukemia* **22**, 2106–2110 (2008).
- Compagno, M. *et al.* Mutations of multiple genes cause deregulation of NF- κ B in diffuse large B-cell lymphoma. *Nature* **459**, 717–721 (2009).
- Ngo, V.N. *et al.* Oncogenically active *MYD88* mutations in human lymphoma. *Nature* **470**, 115–119 (2011).
- Harvey, R.C. *et al.* Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood* **116**, 4874–4884 (2010).
- Hagman, J., Belanger, C., Travis, A., Turk, C.W. & Grosschedl, R. Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev.* **7**, 760–773 (1993).
- Treiber, N., Treiber, T., Zocher, G. & Grosschedl, R. Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. *Genes Dev.* **24**, 2270–2275 (2010).
- Treiber, T. *et al.* Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription-independent poising of chromatin. *Immunity* **32**, 714–725 (2010).
- Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
- Green, M.R. *et al.* Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood* **121**, 1604–1611 (2013).
- Treon, S.P. *et al.* MYD88 L265P somatic mutation in Waldenström's macroglobulinemia. *N. Engl. J. Med.* **367**, 826–833 (2012).
- Cao, K. *et al.* High-resolution mapping of H1 linker histone variants in embryonic stem cells. *PLoS Genet.* **9**, e1003417 (2013).
- Fan, Y. & Skoultschi, A.I. Genetic analysis of H1 linker histone subtypes and their functions in mice. *Methods Enzymol.* **377**, 85–107 (2004).

ONLINE METHODS

Case samples. Ten individuals with follicular lymphoma were selected on the basis of the availability of sequential tumor lymph node biopsies (a sample with follicular lymphoma and a subsequent sample with histological transformation to DLBCL (transformed follicular lymphoma)) and matched germline samples consisting of remission bone marrows or peripheral blood specimens. Clonality for each of the paired follicular lymphoma and transformed follicular lymphoma samples was confirmed by *BCL2-IGH* breakpoint analysis, as previously described⁴⁰. The clinical characteristics of these cases are presented in **Supplementary Table 1**. Overall, 34 samples (10 paired follicular lymphoma–transformed follicular lymphoma–germline samples and 2 additional relapsed follicular lymphoma samples for cases S2 and S7) were evaluated. An extension cohort of 22 additional paired follicular lymphoma and transformed follicular lymphoma samples and 100 independent follicular lymphoma cases were included in our validation experiments (**Supplementary Table 8**). All biopsies were previously histologically reviewed to confirm diagnosis and tumor content before DNA extraction from fresh-frozen tumor samples. Tumor content was further estimated using the ABSOLUTE³⁵ (version 1.0.4) and ASCAT⁴¹ (version 2.1) algorithms. Written consent was obtained for the collection and use of specimens for research purposes, with ethical approval obtained from the London Research Ethics Committee (LREC) of the East London and the City Health Authority (10/H0704/65 and 06/Q0605/69).

Whole-genome sequencing. Genomic DNA (5 µg) from matched tumor and germline samples was fragmented to generate libraries with an average insert size of 300 bp and sequenced using the HiSeq 2000 platform with paired-end reads of 100 bp, according to the manufacturer's protocols (Illumina). Detailed coverage data for all cases are shown in **Supplementary Table 2**.

Whole-genome sequencing read mapping and somatic mutation detection. Raw sequencing data were aligned to the reference human genome (hg19) using the Illumina ELANDv2 aligner. Alignments were sorted and converted to BAM format. PCR duplicates were removed, and clusters of poorly aligned and anomalous reads were *de novo* assembled into contigs using the CASAVA (v1.8) package (see URLs). We implemented the Strelka algorithm⁴² to identify somatic SNVs and indels for matched tumor–germline samples, selecting this algorithm for its sensitivity of mutation detection with varying tumor purities. After candidate indel detection and realignment, we called somatic variants and performed filtering. Only somatic calls originating from homozygous reference alleles in normal samples were considered. A joint somatic quality score (Q) of ≥15 for SNVs and ≥30 for indels was applied for all calls retained after filtering. The complete list of variants is detailed in **Supplementary Table 4**.

Whole-exome sequencing and analysis. Capture libraries were constructed from 3 µg of tumor and germline DNA using the Agilent SureSelect XT Human All Exon V4 kit. Enriched exome libraries were multiplexed and sequenced on the Illumina HiSeq 2500 platform to generate 100-bp paired-end reads (**Supplementary Table 2**). Sequencing reads were aligned to the reference genome using Burrows–Wheeler Aligner (BWA)⁴³. SAM to BAM conversion and marking of PCR duplicates were performed using Picard tools (version 1.86), followed by local realignment around indels and base quality score recalibration using the Genome Analysis Toolkit (GATK)⁴⁴ (v2.3.9). Somatic SNVs and indels were identified using the Strelka pipeline.

