Published in final edited form as:
Nat Genet. 2015 November ; 47(11): 1316–1325. doi:10.1038/ng.3413.

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

A full list of authors and affiliations appears at the end of the article.

[#] These authors contributed equally to this work.

Abstract

In spite of both having features of germinal center B-cells, Burkitt lymphomas and follicular lymphomas are biologically and clinically quite diverse. We here performed whole genome bisulfite, genome and transcriptome sequencing from 13 *IG-MYC*-translocation positive Burkitt lymphoma, 9 *BCL2*-translocation positive follicular lymphoma and four normal germinal center B-cell samples. Comparison of Burkitt and follicular lymphoma samples revealed differential methylation of intragenic regions that strongly correlated with expression of associated genes, e.g. genes active in germinal center dark zone and light zone B-cells. Integrative pathway analyses of regions differentially methylated between Burkitt and follicular lymphoma implicated DNA methylation to cooperate 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 between Burkitt and other germinal center B-cell lymphomas.

Introduction

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 center1–3, where genes encoding

Correspondence: Reiner Siebert, rsiebert@medgen.uni-kiel.de; Bernhard Radlwimmer, b.radlwimmer@dkfz.de; Steve Hoffmann, steve@bioinf.uni-leipzig.de.
[*]These authors jointly supervised this work.
A full list of members and affiliations of the ICGC and Blueprint projects appears in the Supplementary Note.

Author Contributions

A.H., O.A., R.K., P.L., R.S., S.H., 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.L., S.P., I.V., P.R., M.S., 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.H. performed statistical analysis. H.K., S.H.B., G.D., V.H., D.R., F.J., C.O., M.H., M.K., M. Kulis., I.N., M.R., M. Schlesner, S.H. and B.R. analyzed the data. H.K., M.A.W., M.J.B., E.C., G.D., B.H., F.J., Q.J., 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.K., 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.S., M. Schlesner, L.T., H.G.S. and S.H. contributed reagents/materials/analysis tools. H.K., S.H.B., W.W., A.H., P.L., R.S., S.H. and B.R. wrote the paper.

Competing financial interests

The authors declare that they have no competing interests as defined by Nature Publishing Group, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Europe PMC Funders Author Manuscripts

immunoglobulins (IG) are somatically altered to optimize antibody affinity and avidity. The B-cell lymphomas most common in childhood, i.e. Burkitt lymphoma/leukemia, and in adulthood, i.e. diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma, are supposed 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 genetic alterations and gene expression patterns of the tumor cells, Burkitt lymphoma and to a considerable extent also follicular lymphoma are much more homogenous lymphoma subtypes4,5. Burkitt lymphoma is almost invariably associated with an *IG-MYC* chromosomal translocation, i.e. juxtaposition of the *MYC* oncogene encoding a transcription factor next to one of the three immunoglobulin loci. The vast majority of follicular lymphoma, particularly those of grades 1 and 2, carry a t(14;18)(q32;q21) chromosomal translocation juxtaposing the *BCL2* oncogene next-to mostly the immunoglobulin heavy chain *(IGH)* locus6. Both lymphoma subtypes also display a different spectrum of secondary changes with e.g. mutations in cell cycle regulators and in the ID3/TCF3 transcription factor complex being highly prevalent in Burkitt lymphoma7–9. This deregulation of particular transcription factors in Burkitt lymphoma as well as recurrent mutations in chromatin modifiers such as *CREBBP*, *MEF2B*, *EZH2* and *KMT2D* found in follicular lymphoma indicate that lymphoma pathogenesis also involves alterations in physiologic epigenetic modulation10–14. Clinically, patients affected with Burkitt lymphoma are significantly younger at diagnoses than those affected with follicular lymphoma. Moreover, Burkitt lymphoma shows a much higher proliferation rate and, thus, clinical aggressiveness than follicular lymphoma which is a more indolent disease. Intriguingly, despite these considerable differences, both, Burkitt lymphoma and follicular lymphoma, are supposed to derived from the very same cell compartment of the human body, the germinal center B-cell, as described above. 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, i.e. Burkitt lymphoma resembling the dark-zone germinal center B-cell state and follicular lymphoma resembling the light-zone germinal center B-cell state15. This offers the unique opportunity to study to 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 of the both germinal center B-cell lymphomas, we performed whole genome bisulfite (WGBS) and transcriptome sequencing of 13 *IG-MYC*-translocation positive Burkitt lymphoma and 9 *BCL2*-translocation positive follicular lymphoma (including 8 follicular lymphoma grade 1 / 2 and one follicular lymphoma grade 3a/DLBCL) samples. As a reference, FACS-sorted germinal center B-cells from non-neoplastic tonsils of four donors were analyzed in parallel. The source of the sample material as well as the clinical follow-up of all patients are 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 genome16, resulting in a mean coverage of 54x (Supplementary Table 2). Methylation rates called from the sequencing data showed excellent correlation with beta values obtained from Illumina HumanMethylation450K BeadChip analysis (Supplementary Table 3) from 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 30x whole genome sequencing of tumors and matched controls from the same patients, and the germinal center B-cell populations, as well as 8 additional Burkitt lymphoma (Supplementary Table 5).

Analysis of average genome wide CpG methylation revealed all lymphoma entities to be hypomethylated compared 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 DNA-methylation gains were observed in Burkitt lymphoma and follicular lymphoma samples at CpGs with low-level methylation in germinal center B-cells (GC-B<0.3; Fig. 1c). To determine whether DNA methylation patterns vary between genome segments with different function, we compared methylation differences (average of Burkitt lymphoma and follicular lymphoma vs. germinal center B-cells) for 15 chromatin states (Supplementary Table 6) defined in the GM12878 lymphoblastoid cell line, representative of non-neoplastic mature B-cells17. For most chromatin states, individual segments showed either hyper- or hypomethylation largely neutralizing each other in this global analysis. However, changes in heterochromatin were mostly unidirectional, resulting in the greatest relative reduction of DNA methylation levels of any chromatin segment in the lymphoma samples. In stark contrast, poised promoters showed the strongest, but nevertheless small, increase in DNA methylation levels at a global level (Fig. 1d). In line with recent reports18, 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; Supplementary Fig. 2).

DNA methylation downstream of promoters correlates with gene expression

Gene-centered analysis, including coding and non-coding genes, revealed increased levels of gene body methylation for highly expressed genes, which returned to background levels at transcription end sites (TES). High transcription levels typically were associated with the presence of large hypomethylated regions downstream of the transcription start sites (TSS), suggesting regulatory relevance for these regions (Fig. 2a; Supplementary Fig. 3). We applied an unbiased circular binary segmentation approach to our entire WGBS dataset, identifying 90,350 differentially methylated regions (DMRs) with 10 or more CpGs between Burkitt lymphoma, follicular lymphoma and germinal center B-cell controls. These DMRs included 36,775 and 27,607 DMRs between normal germinal center B-cells and follicular lymphoma and Burkitt lymphoma groups, respectively (Supplementary Table 7). The great majority of DMRs were located within transcribed (TSS to TES) rather than in promoter (1,500 nt upstream of TSS) or intergenic regions (Fig. 2b), with strong enrichment of DMRs

