nature genetics

DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control

Helene Kretzmer[1–3,31], Stephan H Bernhart[1–3,31], Wei Wang[4,31], Andrea Haake[5,31], Marc A Weniger[6,31], Anke K Bergmann[5,7,31], Matthew J Betts[8], Enrique Carrillo-de-Santa-Pau[9], Gero Doose[1–3], Jana Gutwein[5], Julia Richter[5], Volker Hovestadt[4], Bingding Huang[10], Daniel Rico[9], Frank Jühling[1–3], Julia Kolarova[5], Qianhao Lu[8], Christian Otto[1–3], Rabea Wagener[5], Judith Arnolds[11], Birgit Burkhardt[12], Alexander Claviez[7], Hans G Drexler[13], Sonja Eberth[13,14], Roland Eils[10,15], Paul Flicek[16], Siegfried Haas[17], Michael Hummel[18], Dennis Karsch[19], Hinrik H D Kerstens[20], Wolfram Klapper[21], Markus Kreuz[22], Chris Lawerenz[10], Dido Lenze[18], Markus Loeffler[22], Cristina López[5], Roderick A F MacLeod[13], Joost H A Martens[20], Marta Kulis[20], José Ignacio Martín-Subero[23], Peter Möller[24], Inga Nagel[5], Simone Picelli[4], Inga Vater[5], Marius Rohde[25], Philip Rosenstiel[26], Maciej Rosolowski[22], Robert B Russell[8], Markus Schilhabel[26], Matthias Schlesner[10], Peter F Stadler[1–3,27–29], Monika Szczepanowski[20], Lorenz Trümper[14], Hendrik G Stunnenberg[20], ICGC MMML-Seq project[30], BLUEPRINT project[30], Ralf Küppers[6,32], Ole Ammerpohl[5,32], Peter Lichter[4,32], Reiner Siebert[5,32], Steve Hoffmann[1–3,32] & Bernhard Radlwimmer[4,32]

Although Burkitt lymphomas and follicular lymphomas both have features of germinal center B cells, they are biologically and clinically quite distinct. Here we performed whole-genome bisulfite, genome and transcriptome sequencing in 13 *IG-MYC* translocation–positive Burkitt lymphoma, nine *BCL2* translocation–positive follicular lymphoma and four normal germinal center B cell samples. Comparison of Burkitt and follicular lymphoma samples showed differential methylation of intragenic regions that strongly correlated with expression of associated genes, for example, genes active in germinal center dark-zone and light-zone B cells. Integrative pathway analyses of regions differentially methylated in Burkitt and follicular lymphomas implicated DNA methylation as cooperating with somatic mutation of sphingosine phosphate signaling, as well as the TCF3-ID3 and SWI/SNF complexes, in a large fraction of Burkitt lymphomas. Taken together, our results demonstrate a tight connection between somatic mutation, DNA methylation and transcriptional control in key B cell pathways deregulated differentially in Burkitt lymphoma and other germinal center B cell lymphomas.

Differentiation of B cells from uncommitted hematopoietic progenitors to antibody-producing plasma cells is a key process in the adaptive immune response. This differentiation is regulated through the coordinated expression of transcription factors along with a wave of epigenetic modulation in the germinal center[1–3], where genes encoding immunoglobulins are somatically altered to optimize antibody affinity and avidity. The B cell lymphomas most common in childhood (Burkitt lymphoma/leukemia) and adulthood (diffuse large B cell lymphoma (DLBCL) and follicular lymphoma) are presumed to derive from or resemble germinal center B cells. They all show molecular features of germinal center B cells, with the mutational mechanisms active in the germinal center, such as somatic hypermutation, contributing to pathogenesis. Whereas DLBCL is heterogeneous with regard to the genetic alterations and gene expression patterns of tumor cells, Burkitt lymphoma and, to a considerable extent, also follicular lymphoma are much more homogenous lymphoma subtypes[4,5]. Burkitt lymphoma is almost invariably associated with an *IG-MYC* chromosomal translocation, involving juxtaposition of the *MYC* oncogene encoding a transcription factor with one of the three immunoglobulin loci. The vast majority of follicular lymphomas, particularly those of grade 1 or 2, carry t(14;18)(q32;q21) chromosomal translocations juxtaposing the *BCL2* oncogene mostly with the *IGH* locus (encoding immunoglobulin heavy chain)[6]. These lymphoma subtypes also display a different spectrum of secondary changes, with, for example, mutations in cell cycle regulators and in the ID3-TCF3 transcription factor complex being highly prevalent in Burkitt lymphoma[7–9]. This deregulation of particular transcription factors in Burkitt lymphoma and the

A full list of author affiliations appears at the end of the paper.

Received 1 June; accepted 8 September; published online 5 October 2015; doi:10.1038/ng.3413

recurrent mutations in chromatin modifiers such as *CREBBP*, *MEF2B*, *EZH2* and *KMT2D* found in follicular lymphoma indicate that lymphoma pathogenesis involves alterations in physiological epigenetic modulation[10–14].

Clinically, patients affected with Burkitt lymphoma are significantly younger at diagnosis than those affected with follicular lymphoma. Moreover, Burkitt lymphoma shows a much higher proliferation rate and, thus, greater clinical aggressiveness than follicular lymphoma, which is a more indolent disease. Intriguingly, despite these considerable differences, both Burkitt lymphoma and follicular lymphoma are presumed to derive from the very same cell compartment of the human body, the germinal center B cell. Physiologically, normal germinal center B cells oscillate between the centroblast state predominant in the dark zone of a germinal center and the centrocyte state predominant in the light zone of the germinal center. Burkitt lymphoma and follicular lymphoma cells have been proposed to be 'frozen' in these functional states, with Burkitt lymphoma resembling the dark-zone germinal center B cell state and follicular lymphoma resembling the light-zone germinal center B cell state[15]. This offers the unique opportunity to study the epigenomic architecture of two neoplasms derived from the same cell of origin but showing considerably divergent genomic evolution.

RESULTS

Burkitt and follicular lymphomas show genome-wide hypomethylation

To better understand the relationship between epigenomic architecture, genetic alterations and phenotype in both germinal center B cell lymphoma subtypes, we performed whole-genome bisulfite sequencing (WGBS) and transcriptome sequencing of 13 *IG-MYC* translocation–positive Burkitt lymphoma samples and nine *BCL2* translocation–positive follicular lymphoma samples (including eight follicular lymphomas falling in grade 1 or 2 and one follicular lymphoma classified as grade 3a/DLBCL). As a reference, FACS-sorted germinal center B cells from the non-neoplastic tonsils of four donors were analyzed in parallel. The source of the sample material, as well as clinical follow-up for all patients, is summarized in **Supplementary Table 1**. In total, 25 billion read pairs, with at least 0.7 billion read pairs per sample, were obtained in the WGBS analysis. On average, 92% of reads successfully mapped to the reference genome[16], resulting in a mean coverage of 54× (**Supplementary Table 2**). The methylation rates called from the sequencing data showed excellent correlation with the β values obtained from Illumina HumanMethylation450K BeadChip analysis (**Supplementary Table 3**) for the same samples (mean $R^2 = 0.921$; **Supplementary Fig. 1**). For transcriptome sequencing, an average of 150 million reads were obtained per case, of which 94.5% mapped to the human genome (**Supplementary Table 4**). Our analyses were complemented by 30× whole-genome sequencing of tumors and matched controls from the same patients and the germinal center B cell populations, as well as eight additional Burkitt lymphomas (**Supplementary Table 5**).

Analysis of average genome-wide CpG methylation showed all lymphoma entities to be hypomethylated in comparison to non-neoplastic germinal center B cells (**Fig. 1a**). However, considerable differences could be seen in the extent and variability of DNA methylation between the Burkitt lymphoma and follicular lymphoma groups, despite their common germinal center B cell derivation (**Fig. 1b**). In addition to global hypomethylation, considerable gains in DNA methylation were observed in Burkitt lymphoma and follicular lymphoma samples at CpGs with low-level methylation in germinal center B cells (methylation <0.3; **Fig. 1c**). To determine whether DNA methylation patterns varied between genome segments with different functions, we compared methylation differences (averages of Burkitt lymphomas and follicular lymphomas versus germinal center B cells) for 15 chromatin states (**Supplementary Table 6**) defined in the GM12878 lymphoblastoid cell line, which is representative of non-neoplastic mature B cells[17]. For most chromatin states, individual segments showed either hyper- or hypomethylation largely neutralizing each other in the global analysis. However, changes in heterochromatin were mostly unidirectional, resulting in the greatest relative reduction in DNA methylation levels of any chromatin segment in the lymphoma samples. In stark contrast, poised promoters showed the strongest but a nevertheless small increase in DNA methylation levels at a global level (**Fig. 1d**). In line with recent reports[18], the observed small gains in poised promoter methylation were at times associated with increases in RNA expression of the associated genes in Burkitt lymphoma and follicular lymphoma samples (**Fig. 1e** and **Supplementary Fig. 2**).

DNA methylation downstream of promoters correlates with gene expression

Gene-centered analysis, including coding and noncoding genes, showed increased levels of gene body methylation for highly expressed genes, which returned to background levels at transcription end sites (TESs). High transcription levels typically were associated with the presence of large hypomethylated regions downstream of transcription start sites (TSSs), suggesting regulatory relevance for these regions (**Fig. 2a** and **Supplementary Fig. 3**). We applied an unbiased circular binary segmentation (CBS) approach to our entire WGBS data set, identifying 90,350 differentially methylated regions (DMRs) with ten or more CpGs between Burkitt lymphoma, follicular lymphoma and germinal center B cell controls. These DMRs included 27,607 and 36,775 DMRs between normal germinal center B cells and the follicular lymphoma and Burkitt lymphoma groups, respectively (**Supplementary Table 7**). The great majority of DMRs were located within transcribed regions (between TSSs and TESs) rather than in promoter regions (1,500-nt regions upstream of TSSs) or intergenic regions (**Fig. 2b**), with strong enrichment of DMRs immediately downstream of TSSs (**Fig. 2c**), analogous to the 'promoter downstream correlated regions' recently described in medulloblastoma[19]. No substantial differences in size were observed between intra- and intergenic DMRs (**Supplementary Fig. 4**).