Annotation and recurrently mutated genes. All somatic SNVs and short indels retained after filtering were annotated using the SNPnexus tool⁴⁵. The Ensembl database (v63) was used for gene and transcript identification and the detection of amino acid changes. Any variants overlapping with common and germline polymorphisms present in the dbSNP132 and HapMap databases were masked. The allele frequencies for individual variants in both germline and tumor samples were determined by read count. Somatic variants identified with both whole-genome and whole-exome sequencing are included in **Supplementary Table 4**.

Copy number and structural variant analyses from whole-genome sequencing data. CNAs were identified using VarScan2 (ref. 46) (v2.3.5), with each

tumor sample compared to the matching germline sample. Samples were segmented using the circular binary segmentation (CBS) algorithm⁴⁷ implemented in the R Bioconductor package DNACopy (v2.12), which allowed splits to be reversed if the copy number ratio for tumor and normal samples was within 3 s.d. Structural variants were called using Pindel⁴⁸ (v0.2.4t), with the exception of translocations, which were called using default parameters. Only calls that had no supporting reads in the normal sample but had ≥8 supporting reads in the tumor sample were retained. Translocations (fusion transcripts) were identified using the DELLY⁴⁹ and Breakdancer⁵⁰ algorithms on paired-end reads, only retaining structural variants called by both methods with ≥5 supporting reads in the tumor sample.

Identification of recurrent mutations in promoter regions. Using Ensembl gene and transcript annotations, the somatic variants in the region from –2,000 to +250 bp (within the 5' UTR) relative to the transcription start site (TSS) for each transcript of protein-coding genes were identified across whole-genome sequencing data for all cases. Promoter regions recurrently mutated in multiple cases were identified, and mutation hotspots in these regions were also noted (**Supplementary Table 6**).

Mutation and indel validation. SNVs and indels were confirmed by bidirectional Sanger sequencing, fluorescence-based fragment size analysis or targeted deep sequencing (Fluidigm–MiSeq) as detailed below (**Supplementary Table 5**). Concordance data for selected variants called by whole-genome sequencing, whole-exome sequencing and validation targeted deep sequencing with estimated VAFs are shown in **Supplementary Figure 14**.

Phylogenetic analysis. Evolutionary trees were constructed for each of the ten cases analyzed from whole-genome or whole-exome sequencing data on the basis of the distance matrix between germline, follicular lymphoma and transformed follicular lymphoma samples derived from the numbers of somatic variants (all variants as well as nonsynonymous SNVs and indels) from each biopsy, using the neighbor-joining algorithm⁵¹ implemented in the PHYLIP package (see URLs). Once the consensus phylogenetic tree was determined, it was redrawn starting from germline and leading to the putative CPC, then progressing to the follicular lymphoma and transformed follicular lymphoma samples, with the branch length proportional to the number of somatic changes, i.e., genetic distance, between the samples. Trees were plotted against the clinical timelines.

High-density SNP arrays. Copy number analysis and identification of regions of copy-neutral loss of heterozygosity (cnLOH, also known as aUPD) were performed on a cohort of 26 paired follicular lymphoma–transformed follicular lymphoma and 3 diagnostic follicular lymphoma DNA samples (including 7 of our discovery cases) using Affymetrix SNP 6.0 microarrays. Further details are given in the **Supplementary Note**.

Targeted resequencing. Twenty-eight candidate genes (**Supplementary Table 8**) were sequenced in the extension cohort cases using the Access Array platform (Fluidigm). All samples were screened in duplicate, with a number of matched germline and normal tonsil DNA samples included as controls in each run. Access Array amplification was performed in a multiplex format with genomic DNA (50 ng) according to the manufacturer's recommendation. Deep sequencing was performed on multiplexed library pools using the Illumina MiSeq platform. After demultiplexing and FASTQ file generation for the raw data, reads were aligned to the reference genome using the Bowtie2 algorithm⁵². SAMtools⁵³ was used to generate sorted BAM files for each sample, and the VarScan2 tool was used to examine the pileup file to call variants, including SNPs and short indels, implementing a minimum threshold VAF of 5% and Fisher's exact test *P* value of <0.05. Only bases meeting the minimum base quality score of 20 from reads meeting the minimum mapping quality score of 20 were considered. A minimum read depth of 10 at a position was required to make calls, and variants with >90% of reads supporting one strand were excluded. Only concordant variants present in both duplicates or confirmed by independent Sanger sequencing were retained. Identified variants were annotated using SNPnexus to exclude variants reported in dbSNP132 and to identify variants that had nonsynonymous consequences or affected splice sites.

Estimating the clonality of mutations. Tumor purity, ploidy and absolute DNA copy numbers were determined from Affymetrix SNP 6.0 data ($n = 26$ follicular lymphoma–transformed follicular lymphoma pairs and 3 diagnostic follicular lymphomas) by implementing the ABSOLUTE algorithm (Supplementary Table 11). The ASCAT algorithm was also used to determine tumor purity and ploidy to ensure high concordance across both algorithms (Supplementary Table 11). These estimates were used to compute and classify individual somatic mutations obtained from our deep sequencing analyses as clonal or subclonal on the basis of the posterior probability that the cancer cell fraction (CCF) exceeded 0.95, as previously reported⁵⁴ (Supplementary Table 12).