immediately downstream of TSS (Fig. 2c), analogous to the so-called "promoter downstream correlated regions" recently described in medulloblastoma19. No substantial size differences 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 methylation levels of intragenic DMRs (contained in the transcribed regions and up to 1,500 bp upstream of the TSS) with expression levels of the associated genes containing these DMRs. 40% of the intragenic DMRs showed significant ($p<0.05$) correlations between methylation and RNA expression (Fig. 2d), with the majority (64%) showing negative correlations (Fig. 2e; Supplementary Table 8). Overall, 85% of correlating DMRs (cDMRs) were located downstream of TSS. Negatively and positively cDMRs did not differ in size; however, the former were highly enriched downstream of TSS while the latter were more evenly distributed across the gene body (Supplementary Fig. 4). Examples of negatively cDMRs are shown in Supplementary Fig. 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 (DZ) or light zone (LZ) of the germinal center15 were examined; around 22% of DZ- and 28% of LZ-specific genes possessed cDMRs. DZ-specific genes were predominantly hypomethylated and upregulated in Burkitt lymphoma, while hypermethylated and downregulated in follicular lymphoma samples ($p<0.001$; Fisher's exact test, Figs. 2f,g); LZ-specific genes showed the opposite pattern ($p<0.001$; Fisher's exact test, Figs. 2f,h). These data are in line with the view that the majority of Burkitt lymphoma are frozen in the DZ cell state while the majority of follicular lymphoma are frozen in the LZ cell state; moreover, they suggest that methylation at cDMRs could act to fix the otherwise physiologically transient functional states of DZ and LZ B-cells in Burkitt lymphoma and follicular lymphoma samples, respectively. Examples include high expression of *TCF3* and *SMARCA4*, higher expressed in DZ than in LZ 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 LZ- and DZ-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 based on (i) all genes differentially expressed between 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 negative control. The first gene set identified 60 pathways showing significant 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 modulation of DNA methylation levels at negatively correlated cDMRs 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 since the latter analysis would have been underpowered and potentially biased due to the smaller number of available germinal center B-cell samples (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 gene expression changes in those activated pathways relative to germinal center B 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), an RNA-binding protein that stabilizes *MYC*20, underscoring its central importance in germinal center B-cell lymphoma biology.

Detailed pathway-based analyses showed that, e.g., DNA methylation was significantly associated with the downregulation of NFκB signaling and the upregulation of cell cycle control genes in Burkitt lymphoma (Supplementary Figs 10, 11), concurrent with the proliferation rates of close to 100%, typical for Burkitt lymphoma (Ki67 staining; Supplementary Table 1), and with previous reports on the molecular pathogenesis of Burkitt lymphoma7–9,21,22. The observed aberrant signaling was linked to differential methylation of key pathway genes in Burkitt lymphoma, including the hypermethylation and downregulation of *JAK3* and *STAT3*7 (Figs. 2h, i, j) in line with low JAK-STAT-pathway activity in Burkitt lymphoma. Another example of differential methylation of key regulatory genes in Burkitt lymphoma was hypomethylation and overexpression of the transcription factor *TCF3* (Fig. 2k–m). Remarkably, the *TCF3* gene even harbors two cDMRs of which cDMR1 contains a TCF3 binding site that is occupied in B-cells17, suggesting the possibility of a positive-feedback loop amplifying *TCF3* transcription (Fig. 2k, l). Analysis of published TCF3-ChIP-Seq data of the Burkitt lymphoma cell lines BL-41 and NAMALWA9 confirmed TCF3 binding within cDMR1 that overlapped a common TCF3 binding motif, and 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 mechanisms8,9,22, *TCF3*-cDMR hypomethylation contributes to TCF3 activation.

DNA mutation and methylation in sphingosine signaling

Pathway analysis also revealed deregulation of sphingosine-1-phosphate signaling in both Burkitt lymphoma and follicular lymphoma when compared to germinal center B-cells (Fig. 3a). Sphingosine-1-phosphate receptor 2 (S1PR2) is frequently affected by loss-of-function mutations in Burkitt lymphoma and DLBCL, leading to increased AKT and migratory activity23. 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), the stimulatory G-alpha-q protein

GNA11, upregulated in Burkitt lymphoma (Fig. 3c), and *S1PR1* and the G-alpha-protein complex G(12/13)-member *GNA12*, which are overexpressed in follicular lymphoma (Figs. 3d, e). Analysis of whole genome sequencing data detected mutually exclusive mutations in several G(12/13)-complex genes in 14 of 21 Burkitt lymphoma samples analyzed (Fig. 3f). These findings were corroborated by our previous description of *RHOA* mutations in 8.5% of Burkitt lymphoma samples in an extended cohort of 82 cases24 and by a recent study functionally implicating G(12/13) mutations in the development of germinal center B lymphomas23. Interestingly, 3 of the remaining 7 samples were affected by somatic mutations in two genes of the Gαi complex, *GNAI1* and *GNAI2*, in 5 Burkitt lymphoma samples, including recurrent (n=3) GNAI2 R179H substitutions. The GαI complex can regulate several signaling pathways including AKT, ERK and protein kinase A (PKA). 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 the Gαs protein GNAS at the position homologous to GNAI2 R179H are known to cause upregulation of AKT signaling and cell migration in pancreatic adenocarcinoma and pituitary tumors25,26. Recently, it was proposed that, in addition to G(12/13) mutations, GNAI2 R179H mutations might increase proliferation and migration of DLBCL cells27. Our observations on sphingosine-phosphate and G-protein signaling suggest that mutation and 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

Since 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 (TFs) driving expression of genes with cDMRs. We intersected the cDMR dataset with TF binding sites (TFBS) data for 46 TF17,28. Significant enrichment of TFBS 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 (Q2) of the radar plot (Fig. 4a). We selected the ten TFs displaying the highest correlation of TF and target gene expression (Supplementary Table 11) for subsequent analyses. Differential methylation of the corresponding TFBS typically was restricted to the TFBS itself and rapidly returned to background levels outside of it (Supplementary Fig. 14).

To understand how cDMR-methylation modulates TF-target gene expression, we plotted average methylation levels of cDMRs against average RNA expression of the associated genes for all 10 TFs (Fig. 4b). Target genes of STAT5A and BCL3, both enriched in Q2 (Fig. 4a), displayed hypermethylation and reduced RNA expression in Burkitt lymphoma when compared to germinal center B-cell and follicular lymphoma samples, in agreement with the low level activity of the NFκB and JAK/STAT pathways previously reported in Burkitt lymphoma7. Similar patterns were observed for the target genes of other TFs in Q2. In Q4, 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

Burkitt lymphoma. As previously noted, mutations affecting *TCF3*, or its negative regulator *ID3*, detected by us and others8,22 in more than two-third of all sporadic Burkitt lymphoma have been proposed to foster TCF3-dependent gene expression driving Burkitt lymphoma-cell proliferation9. Importantly, neither TF nor target gene expression were correlated with cell proliferation differences between Burkitt lymphoma and follicular lymphoma (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 TFBS.

To assess the impact of these TFs 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 TF 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 the polycomb repressive complex 1, which was previously shown to be functionally involved in lymphomagenesis with high expression in follicular lymphoma29. 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-lymphocytes7 in follicular, in contrast to Burkitt, lymphomagenesis. *TCF3* was among the genes strongly activated in Burkitt lymphoma, in line with the already above proposed auto-activation loop of its expression (Fig. 4c, d). Other strongly activated genes included the E2F family member *E2F1*, the *POLD1* gene, which encodes the MYC-interacting polymerase D1, the *TERT* gene, which encodes telomerase as well as the *SMARCA4* gene (Fig. 4c, e) encoding a member of the SWI/SNF nucleosome remodeling complex. Consistent with previous findings21, we could independently confirm overexpression of *SMARCA4* RNA in Burkitt lymphoma using array-based analysis (Supplementary Fig. 16; 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 its binding sites in the target genes, but rather is deregulated 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 (Chr19: 11.07-11.09 Mb) and enhancers to promoters (Chr19: 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, Supplementary Table 13); in contrast, Burkitt lymphoma samples showed ubiquitous, high SMARCA4 protein expression in (Fig. 5c, Supplementary Table 13).

Whole genome sequencing of the ICGC-MMML-Seq cohort identified *SMARCA4* mutations in 9 out of 21 Burkitt lymphoma samples. In line with previous reports30,31, these *SMARCA4* mutations clustered in the helicase domain (Fig. 5d), revealing a mutation pattern that stands in strong contrast to the germline nonsense mutations predisposing to cancer syndromes32–35. Further investigation of the SMARCA4 mutational landscape from 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 wildtype *SMARCA4* transcripts had similar expression levels (data not shown) and immunohistochemistry and immunofluorescence analyses of Burkitt lymphoma samples showed constant co-expression of SMARCA4 and MYC proteins in Burkitt lymphoma, regardless of the presence or absence of *SMARCA4* mutations (Fig. 5c; Supplementary Fig. 17; Supplementary Table 13).