To define the role of these potential regulatory regions in Burkitt lymphoma and follicular lymphoma, we correlated the methylation levels of intragenic DMRs (contained in the transcribed regions and up to 1,500 bp upstream of the TSS) with the expression levels of the associated genes containing these DMRs. Forty percent of the intragenic DMRs showed significant ($P < 0.05$) correlations between methylation and RNA expression (**Fig. 2d**), with the majority (64%) showing negative correlation (**Fig. 2e** and **Supplementary Table 8**). Overall, 85% of correlating DMRs (cDMRs) were located downstream of a TSS. Negatively and positively correlating DMRs did not differ in size; however, the former were highly enriched downstream of a TSS, whereas the latter were more evenly distributed across the gene body (**Supplementary Fig. 4**). Examples of negatively correlating DMRs are shown in **Supplementary Figure 5**. To ascertain the biological relevance of cDMR methylation, genes that previously had been reported as preferentially expressed in B cells of the dark zone or light zone of the germinal center[15] were examined; around 22% of dark zone– and 28% of light zone–specific genes possessed cDMRs. Dark zone–specific genes were predominantly hypomethylated and upregulated in Burkitt lymphoma, whereas they were hypermethylated

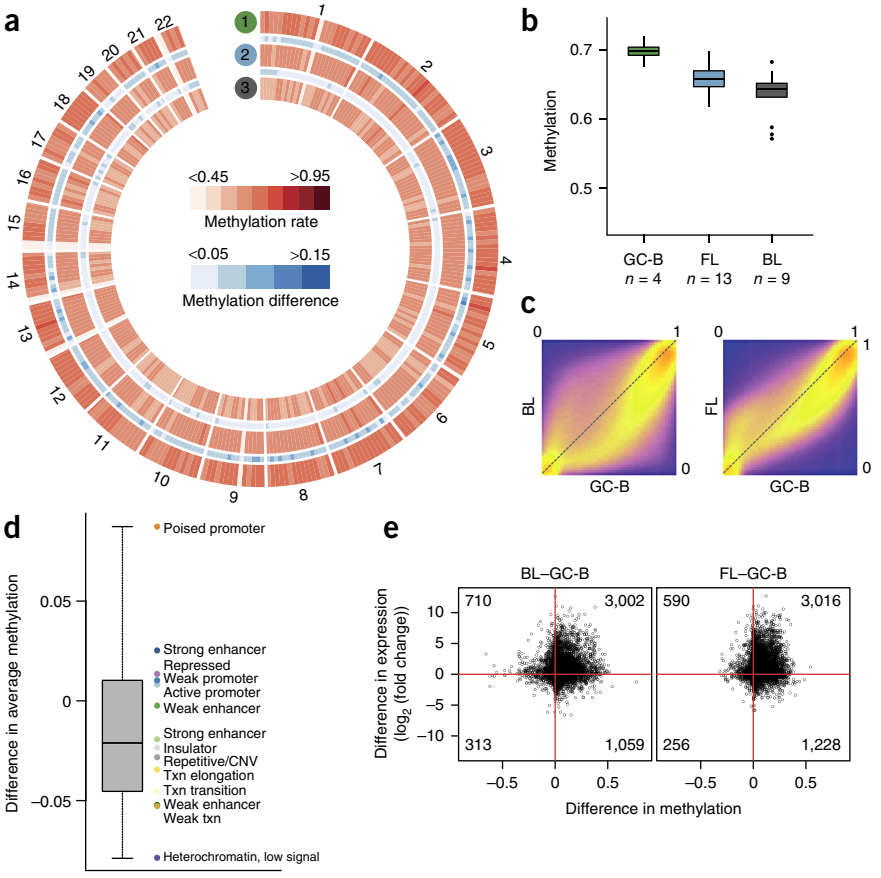

Figure 1 Loss of methylation in lymphoma. (**a**) Circular plot for genome-wide group average DNA methylation signals. A gradual loss of methylation in comparison to normal germinal center B cells is observed (1: germinal center B cell samples, $n = 4$; 2: follicular lymphomas, $n = 9$; 3: Burkitt lymphomas, $n = 13$). (**b**) Box plots of average methylation levels for the lymphomas and normal controls. Methylation differences are significant ($P < 0.05$) for all pairwise comparisons except for follicular lymphoma versus Burkitt lymphoma ($P = 0.13$). Boxes indicate medians and 25th and 75th percentiles; whiskers indicate the 1.5-fold interquartile ranges. GC-B, germinal center B cells; BL, Burkitt lymphoma; FL, follicular lymphoma. (**c**) Correlation of DNA methylation levels in Burkitt lymphoma and follicular lymphoma with those in germinal center B cells at CpG-level resolution (purple, low density; orange, high density). (**d**) Average methylation difference (average for Burkitt lymphoma and follicular lymphoma versus germinal center B cells) in GM12878 chromatin segments. Txn, transcription. (**e**) Scatterplots of DNA methylation of poised promoters versus RNA expression of the associated genes.

and downregulated in follicular lymphoma ($P < 0.001$, Fisher's exact test; **Fig. 2f,g**); light zone–specific genes showed the opposite pattern ($P < 0.001$, Fisher's exact test; **Fig. 2f,h**). These data are in line with the view that the majority of Burkitt lymphomas are frozen in the dark-zone cell state, whereas the majority of follicular lymphomas are frozen in the light-zone cell state. Moreover, they suggest that methylation at cDMRs could act to fix the otherwise physiologically transient functional states of dark-zone and light-zone B cells in Burkitt lymphoma and follicular lymphoma samples, respectively. Examples include high expression of *TCF3* and *SMARCA4*, which are more highly expressed in dark-zone than in light-zone B cells and in Burkitt lymphoma than in follicular lymphoma, as well as *NFKB2* and *TNFAIP3*, showing the opposite expression pattern (**Supplementary Fig. 6**).

Differential methylation is common in key pathways in Burkitt lymphoma

The observed association between light zone– and dark zone–specific gene expression prompted us to more fully evaluate cDMR involvement in differential gene expression between Burkitt lymphoma and follicular lymphoma. Ingenuity Pathway Analysis (IPA) was performed on the basis of (i) all genes differentially expressed in Burkitt lymphoma and follicular lymphoma and (ii) the subset of differentially expressed genes with negatively correlating DMRs. Genes with positive correlations of methylation and expression were used as a negative control. The first gene set identified 60 pathways showing significant ($P < 0.05$) enrichment of differentially expressed genes; these were assigned scores indicating pathway activation (**Supplementary Table 9**). Analysis of genes with negatively correlating DMRs identified 58 of the 60 pathways, all with comparable ratios of gene enrichment and activation scores. Pathway analysis based on positively correlating DMRs identified significantly fewer pathways (17/60; $P < 0.001$, Fisher's exact test). These data strongly indicate that modulation of DNA methylation levels at negatively correlating DMRs is a common mechanism in the regulation of differential gene expression in Burkitt lymphoma and follicular lymphoma. Next, we sought to define the molecular mechanisms deregulated by DNA methylation, particularly in Burkitt lymphoma using follicular lymphoma as a contrasting group. This was achieved by directly comparing Burkitt lymphoma with follicular lymphoma rather than with germinal center B cells, as the latter analysis would have been underpowered and potentially biased owing to the smaller number of germinal center B cell samples available ($n = 3$). We identified 23 pathways with IPA that were differentially activated in Burkitt lymphoma and follicular lymphoma (**Supplementary Table 10**). To determine the directions of the gene expression changes in these activated pathways relative to germinal center B cell controls, we visualized gene expression in Burkitt lymphoma, follicular lymphoma and germinal center B cells for all 23 pathways (**Supplementary Fig. 7**). Gene ontology analysis of genes that were specifically downregulated in Burkitt lymphoma or follicular lymphoma in the 23 pathways identified suppression of processes related to inflammation and immunity in Burkitt lymphoma and cell cycle and DNA repair in follicular lymphoma (**Supplementary Fig. 8**). Remarkably, one of the few genes that, relative to germinal center B cells, were upregulated in Burkitt lymphoma and downregulated in follicular lymphoma was *IGF2BP1* (**Supplementary Fig. 9**), encoding an RNA-binding protein that stabilizes *MYC*[20], underscoring its central importance in germinal center B cell lymphoma biology.

Detailed pathway-based analyses showed, for example, that DNA methylation was significantly associated with the downregulation of nuclear factor (NF)-κB signaling and the upregulation of cell cycle control genes in Burkitt lymphoma (**Supplementary Figs. 10** and **11**), concurrent with the proliferation rates of close to 100%

typical for Burkitt lymphoma (Ki67 staining; **Supplementary Table 1**) and consistent with previous reports on the molecular pathogenesis of Burkitt lymphoma[7–9,21,22]. The observed aberrant signaling was linked to differential methylation of key pathway genes in Burkitt lymphoma, including hypermethylation and downregulation of *JAK3* and *STAT3* (ref. 7) (**Fig. 2h–j**), in line with low JAK-STAT pathway activity in Burkitt lymphoma. Another example of differential methylation for key regulatory genes in Burkitt lymphoma was hypomethylation and overexpression of the transcription factor *TCF3* (**Fig. 2k–m**). Remarkably, the *TCF3* gene even harbored two cDMRs, of which cDMR1 contains a TCF3-binding site that is occupied in B cells[17], suggesting the possibility of a positive feedback loop amplifying *TCF3* transcription (**Fig. 2k,l**). Analysis of published TCF3 chromatin immunoprecipitation and sequencing (ChIP-seq) data from the Burkitt lymphoma cell lines BL-41 and NAMALWA[9] confirmed TCF3 binding within cDMR1 that overlapped a common TCF3 binding motif, and the methylation levels of the CpGs surrounding this motif were inversely correlated with *TCF3* expression in a larger set of 39 samples (**Supplementary Fig. 12**). These results indicate that, in addition to mutational mechanisms[8,9,22], *TCF3* cDMR hypomethylation contributes to TCF3 activation.