Identification of recurrent targets of aberrant somatic hypermutation. SHM hotspots typically occur within the 2-kb window downstream of the TSS, and we focused our analyses on these regions⁵⁵. The total number of SNVs and the average mutation density across the 14 whole-genome sequenced tumor samples were investigated. For each gene, the ratio of transition to transversion mutations, the ratio of mutations at C:G to A:T sites and the percentage of SNVs within the hotspot motif WRCY, their associated *P* values and the SHM indicator were calculated in accordance with previous studies⁵⁶. The probability of observing the actual number of SNVs or more for each target gene was calculated on the basis of the Poisson distribution given the expected number of SNVs per SHM target region if those SNVs were uniformly randomly distributed in all target genes. A list of SHM gene targets is included in Supplementary Table 13.

EBF1 mutagenesis, retroviral expression and target gene analysis. Three follicular lymphoma–derived mutants (p.Gly171Asp and p.Ser238Thr alterations were located in the DBD, and p.Arg381Ser was located in the HLH domain) were generated to assess their ability to activate Ebf1 target gene expression. *Ebf1*-deficient pre-pro-B cell lines (cultured on OP9 feeders in the presence of IL-7, Flt3 ligand and stem cell factor (SCF)) were transduced with bicistronic retroviruses expressing wild-type or mutant Ebf1-GFP as described previously³⁴. GFP-positive cells were sorted 24 h after transduction, and expression was confirmed by immunoblotting (data not shown). Total RNA was extracted using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. qRT-PCR analysis was performed in triplicate on a 7500 Fast Real-Time PCR system (Applied Biosystems) using Power SYBR Green master mix to measure the mRNA levels of the Ebf1 target genes *Igll1* and *Cd79b* (refs. 33,34).

Generation of clones overexpressing wild-type and Ser102Phe histone H1c in histone H1c–histone H1d–histone H1e triple-knockout ESCs. Expression plasmids encoding Flag-tagged wild-type or Ser102Phe human histone H1c were constructed by replacing the mouse histone H1d coding sequence with the sequence encoding Flag-tagged wild-type or mutant human histone H1c in a 5-kb fragment encompassing the mouse *Hist1h1d* upstream and downstream regulatory sequences, which were inserted into a vector containing a blasticidin-resistant gene. Plasmid DNA (20 µg) was transfected into 2×10^7 histone H1 triple-knockout ESCs, as previously described⁵⁷. A total of 48 clones

(24 clones for transfection with each vector) resistant to blasticidin were picked and screened by immunoblotting using an anti-Flag antibody (Sigma-Aldrich, F1804; 1:2,000 dilution). Cell clones with the highest levels of expression of Flag-tagged wild-type and mutant histone H1c were selected for further analysis. (Supplementary Fig. 9)

Preparation and analysis of histones. Chromatin was prepared from ESCs, and histones were extracted with 0.2 N sulfuric acid, according to previously described protocols^{38,39,58}. Approximately 50 µg of total histones was injected into a C18 reverse-phase column (Vydac) on an AKTA UPC10 system (GE Healthcare). Effluent was monitored at 214 nm, and peak areas were analyzed with AKTA UNICORN (v5.11) software. Absorbance at 214 nm for the histone H1 and H2B peaks was adjusted according to the respective peptide bonds, and the histone H1/nucleosome ratio was calculated as previously described^{39,58}.

40. Ladetto, M. *et al.* A validated real-time quantitative PCR approach shows a correlation between tumor burden and successful *ex vivo* purging in follicular lymphoma patients. *Exp. Hematol.* **29**, 183–193 (2001).
41. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. USA* **107**, 16910–16915 (2010).
42. Saunders, C.T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
44. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
45. Dayem Ullah, A.Z., Lemoine, N.R. & Chelala, C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Res.* **40**, W65–W70 (2012).
46. Koboldt, D.C. *et al.* VarScan2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
47. Olshen, A.B., Venkatraman, E.S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
48. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
49. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
50. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* **6**, 677–681 (2009).
51. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
52. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
53. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Landau, D.A. & Wu, C.J. Chronic lymphocytic leukemia: molecular heterogeneity revealed by high-throughput genomics. *Genome Med.* **5**, 47 (2013).
55. Pasqualucci, L. *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346 (2001).
56. Khodabakhshi, A.H. *et al.* Recurrent targets of aberrant somatic hypermutation in lymphoma. *Oncotarget* **3**, 1308–1319 (2012).
57. Zhang, Y. *et al.* Histone H1 depletion impairs embryonic stem cell differentiation. *PLoS Genet.* **8**, e1002691 (2012).
58. Medrzycki, M., Zhang, Y., Cao, K. & Fan, Y. Expression analysis of mammalian linker-histone subtypes. *J. Vis. Exp.* **61**, e3577 (2012).