Intriguingly, genomic mutations of other members of the SWI/SNF complex, including *ARID1A*, *ARID1B* and *SMARCB1* were present in a subset of Burkitt lymphoma samples lacking SMARCA4 mutation (Fig. 5e; Supplementary Table 15), and, moreover, SWI/SNF complex genes showed increased expression levels in Burkitt lymphoma samples (Fig. 5f). Protein modeling36 (Fig. 5g) suggests 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 or by obstructing the interaction of the helicase N and C-terminal domains. It has recently been suggested, that such oncogenic SMARCA4 mutations like those observed herein 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 at comparing 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 levels determined by WGBS in the sorted germinal center B-cells in both studies showed an excellent agreement (correlation of means = 0.9799; see Supplementary Fig. 18a). Principal component analysis of the lymphomas analyzed herein together with the recently investigated normal B-cell populations showed the Burkitt lymphoma and follicular lymphoma to cluster together but distinctly from the normal B-cell subsets (Supplementary Fig. 18b). This clustering of the follicular lymphoma and Burkitt lymphoma separately from the physiologic germinal center B-cells relates most likely to the neoplastic process. To further investigate this, we analyzed the differences in DNA methylation between Burkitt lymphoma and follicular lymphoma, respectively, as compared 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 diffuse large B-cell lymphoma by Kulis et al., we observe 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 extends the previous findings to the fact that also Burkitt

lymphoma and follicular lymphoma frequently acquire methylation changes in regions already undergoing dynamic methylation during normal B cell differentiation. Remarkably, when directly comparing the DNA methylation of Burkitt lymphoma and follicular lymphoma 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 DNA methylation dynamics of normal B-cell differentiation. The only notable exception is module 11, which tends to lower DNA methylation in Burkitt lymphoma than in follicular lymphoma. Remarkably, the CpGs of this module, which are strongly enriched for enhancer sites, are selected by the fact that they switch their DNA methylation state from the naïve B-cell to the germinal center B-cell stage. This 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, respectively, 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 and TCF3/ID3 and Gαi 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 due to its maintained binding competence in the absence of helicase activity.

Accession codes

WGBS data, mRNA-seq data, whole genome DNA-Seq

All sequencing data is available at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession number EGAS00001001067 Please contact the data access committee (DAC; http://www.icgc.org/daco) of the International Cancer Genome Consortium (ICGC, http://www.icgc.org/) to get access to the data.

Chip Seq-Data

The raw sequencing read data is publically available from the European Nucleotide archive (ENA) (https://www.ebi.ac.uk/ena/). All ChIP cell line data is available under study ERP002586. The run identifiers are listed below:

	Input	H3K27ac	H3K27me3	H3K36me3	H3K4me1	H3K4me3	H3K9me3
BL-2	ERR324294	ERR324306	ERR324300	ERR324287	ERR324266	ERR324263	ERR324280
DG-75	ERR324296	ERR324307	ERR324302	ERR324289	ERR324304	ERR324262	ERR324282
KARPAS-422	ERR324297	ERR324271	ERR324276	ERR324290	ERR324268	ERR324270	ERR324283

A complete list of the raw files available from the ftp is listed together with associated meta data in the data index file (indexed with secondary sample ERS accession; ftp://ftp.ebi.ac.uk/pub/databases/blueprint/releases/20140811/homo_sapiens/20140811.data.index).

Online Methods

Tissue samples, basic characterization and sequencing of the ICGC MMML-Seq cohort

The ICGC MMML-Seq study has been approved by Ethics Committee of the Medical Faculty of the University of Kiel (A150/10) and of the recruiting centers. The ICGC MMML-Seq cohort consists of pre-treatment 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 in minors their legal guardian. In addition, sorted GC-B-cells from non-neoplastic tonsils were analyzed. In all tumor samples, basic characterization including histopathologic panel review, immunhistochemical and FISH analyses is being performed as described recently8. The tumor cell content in the cryo-preserved sample material was estimated to be at least 60% in all cases. Experimental procedures for DNA and RNA extraction, detection and sequencing of immunoglobulin rearrangements as well as whole genome and transcriptome sequencing of the ICGC MMML-Seq patient materials have been published previously8. All working steps were performed according to the manufacturer´s instructions or standard protocols unless otherwise stated. The sequences of PCR and sequencing primers are listed in Supplementary Table 16.

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 histopathology review, immunohistochemistry, interphase FISH, gene expression profiling, array-CGH/SNP-arrays and mutation analysis as described recently8. The protocols of the MMML network have been approved by central (University of Göttingen) and local review boards (Institutional Review Board of the Medical Faculty of the University of Kiel, D403/05).

Cell lines

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

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 200–250 bp insert lengths 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 were: 95 °C for 2 min, 8 cycles of 95 °C for 30 s, 65 °C for 20 s and 72 °C for 45 s, ending with 72 °C for 7 min. Library aliquots were pooled per sample and sequenced on a Illumina HiSeq2000 platform, yielding 835 million (±81 million s.d.) 101 bp paired-end reads per sample on average.

HumanMethylation450k BeadChip—For bisulfite conversion of 0.5 to 1.0 μg genomic DNA the EZ DNA Methylation kit (ZymoResearch, Irvine, CA, USA) was applied according to the protocol supplied by the manufacturer. DNA methylation analysis using the Infinium HumanMethylation450k BeadChip (Illumina Inc., San Diego, CA, USA) was performed according the manufacturer's instruction. The HumanMethylation450K BeadChip allows assaying more than 480,000 CpG sites in parallel39. DNA methylation data were processed using GenomeStudio software (v2011.1; methylation module 1.9.0; Illumina, Inc.) 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, H3K27me3 modification sensitive sites was performed on the BL-2, DG-75 and KARPAS-422 cell lines using the standard protocols generated within the Blueprint Consortium (http://www.blueprint-epigenome.eu/index.cfm?p=7BF8A4B6-F4FE-861A-2AD57A08D63D0B58).

Immunohistochemistry for c-MYC and SMARCA4

Conventional and fluorescence double-staining immunohistochemistry for c-MYC and SMARCA4 were done according to standard procedures using the rabbit monoclonal anti-c-MYC antibody ab32072 (clone Y69, abcam plc., Cambridge, UK in 1:100 at pH 6.0) and the mouse monoclonal anti-Brg1 (SMARCA4) antibody sc-17796 (G-7) (Santa Cruz Biotechnology Inc., Heidelberg, Germany in 1:75 at pH 6.0). For conventional immunohistochemistry the secondary antibody Histofine® Simple Stain™ MAX PO (Multi) for simultaneous use with mouse and rabbit primary antibodies (medac GmbH, Wedel, Germany) was used. The secondary antibodies donkey anti-rabbit Alexa Fluor® 488

(A21206) and donkey anti-mouse Alexa Fluor® 555 (A31570) (Life Technologies, Darmstadt, Germany) were used for fluorescence double staining. The photographs were taken using Axiophot (Zeiss, Oberkochen, Germany) and mounted ProgRes CF camera (Jenoptik, Jena, Germany) or mounted SPOT RT Slider digital CCD camera (Diagnostic Instruments Inc., USA).

PKA Inhibition

The BL cell line Ca46 was seeded at a density of 10.000 cells/well in duplicate. For PKA inhibition cells were treated with various concentration of the protein kinase A inhibitor PKI-6-22 (Santa Cruz Biotechnology, Texas, USA) ranging from 25nM to 500 nM. After 72h of treatment the cell viability was measured using the Cell Titer Blue Assay (Promega, Mannheim, Germany).

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 Platypus42 on tumor and matched control. Empirical filters based on the filters built into Platypus, the read depth and variant read fraction in tumor and control, the base qualities and the 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 Genome project database. SNV, indel and mutation calling are described in Jones et al.41 The mutational landscape of the samples is summarized in Supplementary Fig. 20.