DNA mutation and methylation in sphingosine signaling

Pathway analysis also showed deregulation of sphingosine-1-phosphate (S1P) signaling in both Burkitt lymphoma and follicular lymphoma when compared to germinal center B cells (**Fig. 3a**). *S1PR2* (encoding sphingosine-1-phosphate receptor 2) is frequently affected by loss-of-function mutations in Burkitt lymphoma and DLBCL, leading to increased AKT and migratory activity[23]. In our analysis, we additionally identified cDMRs in several key pathway genes, including *PDGFRB*, which is strongly upregulated in both Burkitt lymphoma and follicular lymphoma relative to germinal center B cells (**Fig. 3b**); *GNA11* (encoding the stimulatory $G_q\alpha$ protein), upregulated in Burkitt lymphoma (**Fig. 3c**); and *S1PR1* and *GNA12* (encoding a member

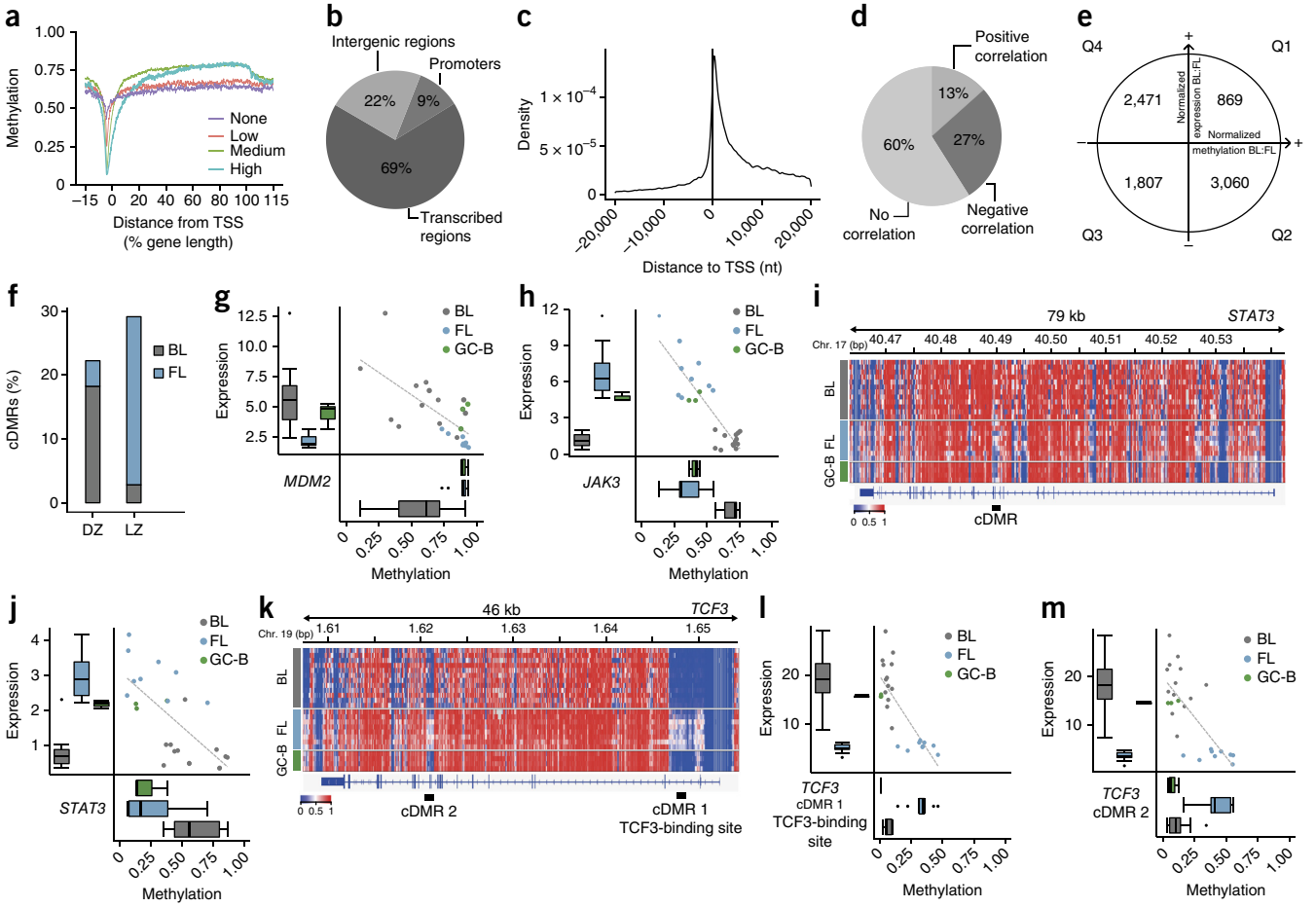

Figure 2 Differentially methylated regions. (**a**) Interdependence of DNA methylation and RNA expression levels across gene regions (the mean for germinal center B cells, follicular lymphomas and Burkitt lymphomas is shown). Genes were grouped by expression level (none, RPKM < 0.1; low, 0.1 ≤ RPKM < 1; medium, 1 ≤ RPKM < 10; high, 10 ≤ RPKM; where PRKM is reads per kilobase per million mapped reads). (**b**) Proportion of DMRs in promoters, transcribed regions and intergenic regions. (**c**) Density plot of DMRs with respect to distance from the TSS ($n = 22$ samples). (**d**) Proportion of DMRs in intragenic regions (1,500 nt upstream of the TSS to the TES) that showed no, positive or negative correlation with DMR methylation and expression of the associated gene. (**e**) Types and distribution of cDMRs. For about one-third of DMRs ($n = 8,207$; top), the degrees of methylation and RNA expression of associated genes correlated ($P < 0.05$, Spearman). Of these cDMR-gene pairs, 67% showed negative correlation (quadrants Q2 and Q4). (**f**) Proportion of germinal center dark-zone (DZ) and light-zone (LZ) genes with negatively correlating DMR–gene pairs in Burkitt lymphoma and in follicular lymphoma. (**g**,**h**) Scatterplots and box plots of cDMRs of the dark zone– and light zone–associated genes *MDM2* (**g**) and *JAK3* (**h**). (**i**) DNA methylation plot for *STAT3*, showing a cDMR in intron 6; red, methylated; blue, unmethylated. (**j**) Correlations of cDMR methylation and expression for *STAT3*. (**k**) DNA methylation plot for *TCF3*, showing two cDMRs within the gene body. cDMR1 in intron 2 contains a TCF3-binding site. (**l**,**m**) Correlation of *TCF3* cDMR 1 (**l**) and cDMR 2 (**m**) methylation with expression of *TCF3*.

of the $G_{12/13}\alpha$ protein complex), which are overexpressed in follicular lymphoma (**Fig. 3d**,**e**). Analysis of whole-genome sequencing data detected mutually exclusive mutations in several $G_{12/13}$ complex genes in 14 of the 21 Burkitt lymphoma samples analyzed (**Fig. 3f**). These findings are corroborated by our previous description of *RHOA* mutations in 8.5% of Burkitt lymphoma samples in an extended cohort of 82 cases[24] and by a recent study functionally implicating $G_{12/13}$ complex mutations in the development of germinal center B lymphomas[23]. Interestingly, three of the remaining seven samples were affected by somatic mutations in two genes of the inhibitory $G_i\alpha$ complex, *GNAI1* and *GNAI2*, in five Burkitt lymphoma samples, including recurrent ($n = 3$) *GNAI2* mutations encoding p.Arg179His (**Fig. 3f**). The $G_i\alpha$ complex can regulate several signaling pathways, including the AKT, extracellular signal–regulated kinase (ERK) and protein kinase A (PKA) pathways. Inhibition of PKA in Burkitt lymphoma cell lines did not affect cell viability (**Supplementary Fig. 13**), suggesting that this pathway might not be essential in Burkitt lymphoma; however, mutations of *GNAS* (encoding the $G_s\alpha$ protein) that affected the position homologous to GNAI2 Arg179 are known to cause upregulation of AKT signaling and cell migration in pancreatic adenocarcinoma and pituitary tumors[25,26]. Recently, it was proposed that, in addition to $G_{12/13}$ complex mutations, *GNAI2* mutations encoding a p.Arg179His substitution might increase the proliferation and migration of DLBCL cells[27]. Our observations on S1P and G protein signaling suggest that the mutation and differential DNA methylation of genes involved in these pathways might complement each other in suppressing $G_{12/13}$ and activating AKT signaling in Burkitt lymphoma.

Differential DNA methylation of transcription factor binding sites

Because cDMRs could be used to identify pathways contributing to the pathogenesis of Burkitt lymphoma through genomic mutation or deregulation by DNA methylation, we sought to identify sets of transcription factors driving the expression of genes with cDMRs. We intersected the cDMR data set with transcription factor binding site data for 46 transcription factors[17,28]. Significant enrichment ($P < 0.05$) of transcription factor binding sites was observed for genes that were expressed at lower levels in Burkitt lymphoma than in follicular lymphoma samples and were associated with hypermethylated (Burkitt lymphoma/follicular lymphoma) cDMRs in quadrant 2 of the radar plot (**Fig. 4a**). We selected the ten transcription factors displaying the highest correlation of transcription factor and target gene expression (**Supplementary Table 11**) for subsequent analyses. Differential methylation of the corresponding transcription factor binding site typically was restricted to the binding site itself and rapidly returned to background levels outside of this site (**Supplementary Fig. 14**).

To understand how cDMR methylation modulates transcription factor target gene expression, we plotted the average methylation levels of cDMRs against average RNA expression of the associated genes for all ten transcription factors (**Fig. 4b**). The target genes of STAT5A and BCL3, both enriched in quadrant 2 (**Fig. 4a**), displayed hypermethylation and reduced RNA expression in Burkitt lymphoma samples when compared to germinal center B cell and follicular lymphoma samples, in agreement with the low-level activity of the nuclear factor (NF)-κB and JAK-STAT pathways previously reported in Burkitt lymphoma[7]. Similar patterns were observed for the target genes of other transcription factors in quadrant 2. In quadrant 4, we observed hypomethylation and activation of TCF3 target genes in Burkitt lymphoma when compared to follicular lymphoma samples. Germinal center B cell samples displayed intermediate methylation levels, more similar to follicular lymphoma methylation levels than those in Burkitt lymphoma. As previously noted, mutations affecting *TCF3* or its negative regulator *ID3* detected by us and others[8,22] in more than two-thirds of all sporadic Burkitt lymphomas have been proposed to foster TCF3-dependent gene expression, driving Burkitt lymphoma cell proliferation[9]. Notably, neither transcription factor nor target gene expression was correlated with cell proliferation differences between Burkitt lymphoma and follicular lymphoma samples (**Supplementary Fig. 15**). Taken together, these findings suggest that both Burkitt lymphoma– and follicular lymphoma–specific gene expression programs are subject to regulation through DNA methylation of transcription factor binding sites.