Whole-genome methylation analysis—The mapping was done using segemehl43,44 in its bisulfite methyl-C seq mode with default parameters and hg19 reference genome. Multiple lanes belonging to a single patient were mapped independently and merged together prior to methylation calling. The methylation calling was done using BAT (BAT - Bisulfite Analysis Toolkit, manuscript in preparation), which calculates the methylation rate, C/(C+T), for each cytosine.

We restricted our analysis to positions that showed a strand-informative coverage between 15 and 150 reads in all samples of our study. This is the coverage calculated on reads that contain information for a C on its respective strand. Only reference CpG positions were taken into account, while reference non-CpG positions were analyzed separately. HumanMethylation450k BeadChips were used for verification. A linear model of the HumanMethylation450k BeadChip beta values and the WGBS methylation rates was performed using the stats package for R (3.1.0)45.

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

Next, we wanted to obtain a genome-wide segmentation of the methylation data we performed a top-down circular binary segmentation (CBS) similar to the method presented in Zhang et al in 201246 (Manuscript in preparation, http://www.bioinf.uni-leipzig.de/Software/metilene/). In a pre-processing step to the CBS methylation gaps larger than 500 nucleotides (e.g. centromeres) have been excluded. Assessment of differential methylation between pBL, pFL and GC B was done by computing pairwise differences of 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. As differentially methylated regions (DMRs) the top 10% and the bottom 10% segments were taken. Furthermore, a DMR was required to have a minimum of 10CpGs. In addition, 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; Illumina, Inc.) and Qlucore (Qlucore AB, Sweden) software, applying the default settings.

Whole-genome transcriptome analysis—Transcriptome data was mapped with segemehl43, 44 allowing for spliced alignments and using a minimum accuracy of 90%. Differential expression was analyzed using DESeq47 with default parameters and a significance criterion of 0.05 (p-value adjusted for multiple testing).

DMR correlation analysis—Correlating DMRs were identified based on prototypic BL (pBL; IG-MYC-translocation-positive; n=12) and prototypic FL (pFL; IG-BCL2-translocation-positive and WHO grade 1 or 2; n=8) cases. To obtain DMRs correlating with the expression of nearby genes (DMR overlapping a gene or at most 1500nt upstream of its TSS), no linear relation was assumed and a spearman correlation test was performed. GENCODE48 V14 gene annotation was used for correlation analysis, including protein coding genes and other biotypes like antisense, lincRNA, miRNA and other noncoding genes (Supplementary Table 17). DMRs with a significant correlation test ($p<0.05$) were declared to be correlating DMRs (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 in Ki67 expression (Supplementary Fig. 22, Supplementary Fig. 15) and homogeneity of DMRs and cDMRs (Supplementary Fig. 23).

To test the hypothesis that transcription factor expression and transcription factor-target gene expression of TCF3, YY1, FOXM1, ZEB1, SMARCA4, STAT5A, PML, BATF and BCL3 is not associated with the proliferation status, we pooled all probesets of the TF genes (Supplementary Figure 15a) and their targets (Supplementary Figure 15b) and correlated them to the Ki-67 staining, a proliferation marker, in germinal center DLBCL (GCB DLBCL; n=160). In both cases (Spearman cor.=0.02 and Spearman cor.=0.043) no meaningful correlation was found. To analyze whether a subset of the TF genes or their targets are associated to the proliferation status we additionally carried out clustering analysis (Supplementary Fig. 15c–f). In both supervised and unsupervised clustering we found no such association.

TFBS enrichment analysis—For all binding sites of each transcription factor annotated in GM1287849 we checked whether they are significantly enriched in one of the four classes of cDMRs. TFBS data was accessed under http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibTfbs/48,50.

If available, replicates were merged. For significance testing bootstrapping was used. As a background model for this test we used the set of all pBL-pFL DMRs that were associated with a gene, i.e. including non-correlating DMRs.

Activity Plot—For each cDMR – gene pair an activation score y (y=rho*log2FC) was calculated. Rho is the spearman correlation coefficient of the cDMR – gene pair. These activation scores were assigned to TFs with a binding site in the cDMR. For TFs 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 TF – gene pairs were plotted using the gplots package for R (3.1.0)45, 46.

Correlation tests, tests for equality of groups and boxplots—To increase robustness and to avoid the violation of assumptions (e.g. variance homogeneity, normal distribution) non-parametric tests were used throughout this study. For all correlations, we used Spearmans'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 boxplots correspond to the 25th and 75th percentiles. The whisker extend from the hinge to the highest\lowest value that is within the 1.5-fold inter-quartile range (IQR) of the hinge. Data beyond the end of the whiskers are outliers and plotted as points

ChiP-segmentation for BL-2, DG-75, KARPAS-422 cell lines—Reads mapped to a gender matched reference (GRCh37) with BWA 0.5.9, with the read trimming parameter (-q) set to 15. The alignments were sorted and duplicates marked with picard. Mappings with a MAPQ below 15 were filtered out with samtools (http://ftp.ebi.ac.uk/pub/databases/blueprint/releases/current_release/homo_sapiens).

We segmented the DG-75, BL-2 and KARPAS-422 cell line genomes with the ChIP-seq data for the 6 histone modifications- described above. For this purpose, we used the implementation described by Ernst et al.51 in ChromHmm software (v1.03). The input data were the ChIP-Seq bed files with the genomic coordinates and strand orientation of mapped sequences (after remove duplicate reads). The genome was divided in consecutive 200 bp non-overlapping intervals and independently assigned present (1) or absent (0) for each of the 6 chromatin modifications. The assignment was based on the count of tags mapping to the interval and on the basis of a Poisson background model using a threshold of 10^{-4}, as explained in Ernst et al. After binarization and for segmentation we used the eleven states model established by the Blueprint Consortium: (http://ftp.ebi.ac.uk/pub/databases/blueprint/paper_data_sets/monocyte_neutrophil_2014/chromatin_states/full_histone_panel/model_11_Blueprint_11.txt). Further, we computed the probability that each location is in a given chromatin state, and then assigned each 200-bp interval to its most likely state for each sample. Lastly, consecutive intervals within the same chromatin state were joined.

SMARCA4 modeling—The structure of the helicase 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 modelled on the structure of the Sulfolobus solfataricus SWI2/SNF2 ATPase core in complex with dsDNA52 using Modeller53. This was superposed on the structure of Mycobacterium tuberculosis SecA54 bound to ADP, to position ADP in the final SMARCA4-dsDNA-ADP model shown in Fig. 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 were available. Data of all gene expression arrays were jointly normalized with the VSN method55 as described previously21. Based on gene expression a "molecular BL (mBL) index"21 was calculated for each individual sample and was assigned one of the following molecular diagnoses; mBL (index $\geq$0.95), non-mBL (index score $\leq$0.05), or molecular intermediate (remaining cases)21. The lymphomas were also stratified according to their "pathway activation patterns"56. The cell of origin was classified according to the methods described by Wright et al42 using a modified classifier21.

To characterize the expression landscape of specific gene sets within 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 lymphoma (n=144). Diffuse large B-cell lymphoma were subdivided into ABC (n=106), GCB (n=176) and unclassified (n=80)57. As normal controls tonsil samples (n=10), naïve B-cells (n=8), germinal center B-cells (n=13) and post GC B-cells (n=9) were used.

In a second more stringent approach lymphomas with mBL signature21, BL-PAP "pathway activation pattern"56, IG-MYC translocation and without HIV infection were classified as pBL (n=58). Lymphomas with diagnosis of follicular lymphoma grade I or II, with t(14;18) translocation and without diffuse large B-cell lymphoma components were classified as pFL (n=75). Germinal center B-cell controls without EBV infection were termed GC-B (n=10).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Helene Kretzmer[#1,2,3,4], Stephan H. Bernhart[#1,2,3,4], Wei Wang[#5], Andrea Haake[#4,6], Marc A. Weniger[#4,7], Anke K. Bergmann[#6,8,9], Matthew J. Betts[10], Enrique Carrillo-de-Santa-Pau[9,11], Gero Doose[1,2,3,4], Jana Gutwein[6], Julia Richter[4,6], Volker Hovestadt[5], Bingding Huang[12], Daniel Rico[9,11], Frank Jühling[1,2,3], Julia Kolarova[6], Qianhao Lu[10], Christian Otto[1,2,3], Rabea Wagener[4,6], Judith Arnolds[13], Birgit Burkhardt[4,14], Alexander Claviez[4,8], Hans G. Drexler[15], Sonja Eberth[4,15,16], Roland Eils[4,12,17], Paul Flicek[9,18], Siegfried Haas[4,19], Michael Humme[4,20], Dennis Karsch[4,21], Hinrik H.D. Kerstens[9,22], Wolfram Klapper[4,23],

Markus Kreuz[4,9,24], Chris Lawerenz[4,12], Dido Lenzek[4,20], Markus Loeffler[4,9,24], Cristina López[4,6], Roderick A.F. MacLeod[15], Joost H.A. Martens[9,22], Marta Kulis[9,22], José Ignacio Martín-Subero[9,25], Peter Möller[4,26], Inga Nage[4,6], Simone Picelli[5], Inga Vater[4,6], Marius Rohde[4,27], Philip Rosenstiel[4,28], Maciej Rosolowski[4,24], Robert B. Russell[10], Markus Schilhabel[4,28], Matthias Schlesner[4,12], Peter F. Stadler[1,2,3,4,29,30,31], Monika Szczepanowski[4], Lorenz Trümper[4,16], Hendrik G. Stunnenberg[9,22], Ralf Küppers[4,7,9,*], Ole Ammerpohl[4,6,*], Peter Lichter[4,5,*], Reiner Siebert[4,6,9,*], Steve Hoffmann[1,2,3,4,9,*], and Bernhard Radlwimmer[4,5,*]

Affiliations

[1]Transcriptome Bioinformatics, LIFE Research Center for Civilization Diseases, 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]German ICGC MMML-Seq-project [5]German Cancer Research Center (DKFZ), Division Molecular Genetics, Heidelberg, Germany [6]Institute of Human Genetics, Christian-Albrechts-University, Kiel, Germany [7]Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Essen, Germany [8]Department of Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany [9]BLUEPRINT project [10]Cell Networks, Bioquant, University of Heidelberg, Heidelberg, Germany [11]Structural Biology and BioComputing Programme, Spanish National Cancer Research Center (CNIO), Madrid, Spain [12]Deutsches Krebsforschungszentrum Heidelberg (DKFZ), Division Theoretical Bioinformatics, Heidelberg, Germany [13]Department of Otorhinolaryngology, University of Duisburg-Essen, Essen, Germany [14]University Hospital Muenster - Pediatric Hematology and Oncology, Münster Germany [15]Leibniz-Institut DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany [16]Department of Hematology and Oncology, Georg-Augusts-University of Göttingen, Göttingen, Germany [17]Institute of Pharmacy and Molecular Biotechnology, Bioquant, University of Heidelberg, Heidelberg, Germany [18]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK [19]Friedrich-Ebert Hospital Neumuenster, Clinics for Haematology, Oncology and Nephrology, Neumünster, Germany [20]Institute of Pathology, Charité – University Medicine Berlin, Berlin, Germany [21]Department of Internal Medicine II: Hematology and Oncology, University Medical Centre, Campus Kiel, Kiel, Germany [22]Radboud University, Department of Molecular Biology, Faculty of Science, Nijmegen, Netherlands [23]Hematopathology Section, Christian-Albrechts-University, Kiel, Germany [24]Institute for Medical Informatics Statistics and Epidemiology, University of Leipzig, Leipzig, Germany [25]Departamento de Anatomía Patológica, Farmacología y Microbiología, Universitat de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain [26]Institute of Pathology, Medical Faculty of the Ulm University, Ulm, Germany [27]University Hospital Giessen, Pediatric Hematology and Oncology, Giessen, Germany [28]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany [29]RNomics Group, Fraunhofer Institute for Cell Therapy and Immunology IZI, Leipzig, Germany [30]Santa

Fe Institute, Santa Fe, New Mexico, United States of America [31]Max-Planck-Institute for Mathematics in Sciences, Leipzig, Germany

Acknowledgements

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 (HEALTH-F5-2011-282510) project, and the KinderKrebsInitiative Buchholz/Holm-Seppensen and LIFE (Leipzig Research Center for Civilization Diseases), Leipzig University. LIFE is funded by the European Union, by the European Regional Development Fund (ERDF), the European Social Fund (ESF) and by the Free State of Saxony. Whole genome bisulfite sequencing additionally was supported by NGFNplus (BMBF #01GS0883) and the DKFZ-Heidelberg Center for Personalized Oncology (DKFZ-HIPO). Former grant support of the MMML by the Deutsche Krebshilfe (2003-2011) is gratefully acknowledged. J. Richter 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).

References

1. Lai AY, 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. 2013; 23:2030–41. [PubMed: 24013550]
2. Lee ST, 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. 2012; 40:11339–51. [PubMed: 23074194]
3. Shaknovich R, et al. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. Blood. 2011; 118:3559–69. [PubMed: 21828137]
4. Basso K, Dalla-Favera R. Germinal centres and B cell lymphomagenesis. Nat Rev Immunol. 2015; 15:172–84. [PubMed: 25712152]
5. Lenz G, Staudt LM. Aggressive lymphomas. N Engl J Med. 2010; 362:1417–29. [PubMed: 20393178]
6. Kuppers R, Dalla-Favera R. Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene. 2001; 20:5580–94. [PubMed: 11607811]
7. Dave SS, et al. Molecular diagnosis of Burkitt's lymphoma. N Engl J Med. 2006; 354:2431–42. [PubMed: 16760443]
8. Richter J, et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat Genet. 2012; 44:1316–20. [PubMed: 23143595]
9. Schmitz R, et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature. 2012; 490:116–20. [PubMed: 22885699]
10. Loeffler M, et al. Genomic and epigenomic co-evolution in follicular lymphomas. Leukemia. 2015; 29:456–63. [PubMed: 25027518]
11. Morin RD, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet. 2010; 42:181–5. [PubMed: 20081860]
12. Morin RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011; 476:298–303. [PubMed: 21796119]
13. Okosun J, et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. Nat Genet. 2014; 46:176–81. [PubMed: 24362818]
14. Pasqualucci L, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature. 2011; 471:189–95. [PubMed: 21390126]
15. Victora GD, et al. Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. Blood. 2012; 120:2240–8. [PubMed: 22740445]