To assess the impact of the transcription factors on target gene expression at a single-gene level, we generated activity plots integrating and visualizing the correlation of target gene expression with cDMR methylation and the level of transcription factor expression (**Fig. 4c**). Individual genes displaying the strongest activation in follicular lymphoma as compared to Burkitt lymphoma included the hallmark gene of follicular lymphomagenesis, *BCL2*, as well as *CBX7*, encoding a chromobox protein of Polycomb repressive complex 1 (PRC1) that was previously shown to be functionally involved in lymphomagenesis, with high expression in follicular lymphoma[29]. Notably, strong activation in follicular lymphoma was also observed for a set of genes typically expressed in T cells, including *CD3E*, *CD2*, *GIMAP1* and *ITK*, in line with the differential role of bystander T lymphocytes[7] in follicular in contrast to Burkitt lymphomagenesis. *TCF3* was among the genes strongly activated in Burkitt lymphoma, in line with the already proposed autoactivation loop regulating its expression (**Fig. 4c**,**d**). Other strongly activated genes included the E2F family member *E2F1*, the *POLD1* gene, which encodes the MYC-interacting DNA polymerase δ_1, the *TERT* gene, which encodes telomerase, and the *SMARCA4* gene (**Fig. 4c**,**e**) encoding a member of the SWI/SNF nucleosome-remodeling complex. Consistent with previous findings[21], we could independently confirm overexpression of *SMARCA4* RNA in Burkitt lymphoma using array-based analysis (**Supplementary Fig. 16** and **Supplementary Table 12**).

DNA mutation and methylation affect *SMARCA4* in Burkitt lymphoma

Unexpectedly, the observed strong *SMARCA4* transcriptional activation did not result in the upregulation of SMARCA4 target genes in Burkitt lymphoma, despite the fact that SMARCA4 binding sites were hypomethylated (**Fig. 4a**,**f**). These data suggest that suppression of SMARCA4-dependent expression does not occur through hypermethylation of *SMARCA4* binding sites in target genes but rather through deregulation by DNA methylation–independent mechanisms in Burkitt lymphoma.

To elucidate the role of *SMARCA4* in Burkitt lymphomagenesis, we performed an integrative data analysis (**Fig. 5a**) that showed high expression of *SMARCA4* in Burkitt lymphoma associated with both decreased DNA methylation and chromatin reprogramming to active promoters (chr. 19: 11.07–11.09 Mb) and enhancers to promoters (chr. 19: 11.15–11.16 Mb). Immunohistochemical staining found strong SMARCA4 protein (BRG1) expression in normal germinal center B cells, but much less pronounced signal was seen in germinal center mantle and other B cell compartments of the tonsils (**Fig. 5b** and **Supplementary Table 13**); in contrast, Burkitt lymphoma samples showed ubiquitous, high SMARCA4 protein expression (**Fig. 5c** and **Supplementary Table 13**).

Whole-genome sequencing of the International Cancer Genome Consortium (ICGC) MMML-Seq cohort identified *SMARCA4* mutations in nine of 21 Burkitt lymphoma samples. In line with previous

reports[30,31], these *SMARCA4* mutations clustered in the portion of the gene encoding the helicase domain (**Fig. 5d**), showing a mutation pattern that stands in strong contrast to the pattern of germline nonsense mutations predisposing to cancer syndromes[32–35]. Further investigation of the *SMARCA4* mutational landscape in bona fide Burkitt lymphoma samples, including data from cell lines and published data, confirmed these observations (**Supplementary Table 14**). In the Burkitt lymphoma samples, mutant and wild-type *SMARCA4* transcripts had similar expression levels (data not shown), and immunohistochemistry and immunofluorescence analyses of Burkitt lymphoma samples showed constant coexpression of SMARCA4 and MYC proteins in Burkitt lymphoma, regardless of the presence or absence of *SMARCA4* mutations (**Fig. 5c**, **Supplementary Fig. 17** and **Supplementary Table 13**).

Intriguingly, genomic mutations of other members of the SWI/SNF complex, including *ARID1A*, *ARID1B* and *SMARCB1*, were present

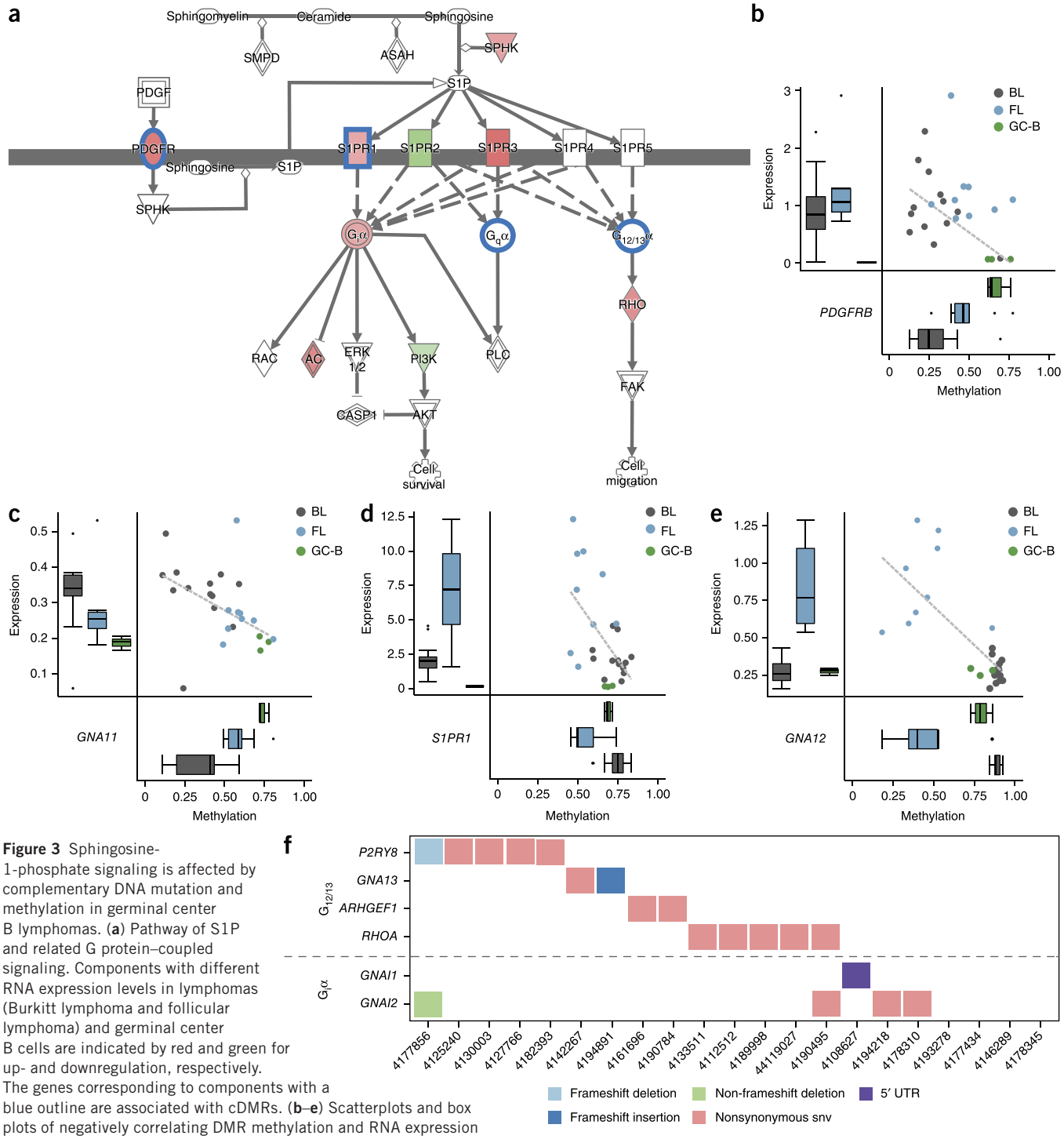

Figure 3 Sphingosine-1-phosphate signaling is affected by complementary DNA mutation and methylation in germinal center B lymphomas. (**a**) Pathway of S1P and related G protein–coupled signaling. Components with different RNA expression levels in lymphomas (Burkitt lymphoma and follicular lymphoma) and germinal center B cells are indicated by red and green for up- and downregulation, respectively. The genes corresponding to components with a blue outline are associated with cDMRs. (**b**–**e**) Scatterplots and box plots of negatively correlating DMR methylation and RNA expression affecting genes involved in S1P and G protein signaling: *PDGFRB* (**b**), *GNA11* (**c**), *S1PR1* (**d**) and *GNA12* (**e**). (**f**) Mutations identified in Burkitt lymphoma samples affecting the $G_{12/13}$ (above the dashed line) and $G_i\alpha$ (below the dashed line) complexes. SNV, single-nucleotide variant.

in a subset of Burkitt lymphoma samples lacking *SMARCA4* mutation (**Fig. 5e** and **Supplementary Table 15**), and, moreover, SWI/SNF complex genes showed increased expression levels in Burkitt lymphoma samples (**Fig. 5f**). Protein modeling[36] (**Fig. 5g**) suggested that *SMARCA4* mutations likely do not obstruct binding of the mutant protein to DNA, but some mutations might ablate helicase function directly, through interference with ATP binding, or indirectly, by obstructing the interaction of the helicase N- and C-terminal domains. It has recently been suggested that oncogenic *SMARCA4* mutations such as those observed here in Burkitt lymphoma compromise TOP2A chromatin binding, which is dependent on the ATPase activity of SMARCA4.