16. Otto C, Stadler PF, Hoffmann S. Fast and sensitive mapping of bisulfite-treated sequencing data. Bioinformatics. 2012; 28:1698–704. [PubMed: 22581174]
17. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–9. [PubMed: 21441907]
18. Hahn MA, et al. Loss of the polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors. Cancer Res. 2014; 74:3617–29. [PubMed: 24786786]
19. Hovestadt V, et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature. 2014; 510:537–41. [PubMed: 24847876]
20. Weidensdorfer D, et al. Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. RNA. 2009; 15:104–15. [PubMed: 19029303]
21. Hummel M, et al. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. N Engl J Med. 2006; 354:2419–30. [PubMed: 16760442]
22. Love C, et al. The genetic landscape of mutations in Burkitt lymphoma. Nat Genet. 2012; 44:1321–5. [PubMed: 23143597]
23. Muppidi JR, et al. Loss of signalling via Galpha13 in germinal centre B-cell-derived lymphoma. Nature. 2014
24. Rohde M, et al. Recurrent RHOA mutations in pediatric Burkitt lymphoma treated according to the NHL-BFM protocols. Genes Chromosomes Cancer. 2014; 53:911–6. [PubMed: 25044415]
25. O'Hayre M, et al. The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. Nat Rev Cancer. 2013; 13:412–24. [PubMed: 23640210]
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. 2011; 1:460–81. [PubMed: 21984966]
27. Morin RD, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood. 2013; 122:1256–65. [PubMed: 23699601]
28. Abraham BJ, Cui K, Tang Q, Zhao K. Dynamic regulation of epigenomic landscapes during hematopoiesis. BMC Genomics. 2013; 14:193. [PubMed: 23510235]
29. Scott CL, et al. Role of the chromobox protein CBX7 in lymphomagenesis. Proc Natl Acad Sci U S A. 2007; 104:5389–94. [PubMed: 17374722]
30. Dykhuizen EC, et al. BAF complexes facilitate decatenation of DNA by topoisomerase IIalpha. Nature. 2013; 497:624–7. [PubMed: 23698369]
31. Klapper W, et al. Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma. Blood. 2012; 119:1882–7. [PubMed: 22238326]
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. 2011; 35:933–5. [PubMed: 21566516]
33. Hasselblatt M, et al. SMARCA4-mutated atypical teratoid/rhabdoid tumors are associated with inherited germline alterations and poor prognosis. Acta Neuropathol. 2014; 128:453–6. [PubMed: 25060813]
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. 2010; 86:279–84. [PubMed: 20137775]
35. Witkowski L, et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. Nat Genet. 2014; 46:438–43. [PubMed: 24658002]
36. Betts MJ, et al. Mechismo: predicting the mechanistic impact of mutations and modifications on molecular interactions. Nucleic Acids Res. 2014
37. Kulis M, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. Nat Genet. 2015; 47:746–56. [PubMed: 26053498]
38. Lister R, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011; 471:68–73. [PubMed: 21289626]
39. Bibikova M, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011; 98(4):288–95. [PubMed: 21839163]

40. Li H, Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. [PubMed: 19451168]
41. Jones DT, et al. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. Nat Genet. 2013; 45(8):927–32. [PubMed: 23817572]
42. Rimmer A, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014; 46(8):912–8. [PubMed: 25017105]
43. Hoffmann S, et al. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS computational biology. 2009; 5(9)
44. Otto C, Stadler Peter F, Hoffmann S. Fast and sensitive mapping of bisulfitetreated sequencing data. Bioinformatics. 2012; 28(13):1698–1704. [PubMed: 22581174]
45. R Core Team R. A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. URL http://www.R-project.org/
46. Zhang, Nancy R., Siegmund, David O. Model selection for high-dimensional, multisequence change-point problems. Statistica Sinica. 2012; 22.4:1507.
47. Anders S, Huber W. Differential expression analysis for sequence count data. Genome biol. 2010; 11(10):R106. [PubMed: 20979621]
48. Karolchik D, et al. The UCSC Genome Browser database: 2014 update. Nucleic Acids Res. 2014
49. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74. [PubMed: 22955616]
50. Flicek P, et al. Ensembl 2014. Nucleic Acids Res. 2014
51. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature methods. 2012; 9(3):215–216. [PubMed: 22373907]
52. Dürr H, et al. X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. Cell. 2005; 121(3):363–73. [PubMed: 15882619]
53. Eswar N, et al. Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics. 2006:5.6.
54. Sharma V, et al. Crystal structure of Mycobacterium tuberculosis SecA, a preprotein translocating ATPase. Proc Natl Acad Sci U S A. 2003; 100(5):2243–8. [PubMed: 12606717]
55. Huber W, et al. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 2002; 18:S96–S104. [PubMed: 12169536]
56. Bentink S, et al. Pathway activation patterns in diffuse large B-cell lymphomas. Leukemia. 2008; 22(9):1746–54. [PubMed: 18580954]
57. 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. 2003; 100(17):9991–6. [PubMed: 12900505]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

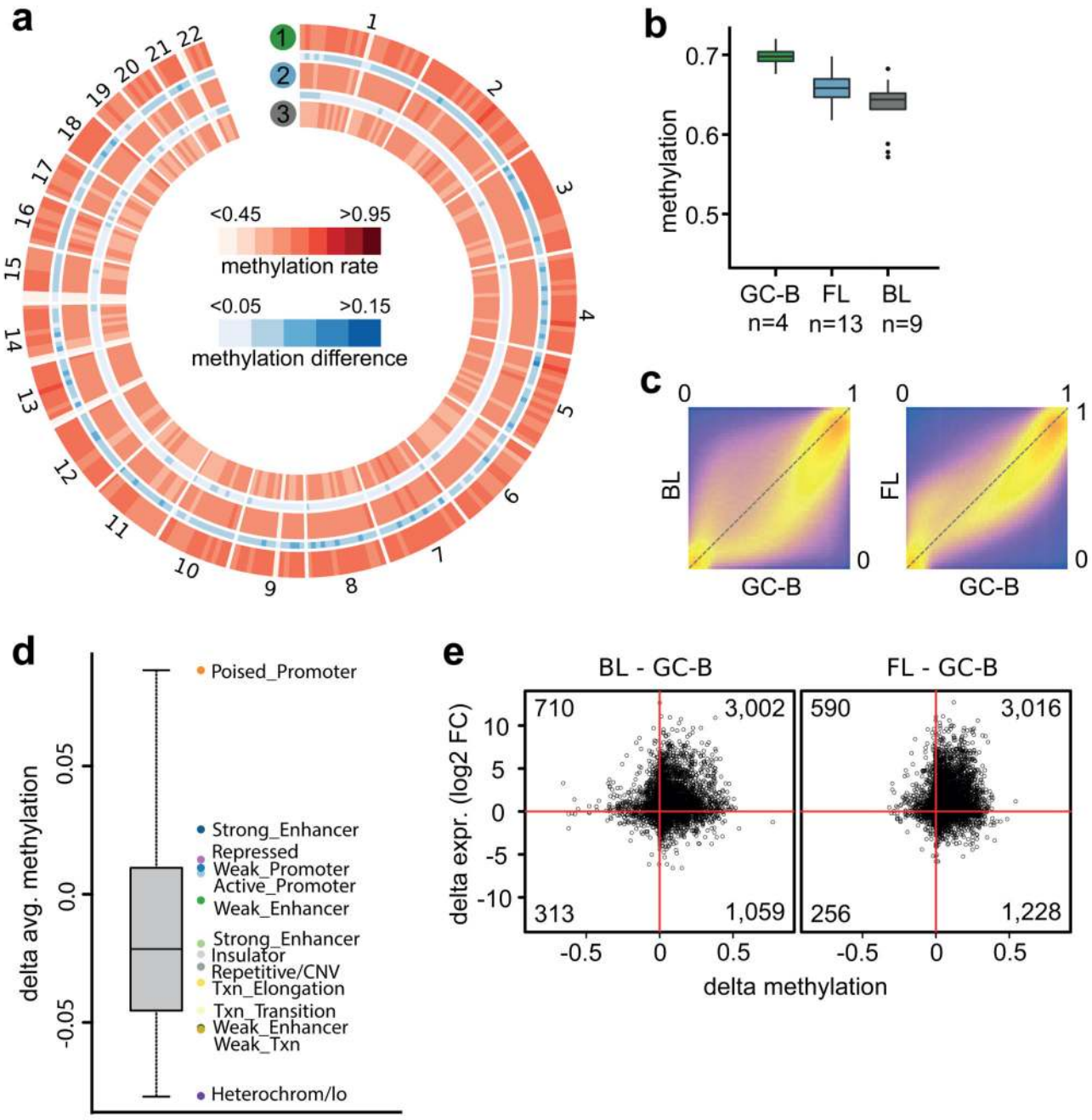

Figure 1. Loss of methylation in lymphoma.