DNA methylation in lymphomas versus normal B cell development

Finally, we aimed to compare the DNA methylation patterns in Burkitt lymphoma and follicular lymphoma with the DNA methylome changes observed during human B cell differentiation recently described by Kulis *et al.*[37]. Comparison of the DNA methylation

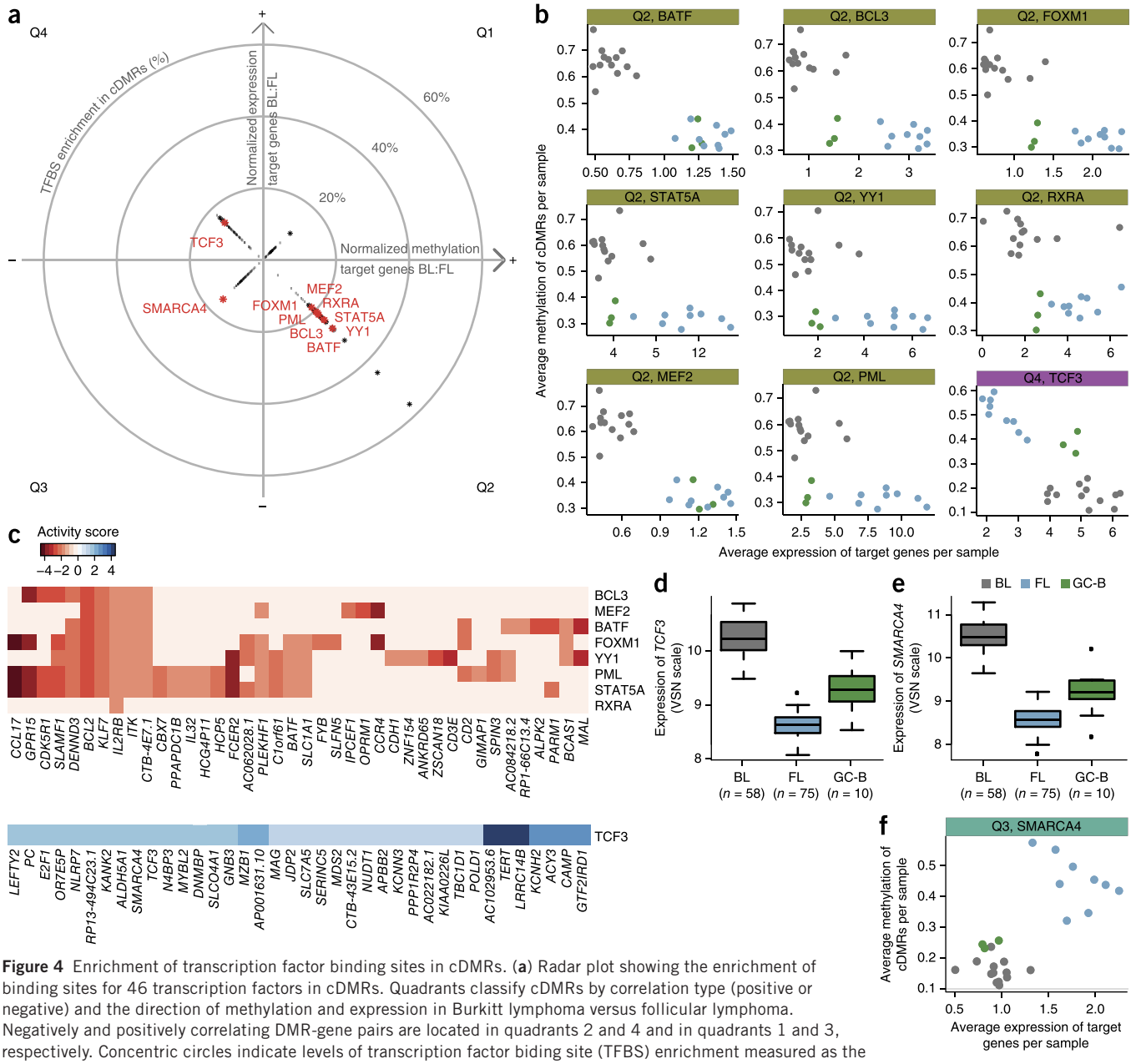

Figure 4 Enrichment of transcription factor binding sites in cDMRs. (**a**) Radar plot showing the enrichment of binding sites for 46 transcription factors in cDMRs. Quadrants classify cDMRs by correlation type (positive or negative) and the direction of methylation and expression in Burkitt lymphoma versus follicular lymphoma. Negatively and positively correlating DMR-gene pairs are located in quadrants 2 and 4 and in quadrants 1 and 3, respectively. Concentric circles indicate levels of transcription factor biding site (TFBS) enrichment measured as the percentage of binding sites of a particular transcription factor found in cDMRs relative to all binding sites of this transcription factor found in DMRs. Asterisks indicate transcription factor binding sites that are significantly enriched ($P < 0.05$, permutation test) in cDMRs. Red asterisks indicate the ten transcription factors showing the best correlation of transcription factor and average target gene expression. (**b**) Correlations of sample-wise average cDMR methylation and average target gene expression for the nine of the top ten transcription factors showing negative correlations. (**c**) Activity map integrating the correlation of transcription factor binding site methylation and differential expression of target genes. Color and intensity indicate the strength of inactivation (red) or activation (blue) in Burkitt lymphoma as compared to follicular lymphoma samples. The top 40 target genes are shown for both quadrants 2 and 4. (**d**,**e**) RNA expression of *TCF3* (**d**) and *SMARCA4* (**e**) in Burkitt lymphoma, follicular lymphoma and germinal center B cells as determined by microarray analysis in an extended lymphoma cohort. VSN, variance-stabilizing normalization. (**f**) Correlation of sample-wise average cDMR methylation and average target gene expression for *SMARCA4*.

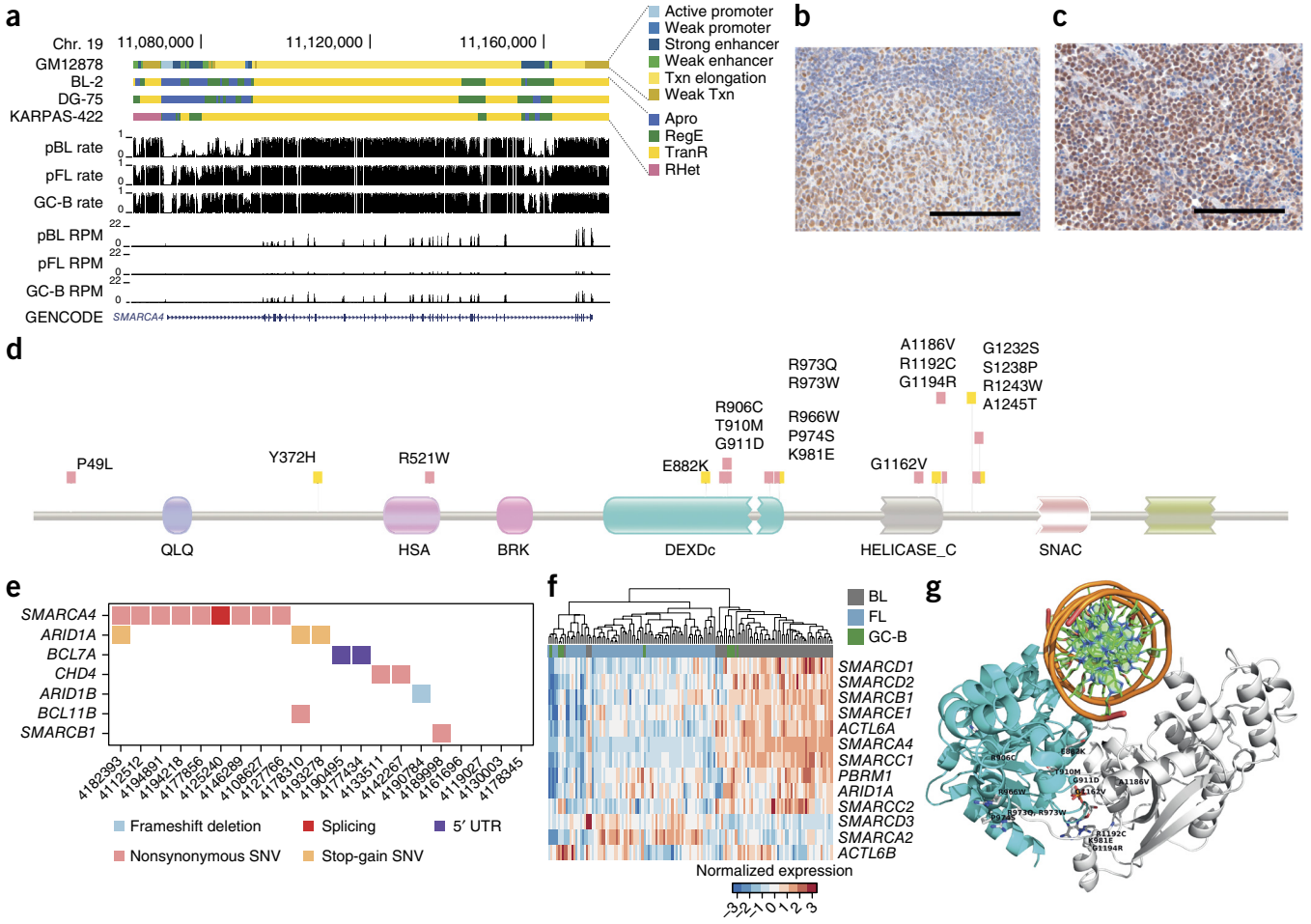

Figure 5 SMARCA4 genome architecture and protein expression. (**a**) Genome browser view of the *SMARCA4* locus. Chromatin segmentations in GM12878, BL-2, DG-75 and KARPAS-422 cell lines (top) are shown along with average rate of CpG methylation (middle) and RNA expression (bottom) in Burkitt lymphoma, follicular lymphoma and germinal center B cell samples. pBL, prototypical Burkitt lymphoma; pFL, prototypical follicular lymphoma; RPM, reads per million. (**b**,**c**) Staining for SMARCA4 (BRG1) by immunohistochemistry in a normal reactive follicle (tonsil) (**b**) and a Burkitt lymphoma with *SMARCA4* mutation encoding p.Arg973Trp (infiltrated lymph node) (**c**). Original magnification, 400×; scale bars, 200 µm. (**d**) Accumulation of missense mutations in *SMARCA4* affecting the DEXDc helicase (cyan) and helicase conserved C-terminal (gray) domains in Burkitt lymphoma samples. The heights of the lollipops indicate the relative number of alterations at each position. Red squares indicate severe changes in physicochemical properties (QLQ score < 0); yellow squares indicate 0 < QLQ score < 4. HAS, helicase/SANT-associated domain; BRK, BRK domain; SNAC, Snf2-ATP coupling chromatin remodeling complex. (**e**) The distribution of somatic mutations (SNVs and indels) in the SWI/SNF complex in Burkitt lymphoma suggests mutually exclusive mechanisms of inactivation. (**f**) Unsupervised analysis of SWI/SNF complex RNA expression data (Affymetrix U133A) segregates Burkitt lymphoma and follicular lymphoma. (**g**) Model of SMARCA4 interaction with a DNA helix. The amino acid residues mutated in Burkitt lymphoma are indicated, but the model suggests that they ablate helicase function rather than DNA binding by interfering with ATP binding, either directly or by obstructing the interaction of the helicase N- and C-terminal domains. Cyan, DEXDc domain; gray, C-terminal helicase domain. ATP is shown with sticks colored by atom type.