(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-cells, n=4; 2: follicular lymphoma, n=9; 3: Burkitt lymphoma, n=13). **(b)** Boxplot of average methylation levels for lymphoma and normal controls. Methylation differences are significant (<0.05) for all pairwise comparisons except for follicular lymphoma vs. Burkitt lymphoma (p=0.13). Boxes indicate medians, 25th and 27th percentiles. Whiskers indicate the 1.5-fold inter-quartile ranges. GC-B, germinal center B-cells; BL, Burkitt

lymphoma; FL; follicular lymphoma. **(c)** Correlation of DNA methylation levels in Burkitt lymphoma, follicular lymphoma and germinal center B-cells at CpG resolution (blue=low density, orange=high density). **(d)** Average methylation difference (average of Burkitt lymphoma and follicular lymphoma vs. germinal center B-cells) in GM12878 chromatin segments. **(e)** Scatter plots of DNA methylation of poised promoters versus RNA expression of the associated genes.

Europe PMC Funders Author Manuscripts

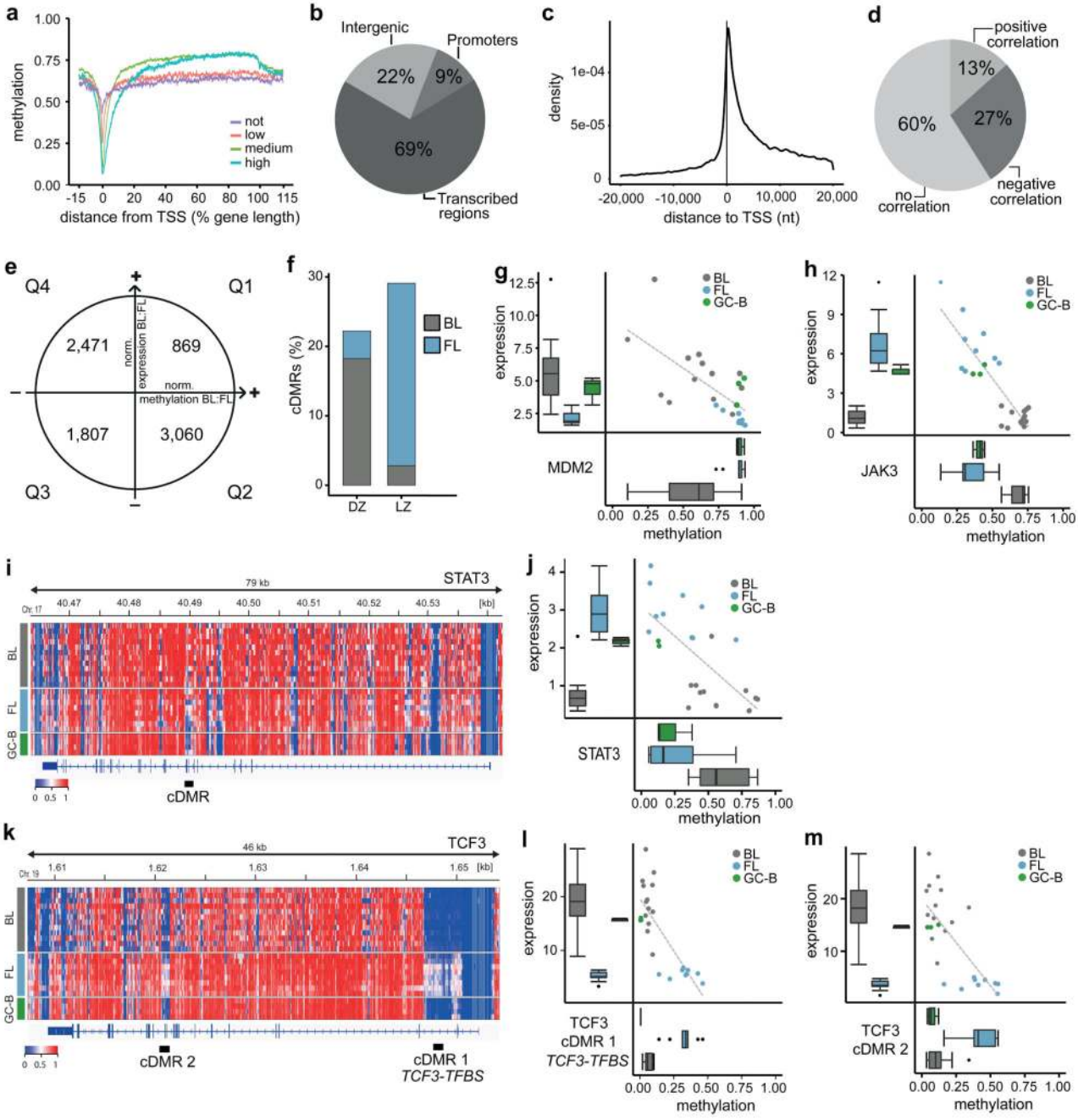

Figure 2. Differentially methylated regions (DMRs).

(a) Interdependence of DNA methylation and RNA expression levels across gene regions (mean of germinal center B-cells, follicular lymphoma and Burkitt lymphoma is shown). Genes were grouped by expression level (not: RPKM<0.1; low: 0.1<=RPKM<1; medium: 1<=RPKM<10; high: 10<=RPKM) **(b)** Proportion of DMRs in promoters, transcribed- 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 upstreams of TSS to TES) that showed no, positive or negative correlations of DMR methylation and expression of the

Europe PMC Funders Author Manuscripts

associated gene. **(e)** Types and distribution of correlating DMRs (cDMRs). For about one-third of DMRs (n=8,207; upper panel) degrees of methylation and RNA expression of associated genes correlated ($p<0.05$; Spearman). Of these cDMR-gene pairs, 67% showed negative correlation (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)** Scatter and box plots of correlating DMRs of the DZ and LZ-associated genes *MDM1* and *JAK3*, respectively. (**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)** Correlations of cDMR methylation and expression for *TCF3*. GC-B, germinal center B-cells; BL, Burkitt lymphoma; FL; follicular lymphoma.