levels determined by WGBS in sorted germinal center B cells in both studies showed excellent agreement (correlation of means = 0.9799; **Supplementary Fig. 18a**). Principal-component analysis of the lymphomas analyzed herein together with the recently investigated normal B cell populations showed that the Burkitt lymphoma and follicular lymphoma subsets clustered together but distinctly from the normal B cell subsets (**Supplementary Fig. 18b**). This clustering of the follicular lymphomas and Burkitt lymphomas separately from the physiological germinal center B cells most likely relates to the neoplastic process. To further investigate this, we analyzed the differences in DNA methylation separately for Burkitt lymphoma and follicular lymphoma in comparison to normal germinal center B cells for the 20 major modules of dynamic CpG methylation during B cell development defined by Kulis *et al.* using both WGBS and array-based data (**Supplementary Fig. 18c**,**d**). In line with what has been described for DLBCL by Kulis *et al.*, we observed strongly increased DNA methylation of CpGs related to Polycomb-repressed regions (modules 17–20) and decreased DNA methylation of CpGs related to heterochromatin (modules 8 and 9) in both follicular lymphoma and Burkitt lymphoma as compared to germinal center B cells. This analysis extends the previous findings to reflect the fact that Burkitt lymphomas and follicular lymphomas also frequently acquire methylation changes in regions already undergoing dynamic methylation during normal B cell differentiation. Remarkably, when directly comparing the DNA methylation of Burkitt lymphomas and follicular lymphomas for the 20 modules, we hardly observed any differences. This suggests that the DNA methylation differences between these two germinal center B-lymphoma subtypes indeed reflect the separate oncogenic paths driving these lymphomas rather than the DNA methylation dynamics of normal B cell differentiation. The only notable exception was

module 11, which tended to have lower DNA methylation in Burkitt lymphomas than in follicular lymphomas. Remarkably, the CpGs of this module, which are strongly enriched for enhancer sites, are characterized by the fact that they switch their DNA methylation state from the naive B cell to the germinal center B cell stage[37]. This switching, as well as the enrichment for TCF3-binding sites in this module, might to a certain extent relate to the above-described freezing of the functional states of germinal center B cells in Burkitt lymphoma and follicular lymphoma, as well as to the differential activation of oncogenic pathways in these lymphomas (**Supplementary Fig. 18e**).

DISCUSSION

The presented DNA methylation analyses of Burkitt lymphoma and follicular lymphoma, both derived from germinal center B cells, identified substantial differences in DNA methylation patterns. By combination with genome and transcriptome analyses, these differences in DNA methylation can be linked to transcription factor activity, and they are critical determinants of follicular lymphoma– and Burkitt lymphoma–specific gene expression patterns. In addition, lymphoma subtype-specific patterns of DNA methylation, integrated with mutation analysis, helped us identify complementary aberrant regulation of the SWI/SNF, TCF3-ID3 and $G_i\alpha$ complexes as important contributors to Burkitt lymphomagenesis. Through our integrated analysis, we now add a model for a potential dominant function of mutated SMARCA4 in Burkitt lymphomagenesis (**Supplementary Fig. 19**). In this model, overexpression of TCF3, initiated by the combination of cDMR hypomethylation and mutation of *TCF3* or its inhibitor *ID3*, leads to high expression of the inactive SMARCA4 protein, which prevents methylation of its binding site and expression of its target genes as a result of maintained binding competence in the absence of helicase activity.

URLs. ChIP-seq library preparation protocols, http://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58; R software, http://www.R-project.org/; segmentation of methylation data, http://www.bioinf.uni-leipzig.de/Software/metilene/; source of transcription factor binding site data, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibTfbs; SAMtools, http://www.htslib.org/; model of chromatin states, http://ftp.ebi.ac.uk/pub/databases/blueprint/paper_data_sets/monocyte_neutrophil_2014/chromatin_states/full_histone_panel/model_11_Blueprint_11.txt.

METHODS

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

Accession codes. WGBS data, mRNA sequencing data and whole-genome DNA sequencing data: all sequencing data are available at the European Genome-phenome Archive (EGA) under accession EGAS00001001067. Please contact the data access committee (DAC; http://www.icgc.org/daco) of the International Cancer Genome Consortium (ICGC; http://www.icgc.org/) for access to the data. ChIP-seq data: the raw sequencing read data are publically available from the European Nucleotide Archive (ENA); all ChIP cell line data are available under study accession ERP002586. The run identifiers are listed in **Supplementary Table 16**. A complete list of the raw files available from the ftp site is provided together with associated metadata in the data index file (indexed with secondary sample ERS accession codes; ftp://ftp.ebi.ac.uk/pub/databases/blueprint/releases/20140811/homo_sapiens/20140811.data.index).

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

ACKNOWLEDGMENTS

This study has been supported by the German Ministry of Science and Education (BMBF) in the framework of the ICGC MMML-Seq project (01KU1002A-J) and the MMML-MYC-SYS project (036166B), the European Union in the framework of the BLUEPRINT Project (HEALTH-F5-2011-282510) and the KinderKrebsInitiative Buchholz/Holm-Seppensen and LIFE (Leipzig Research Center for Civilization Diseases), Leipzig University. LIFE is funded by the European Union, the European Regional Development Fund (ERDF), the European Social Fund (ESF) and the Free State of Saxony. WGBS was additionally supported by NGFNplus (BMBF, 01GS0883) and the DKFZ–Heidelberg Center for Personalized Oncology (DKFZ-HIPO). Former grant support of MMML by the Deutsche Krebshilfe (2003–2011) is gratefully acknowledged. J.R. is supported by the Dr. Werner Jackstädt Foundation in the framework of a Junior Excellence Research Group on 'Mechanisms of B-Cell Lymphomagenesis in the Senium as Basic Principle for the Development of Age-Adjusted Therapy Regimes' (S134-10.100).

AUTHOR CONTRIBUTIONS

A.H., O.A., R.K., P.L., R.S., S. Hoffmann and B.R. conceived and designed the experiments. W.W., A.H., A.K.B., J.G., J.R., J.K., R.W., S.E., H.H.D.K., W.K., D.L., C. López, S.P., I.V., P.R., M. Schilhabel, M. Szczepanowski, L.T. and R.K. performed the experiments. H.K., S.H.B., M.J.B., B.H., R.E., P.F., C. Lawerenz, J.H.A.M., M. Schlesner, P.F.S., H.G.S. and S. Hoffmann performed statistical analysis. H.K., S.H.B., G.D., V.H., D.R., F.J., C.O., M.H., M. Kreuz, M. Kulis, I.N., M. Rosolowski, R.B.R., M. Schlesner, S. Hoffmann and B.R. analyzed the data. H.K., M.A.W., M.J.B., E.C.-d.-S.-P., G.D., B.H., F.J., Q.L., C.O., J.A., B.B., A.C., H.G.D., S.E., R.E., P.F., S. Haas, D.K., H.H.D.K., W.K., M. Kreuz, C. Lawerenz, D.L., M.L., R.A.F.M., J.H.A.M., J.I.M.-S., P.M., M. Rohde, P.R., M. Schilhabel, M. Schlesner, L.T., H.G.S. and S. Hoffmann contributed reagents, materials and/or analysis tools. H.K., S.H.B., W.W., A.H., P.L., R.S., S. Hoffmann and B.R. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Lai, A.Y. *et al.* DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at Alu elements during B-cell activation. *Genome Res.* **23**, 2030–2041 (2013).
2. Lee, S.T. *et al.* A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res.* **40**, 11339–11351 (2012).
3. Shaknovich, R. *et al.* DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. *Blood* **118**, 3559–3569 (2011).
4. Basso, K. & Dalla-Favera, R. Germinal centres and B cell lymphomagenesis. *Nat. Rev. Immunol.* **15**, 172–184 (2015).
5. Lenz, G. & Staudt, L.M. Aggressive lymphomas. *N. Engl. J. Med.* **362**, 1417–1429 (2010).
6. Küppers, R. & Dalla-Favera, R. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* **20**, 5580–5594 (2001).
7. Dave, S.S. *et al.* Molecular diagnosis of Burkitt's lymphoma. *N. Engl. J. Med.* **354**, 2431–2442 (2006).
8. Richter, J. *et al.* Recurrent mutation of the *ID3* gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat. Genet.* **44**, 1316–1320 (2012).
9. Schmitz, R. *et al.* Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* **490**, 116–120 (2012).
10. Loeffler, M. *et al.* Genomic and epigenomic co-evolution in follicular lymphomas. *Leukemia* **29**, 456–463 (2015).
11. 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).
12. Morin, R.D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
13. Okosun, J. *et al.* Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat. Genet.* **46**, 176–181 (2014).
14. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189–195 (2011).
15. Victora, G.D. *et al.* Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. *Blood* **120**, 2240–2248 (2012).
16. Otto, C., Stadler, P.F. & Hoffmann, S. Fast and sensitive mapping of bisulfite-treated sequencing data. *Bioinformatics* **28**, 1698–1704 (2012).

17. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
18. Hahn, M.A. *et al.* Loss of the Polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors. *Cancer Res.* **74**, 3617–3629 (2014).
19. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
20. Weidensdorfer, D. *et al.* Control of c-*myc* mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA* **15**, 104–115 (2009).
21. Hummel, M. *et al.* A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. *N. Engl. J. Med.* **354**, 2419–2430 (2006).
22. Love, C. *et al.* The genetic landscape of mutations in Burkitt lymphoma. *Nat. Genet.* **44**, 1321–1325 (2012).
23. Muppidi, J.R. *et al.* Loss of signalling via Galpha13 in germinal centre B-cell–derived lymphoma. *Nature* **516**, 254–258 (2014).
24. Rohde, M. *et al.* Recurrent *RHOA* mutations in pediatric Burkitt lymphoma treated according to the NHL-BFM protocols. *Genes Chromosom. Cancer* **53**, 911–916 (2014).
25. O'Hayre, M. *et al.* The emerging mutational landscape of G proteins and G-protein–coupled receptors in cancer. *Nat. Rev. Cancer* **13**, 412–424 (2013).
26. Takuwa, N. *et al.* Tumor-suppressive sphingosine-1-phosphate receptor-2 counteracting tumor-promoting sphingosine-1-phosphate receptor-1 and sphingosine kinase 1—Jekyll hidden behind Hyde. *Am. J. Cancer Res.* **1**, 460–481 (2011).
27. Morin, R.D. *et al.* Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* **122**, 1256–1265 (2013).
28. Abraham, B.J., Cui, K., Tang, Q. & Zhao, K. Dynamic regulation of epigenomic landscapes during hematopoiesis. *BMC Genomics* **14**, 193 (2013).
29. Scott, C.L. *et al.* Role of the chromobox protein CBX7 in lymphomagenesis. *Proc. Natl. Acad. Sci. USA* **104**, 5389–5394 (2007).
30. Dykhuizen, E.C. *et al.* BAF complexes facilitate decatenation of DNA by topoisomerase IIα. *Nature* **497**, 624–627 (2013).
31. Klapper, W. *et al.* Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma. *Blood* **119**, 1882–1887 (2012).
32. Hasselblatt, M. *et al.* Nonsense mutation and inactivation of SMARCA4 (BRG1) in an atypical teratoid/rhabdoid tumor showing retained SMARCB1 (INI1) expression. *Am. J. Surg. Pathol.* **35**, 933–935 (2011).
33. Hasselblatt, M. *et al. SMARCA4*-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. *Acta Neuropathol.* **128**, 453–456 (2014).
34. Schneppenheim, R. *et al.* Germline nonsense mutation and somatic inactivation of *SMARCA4*/*BRG1* in a family with rhabdoid tumor predisposition syndrome. *Am. J. Hum. Genet.* **86**, 279–284 (2010).
35. Witkowski, L. *et al.* Germline and somatic *SMARCA4* mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat. Genet.* **46**, 438–443 (2014).
36. Betts, M.J. *et al.* Mechismo: predicting the mechanistic impact of mutations and modifications on molecular interactions. *Nucleic Acids Res.* **43**, e10 (2015).
37. Kulis, M. *et al.* Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat. Genet.* **47**, 746–756 (2015).

[1]Transcriptome Bioinformatics, Research Center for Civilization Diseases (LIFE), University of Leipzig, Leipzig, Germany. [2]Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany. [3]Bioinformatics Group, Department of Computer Science, University of Leipzig, Leipzig, Germany. [4]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [5]Institute of Human Genetics, Christian Albrechts University, Kiel, Germany. [6]Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Kiel, Germany. [7]Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. [8]Cell Networks, Bioquant, University of Heidelberg, Heidelberg, Germany. [9]Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [10]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [11]Department of Otorhinolaryngology, University of Duisburg-Essen, Essen, Germany. [12]Pediatric Hematology and Oncology, University Hospital Münster, Münster, Germany. [13]Leibniz Institut, German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany. [14]Department of Hematology and Oncology, Georg Augusts University of Göttingen, Göttingen, Germany. [15]Institute of Pharmacy and Molecular Biotechnology, Bioquant, University of Heidelberg, Heidelberg, Germany. [16]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, UK. [17]Friedrich Ebert Hospital Neumünster, Clinics for Haematology, Oncology and Nephrology, Neumünster, Germany. [18]Institute of Pathology, Charité, University Medicine Berlin, Berlin, Germany. [19]Hematology and Oncology, Department of Internal Medicine II, University Medical Centre, Campus Kiel, Kiel, Germany. [20]Department of Molecular Biology, Radboud University, Faculty of Science, Nijmegen, the Netherlands. [21]Hematopathology Section, Christian Albrechts University, Kiel, Germany. [22]Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig, Germany. [23]Departamento de Anatomía Patológica, Farmacología y Microbiología, Universitat de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. [24]Institute of Pathology, Medical Faculty of the Ulm University, Ulm, Germany. [25]Pediatric Hematology and Oncology, University Hospital Giessen, Giessen, Germany. [26]Institute of Clinical Molecular Biology, Christian Albrechts University, Kiel, Germany. [27]RNomics Group, Fraunhofer Institute for Cell Therapy and Immunology (IZI), Leipzig, Germany. [28]Santa Fe Institute, Santa Fe, New Mexico, USA. [29]Max Planck Institute for Mathematics in Sciences, Leipzig, Germany. [30]A full list of members and affiliations appears in the **Supplementary Note**. [31]These authors contributed equally to this work. [32]These authors jointly supervised this work. Correspondence should be addressed to R.S. (rsiebert@medgen.uni-kiel.de), B.R. (b.radlwimmer@dkfz.de) or S. Hoffmann (steve@bioinf.uni-leipzig.de).

ONLINE METHODS

Tissue samples, basic characterization and sequencing of the ICGC MMML-Seq cohort. The ICGC MMML-Seq study has been approved by the ethics committees of the Medical Faculty of the University of Kiel (A150/10) and of the recruiting centers. The ICGC MMML-Seq cohort consists of pretreatment tumor tissue and corresponding germline material (peripheral blood, buffy coats shown to be tumor free by clonality analyses) obtained with informed consent of the respective patients and/or their legal guardians, in the case of minors. In addition, sorted germinal center B cells from non-neoplastic tonsils were analyzed. In all tumor samples, basic characterization including histopathological panel review and immunohistochemical and FISH analyses was performed as described recently[8]. The tumor cell content in the cryopreserved sample material was estimated to be at least 60% in all cases. Experimental procedures for DNA and RNA extraction, the detection and sequencing of immunoglobulin rearrangements, and whole-genome and transcriptome sequencing of the ICGC MMML-Seq patient materials have been published previously[8]. All working steps were performed according to the manufacturer's instructions or standard protocols unless otherwise stated. The sequences for PCR and sequencing primers are listed in **Supplementary Table 17**.

Validation sample cohort. A cohort of 103 previously characterized mature aggressive B cell lymphomas from the Deutsche Krebshilfe Network 'Molecular Mechanisms in Malignant Lymphomas' (MMML) in which previous studies had identified an *IG-MYC* fusion was used for validation purposes. These cases have been extensively characterized by histopathological review, immunohistochemistry, interphase FISH, gene expression profiling, array comparative genomic hybridization (CGH) or SNP arrays, and mutation analysis as described recently[8]. The protocols of the MMML network have been approved by central (University of Göttingen) and local (Institutional Review Board of the Medical Faculty of the University of Kiel, D403/05) review boards.

Cell lines. The DG-75, BL-2, KARPAS-422 and Ca46 cell lines were obtained from the German Cell Culture Collection (DSMZ). Cell lines tested negative for mycoplasma contamination, and their authenticity was confirmed by STR analysis using the StemElite ID System (Promega).

DNA methylation analysis. *Whole-genome bisulfite sequencing.* Strand-specific MethylC-seq libraries were prepared using the approach described by Lister *et al.*[38] with the following modifications. Adaptor-ligated DNA fragments with insert lengths of 200–250 bp were isolated, and bisulfite conversion was performed using the EZ DNA Methylation kit (Zymo Research). PCR amplification of the fragments was performed in six parallel reactions per sample using the FastStart High-Fidelity PCR kit (Roche). The thermocycling conditions included an initial incubation at 95 °C for 2 min, eight cycles at 95 °C for 30 s, 65 °C for 20 s and 72 °C for 45 s, and a final incubation at 72 °C for 7 min. Library aliquots were pooled for each sample and sequenced on the Illumina HiSeq 2000 platform, yielding 835 million (s.d. = 81 million) 101-bp paired-end reads per sample on average.

HumanMethylation450k BeadChip. For bisulfite conversion of 0.5 to 1.0 µg of genomic DNA, the EZ DNA Methylation kit was applied according to the protocol supplied by the manufacturer. DNA methylation analysis using the Infinium HumanMethylation450k BeadChip (Illumina) was performed according to the manufacturer's instructions. The HumanMethylation450k BeadChip allows the assay of more than 480,000 CpG sites in parallel[39]. DNA methylation data were processed using GenomeStudio software (v2011.1; methylation module 1.9.0; Illumina), applying the default settings. The intrinsic controls present on the array were used for data normalization.

ChIP-seq library preparation protocols. ChIP-seq data on H3K4me1, H3K4me3, H3K27ac, H3K9me3, H3K36me3 and H3K27me3 modification-sensitive sites were obtained from the BL-2, DG-75 and KARPAS-422 cell lines using standard protocols generated within the BLUEPRINT Consortium.

Immunohistochemistry for MYC and SMARCA4. Conventional and fluorescent double-staining immunohistochemistry for MYC and SMARCA4 was carried out according to standard procedures using rabbit monoclonal antibody to c-MYC (ab32072, clone Y69, Abcam; 1:100 dilution at pH 6.0) and mouse monoclonal antibody to Brg1 (SMARCA4) (sc-17796, clone G-7, Santa Cruz Biotechnology; 1:75 dilution at pH 6.0). For conventional immunohistochemistry, the Histofine Simple Stain MAX PO (Multi) secondary antibody for simultaneous use with mouse and rabbit primary antibodies (Medac) was used. Alexa Fluor 488–conjugated donkey anti-rabbit (A21206) and Alexa Fluor 555–conjugated donkey anti-mouse (A31570) secondary antibodies (Life Technologies) were used for fluorescent double staining. Photographs were taken using an Axiophot (Zeiss) and a mounted ProgRes CF camera (Jenoptik) or a mounted SPOT RT Slider digital CCD camera (Diagnostic Instruments).

PKA inhibition. The Ca46 Burkitt lymphoma cell line was seeded at a density of 10,000 cells/well in duplicate. For PKA inhibition, cells were treated with various concentrations of the PKA inhibitor PKI-6-22 (Santa Cruz Biotechnology), ranging from 25 to 500 nM. After 72 h of treatment, cell viability was assessed using the CellTiter-Blue assay (Promega).

Bioinformatics and statistical analysis. *Whole-genome sequencing data processing.* Sequencing reads were mapped and aligned to the hg19 reference assembly plus decoy sequences (hs37d5) using Burrows-Wheeler Aligner (BWA)[40] (version 0.6.1), and somatic SNVs were identified as described in Jones *et al.*[41]. Indels were called by running Platypus[42] on tumor and matched control samples. Empirical filters based on the filters built into Platypus, the read depth and variant read fraction in tumor and control samples, base qualities and genotype qualities were applied to extract high-confidence somatic indels.

SNVs and indels were annotated using ANNOVAR according to GENCODE gene annotation (v17) and overlapped with variants from dbSNP (build 135) and the 1000 Genomes Project database. SNV, indel and mutation calling are described in Jones *et al.*[41]. The mutational landscape of the samples is summarized in **Supplementary Figure 20**.

Whole-genome methylation analysis. Mapping was carried out using segemehl[16,43] in its bisulfite MethylC-seq mode with default parameters and the hg19 reference genome. Multiple lanes belonging to a single patient were mapped independently and merged together before methylation calling. Methylation calling was carried out using BAT (the Bisulfite Analysis Toolkit; H.K., C.O., A. Viehweger, P.F.S. and S. Hoffmann, unpublished data), which calculates the methylation rate, C/(C + T), for each cytosine.

We restricted our analysis to positions that showed a strand-informative coverage of between 15 and 150 reads in all samples of our study. This is coverage calculated on the basis of reads that contain information for a cytosine with regard to its respective strand. Only reference CpG positions were taken into account, and reference non-CpG positions were analyzed separately. HumanMethylation450k BeadChips were used for verification. A linear model of the HumanMethylation450k BeadChip β values and the WGBS methylation rates was generated using the stats package in R (3.1.0).

Assessment of global, genome-wide methylation as well as chromatin state–specific methylation was performed using average (arithmetic mean) methylation levels.

Next, to obtain genome-wide segmentation of the methylation data, we performed top-down CBS similar to the method presented in Zhang *et al.*[44] (F.J., H.K., S.H.B., C.O. and P.F.S. *et al.*, unpublished data). In a preprocessing step, CBS methylation gaps larger than 500 nt (for example, centromeres) were excluded. Assessment of differential methylation between prototypical Burkitt lymphoma, prototypical follicular lymphoma and germinal center B cells was carried out by computing pairwise differences in the group-wise average methylation. These differences were segmented using the modified CBS approach and tested for significance using sample methylation rates with a non-parametric test. DMRs were defined as the top 10% and the bottom 10% of segments. Furthermore, a DMR was required to have a minimum of ten CpGs; our method did not impose further size constraints.

HumanMethylation450k BeadChip methylation analysis. DNA methylation data were processed using GenomeStudio (v2011.1; methylation module 1.9.0) and Qlucore software, applying the default settings.

Whole-genome transcriptome analysis. Transcriptome data were mapped with segemehl[16,43], allowing for spliced alignments and using a minimum accuracy of 90%. Differential expression was analyzed using DESeq[45] with default parameters and a significance criterion of 0.05 (*P* value adjusted for multiple testing).

DMR correlation analysis. cDMRs were identified on the basis of prototypical Burkitt lymphoma (*IG-MYC* translocation positive; $n = 12$) and prototypical follicular lymphoma (*IG-BCL2* translocation positive and World Health Organization (WHO) grade 1 or 2; $n = 8$) cases. To obtain DMRs correlating with the expression of nearby genes (DMRs overlapping a gene or at most 1,500 nt upstream of its TSS), no linear relationship was assumed and a Spearman correlation test was performed. GENCODE[46] v14 gene annotation was used for correlation analysis, including protein-coding genes and other biotypes, such as antisense genes, long intergenic noncoding RNAs (lincRNAs), microRNAs and other noncoding genes (**Supplementary Table 18**). DMRs with a significant correlation test ($P < 0.05$) were declared to be cDMRs.

Control of confounding factors. Control of confounding factors was performed for age bias (**Supplementary Fig. 21**), tumor cell content (**Supplementary Fig. 22**), proliferation rate as measured by Ki67 expression (**Supplementary Figs. 15** and **22**), and homogeneity of DMRs and cDMRs (**Supplementary Fig. 23**).

To test the hypothesis that transcription factor expression and transcription factor target gene expression for TCF3, YY1, FOXM1, ZEB1, SMARCA4, STAT5A, PML, BATF and BCL3 are not associated with proliferation status, we pooled all probe sets of the transcription factor genes (**Supplementary Fig. 15a**) and their targets (**Supplementary Fig. 15b**) and correlated them to Ki67 staining, a marker of proliferation, in germinal center DLBCL (GCB DLBCL; $n = 160$). In both cases, no meaningful correlation (Spearman correlation = 0.02 and 0.043, respectively) was found. To analyze whether a subset of the transcription factor genes or their targets are associated with proliferation status, we additionally carried out clustering analysis (**Supplementary Fig. 15c–f**). In both supervised and unsupervised clustering, we found no such association.

Transcription factor binding site enrichment analysis. For all binding sites of each transcription factor annotated in GM12878 (refs. 46–48), we checked whether they were significantly enriched in one of the four classes of cDMRs.

If available, replicates were merged. For significance testing, bootstrapping was used. As a background model for this test, we used the set of all prototypical Burkitt lymphoma and follicular lymphoma DMRs that were associated with a gene, including noncorrelating DMRs.

Activity plots. For each cDMR-gene pair, an activation score y (where $y = \rho \times \log_2$ (fold change)) was calculated, where ρ was the Spearman correlation coefficient of the cDMR-gene pair. These activation scores were assigned to transcription factors with a binding site in the cDMR. For transcription factors with more than one binding site in different cDMRs belonging to the same gene, the average of the activation scores was used. The activation scores for the top 40 transcription factor–gene pairs were plotted using the gplots package in R (3.1.0)[44].

Correlation tests, tests for equality of groups and box plots. To increase robustness and to avoid the violation of assumptions (for example, variance homogeneity and normal distribution), non-parametric tests were used throughout this study. For all correlations, we used Spearman's rank correlation method. The equality of the distributions of mean methylation values between the groups (**Fig. 1b**) was tested using a two-tailed non-parametric Wilcoxon-Mann-Whitney test. The upper and lower 'hinges' of all box plots correspond to 25th and 75th percentiles. The whiskers extend from the hinge to the highest or lowest value that is within the 1.5-fold interquartile range of the hinge. Data falling beyond the whiskers are outliers and are plotted as points.

ChIP segmentation for the BL-2, DG-75 and KARPAS-422 cell lines. Reads were mapped to a sex-matched reference (GRCh37) using BWA 0.5.9, with the read trimming parameter (-q) set to 15. Alignments were sorted and duplicates were marked with Picard. Mappings with a MAPQ score below 15 were filtered out with SAMtools.

We segmented the DG-75, BL-2 and KARPAS-422 cell line genomes with the ChIP-seq data for the six histone modifications described above. For this purpose, we used the implementation described by Ernst *et al.*[49] in ChromHMM software (v1.03). The input data were ChIP-seq BED files with the genomic coordinates and strand orientation of mapped sequences (after removing duplicate reads). The genome was divided into consecutive 200-bp non-overlapping intervals, each of which was independently assigned a notation of present (1) or absent (0) for each of the six chromatin modifications. Assignment was based on the count of tags mapping to the interval and on the Poisson background model using a threshold of 1×10^{-4}, as explained in Ernst *et al.* After binarization and for segmentation, we used the 11-state model established by the BLUEPRINT Consortium. Further, we computed the probability that each location was in a given chromatin state and then assigned each 200-bp interval to its most likely state for each sample. Lastly, consecutive intervals in the same chromatin state were joined.

SMARCA4 modeling. The structure of the N-terminal helicase domain (DEXDc) and conserved C-terminal helicase domain of human SMARCA4 (residues Gln756 to Val1224) interacting with double-stranded DNA (dsDNA) was modeled on the structure of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core in complex with dsDNA[50] using Modeller[51]. This structure was superposed on the structure of *Mycobacterium tuberculosis* SecA[52] bound to ADP, to position ADP in the final SMARCA4-dsDNA-ADP model shown in **Figure 5g**.

Array-based gene expression analysis. For the MMML cohort comprising 935 samples, including primary lymphomas, lymphoma cell lines and normal B cell controls, array-based gene expression data from Affymetrix GeneChip U133A arrays were available. Data from all gene expression arrays were jointly normalized with the variance-stabilizing normalization (VSN) method[53] as described previously[21]. On the basis of gene expression, a 'molecular Burkitt lymphoma (mBL) index' (ref. 21) was calculated for each individual sample and was assigned one of the following molecular diagnoses; mBL (index score ≥ 0.95), non-mBL (index score ≤ 0.05) or molecular intermediate (remaining cases)[21]. The lymphomas were also stratified according to their 'pathway activation patterns' (ref. 54). The cell of origin was classified according to the methods described by Wright *et al.*[42] using a modified classifier[21].

To characterize the expression landscape of specific gene sets in lymphoma subgroups, the samples were classified in two ways. First, cell lines ($n = 32$) were compared with lymphomas classified as mBL ($n = 84$)[21] and follicular lymphomas ($n = 144$). DLBCLs were subdivided into activated B cell (ABC; $n = 106$), germinal center B cell (GCB; $n = 176$) and unclassified ($n = 80$) subtypes[55]. Tonsil samples ($n = 10$), naive B cells ($n = 8$), germinal center B cells ($n = 13$) and post–germinal center B cells ($n = 9$) were used as normal controls.

In a second, more stringent approach, lymphomas with the mBL signature[21], the Burkitt lymphoma pathway activation pattern[54] and *IG-MYC* translocation and without HIV infection were classified as prototypical Burkitt lymphomas ($n = 58$). Lymphomas with a diagnosis of follicular lymphoma grade 1 or 2, with t(14;18) translocation and without DLBCL components were classified as prototypical follicular lymphomas ($n = 75$). Germinal center B cell controls did not have Epstein-Barr virus (EBV) infection ($n = 10$).

38. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
39. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288–295 (2011).
40. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
41. Jones, D.T. *et al.* Recurrent somatic alterations of *FGFR1* and *NTRK2* in pilocytic astrocytoma. *Nat. Genet.* **45**, 927–932 (2013).
42. Rimmer, A. *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).

doi:10.1038/ng.3413

43. Hoffmann, S. *et al.* Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.* **5**, e1000502 (2009).
44. Zhang, N.R. & David, O.S. Model selection for high-dimensional, multisequence change-point problems. *Stat. Sin.* **22**, 1507 (2012).
45. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
46. Karolchik, D. *et al.* The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* **42**, D764–D770 (2014).
47. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
48. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755 (2014).
49. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
50. Dürr, H. *et al.* X-ray structures of the *Sulfolobus solfataricus* SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**, 363–373 (2005).
51. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **47**, 5.6.1–5.6.32 (2006).
52. Sharma, V. *et al.* Crystal structure of *Mycobacterium tuberculosis* SecA, a preprotein translocating ATPase. *Proc. Natl. Acad. Sci. USA* **100**, 2243–2248 (2003).
53. Huber, W. *et al.* Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**, S96–S104 (2002).
54. Bentink, S. *et al.* Pathway activation patterns in diffuse large B-cell lymphomas. *Leukemia* **22**, 1746–1754 (2008).
55. Wright, G. *et al.* A gene expression–based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc. Natl. Acad. Sci. USA* **100**, 9991–9996 (2003).