Europe PMC Funders Author Manuscripts

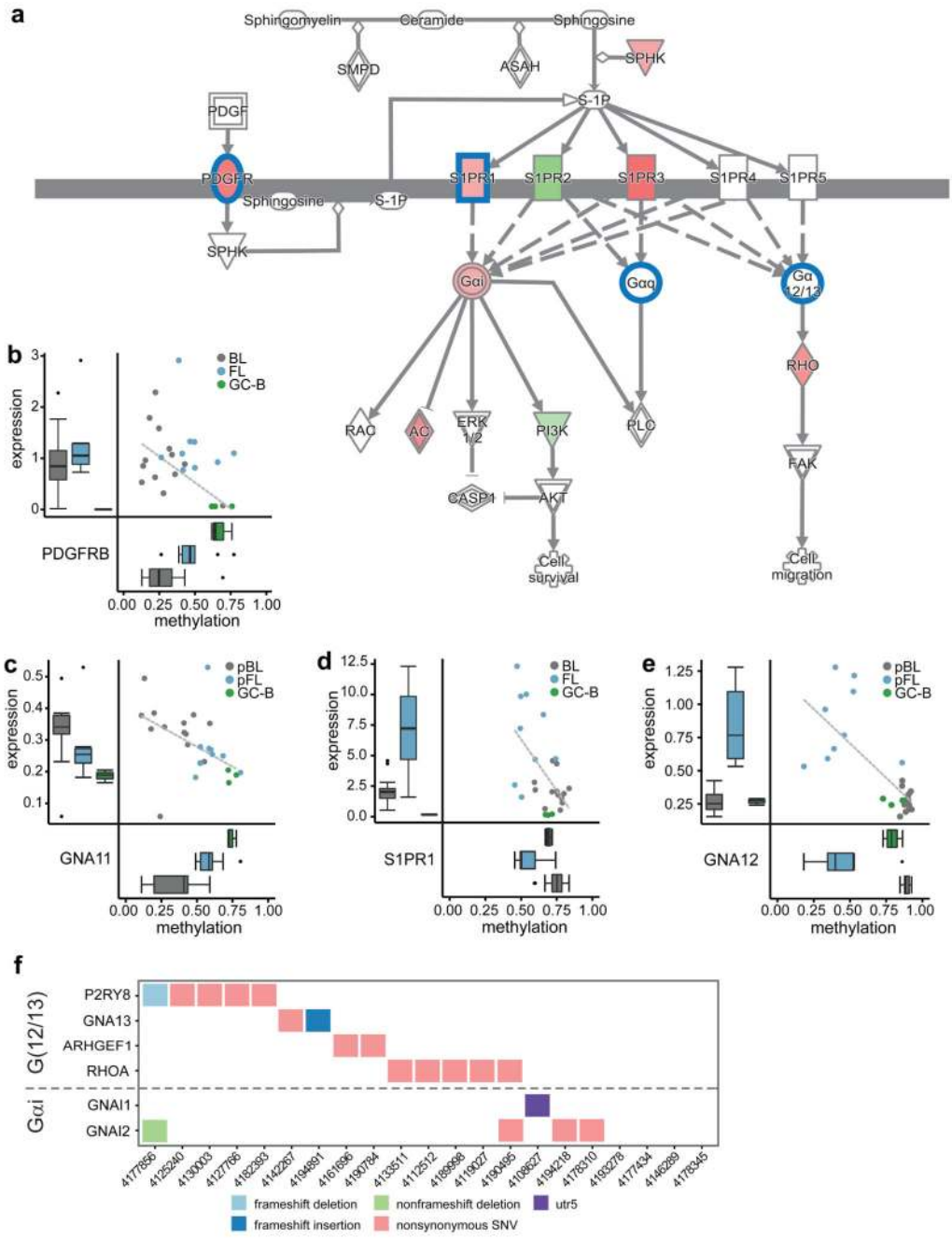

Figure 3. Sphingosine-1-phosphate signaling is affected by complementary DNA mutation and methylation in germinal center B-lymphomas.

(a) Pathway of sphingosine-1-phosphate and related G-protein-coupled signaling. RNA expression of genes differentially expressed between lymphoma (Burkitt lymphoma and follicular lymphoma) and germinal center B-cells are indicated in red and green for up- and downregulation, respectively. Genes marked with blue outlines are associated with cDMRs. **(b-e)** Scatter and box plots of negatively correlating DMR methylation and RNA expression affecting genes involved in sphingosine-1-phosphate and G-protein signaling. **(b)** *PDGFRB*

(c) *GNA11* **(d)** *S1PR1* **(e)** *GNA12* **(f)** Mutations identified in Burkitt lymphoma samples affecting the G12/13 (above dashed line) and Gαi (below dashed line) complexes. GC-B, germinal center B-cells; BL, Burkitt lymphoma; FL, follicular lymphoma.

Europe PMC Funders Author Manuscripts

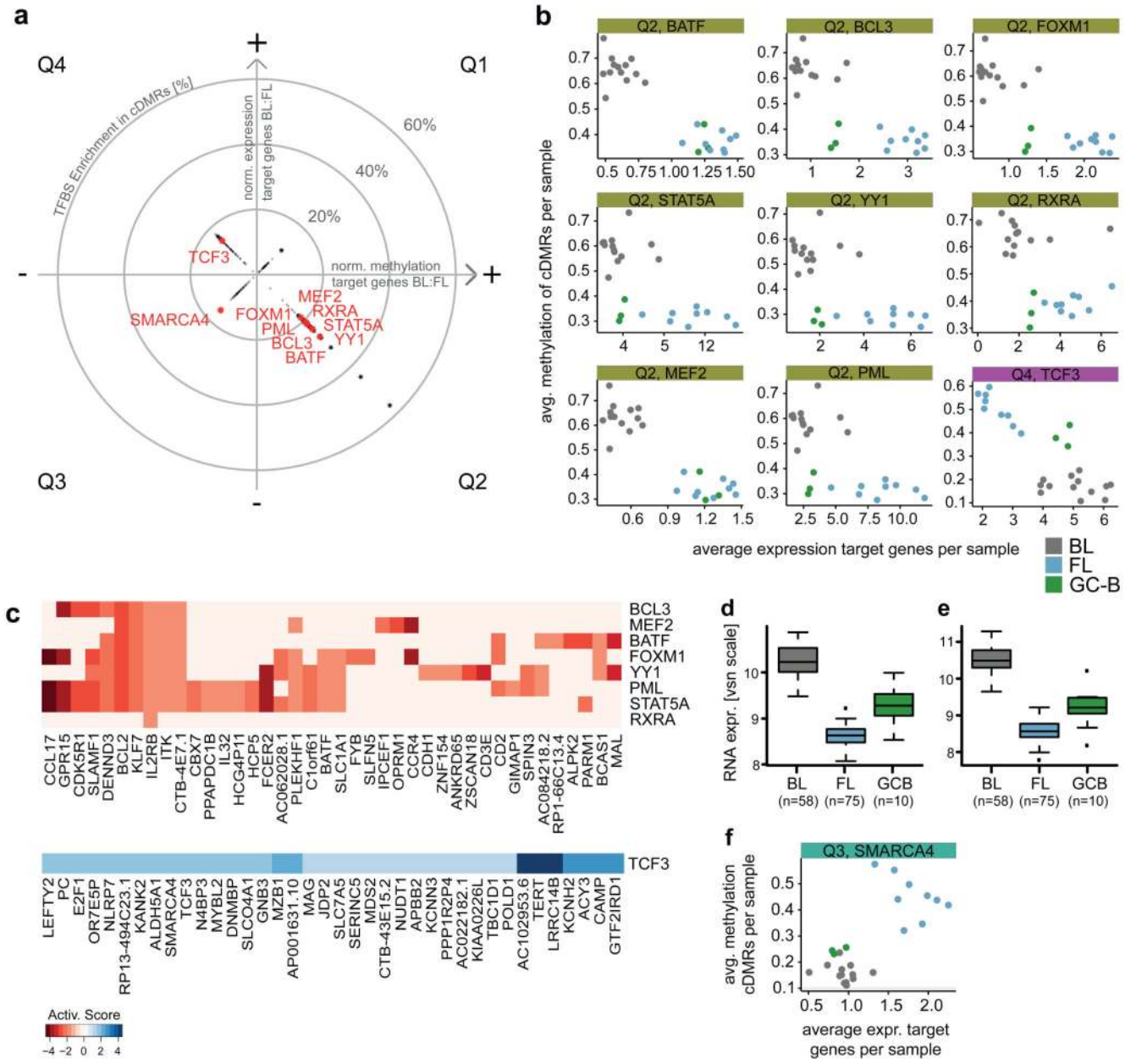

Figure 4. Enrichment of transcription factor binding sites (TFBS) in cDMRs.

(a) Radar plot showing the enrichment of binding sites of 46 transcription factors in cDMRs. Quadrants (Q1-Q4) classify cDMRs by correlation type (positive or negative) and direction of methylation and expression in Burkitt lymphoma vs. follicular lymphoma. Negatively and positively correlating DMR-gene pairs are located in Q2/Q4, and Q1/Q3, respectively. Concentric circles indicate levels of TFBS enrichment measured as percent binding sites of a particular TF found in cDMRs relative to all binding sites of this TF found in DMRs. Stars indicate TFBS that are significantly enriched (permutation test) in cDMRs. Red stars indicate the 10 TFs showing the best correlation of TF and average target gene expression. **(b)** Correlations of sample-wise average cDMR methylation and average target gene

Europe PMC Funders Author Manuscripts

expression for the 9 of the top 10 TFs showing negative correlations. **(c)** Activity map integrating the correlation of TFBS methylation and differential expression of target genes. Color and hue indicate the strength of inactivation (red) or activation (blue) in Burkitt lymphoma compared to follicular lymphoma samples. The top 40 target genes are shown for quadrants Q2 and Q4, each. **(d,e)** RNA expression of *TCF3* **(d)** and *SMARCA4* **(e)** in Burkitt lymphoma, follicular lymphoma and germinal center B-cells determined by microarray analysis in an extended lymphoma cohort. **(f)** Correlations of sample-wise average cDMR methylation and average target gene expression for *SMARCA4*. GC-B, germinal center B-cells; BL, Burkitt lymphoma; FL; follicular lymphoma.

Europe PMC Funders Author Manuscripts

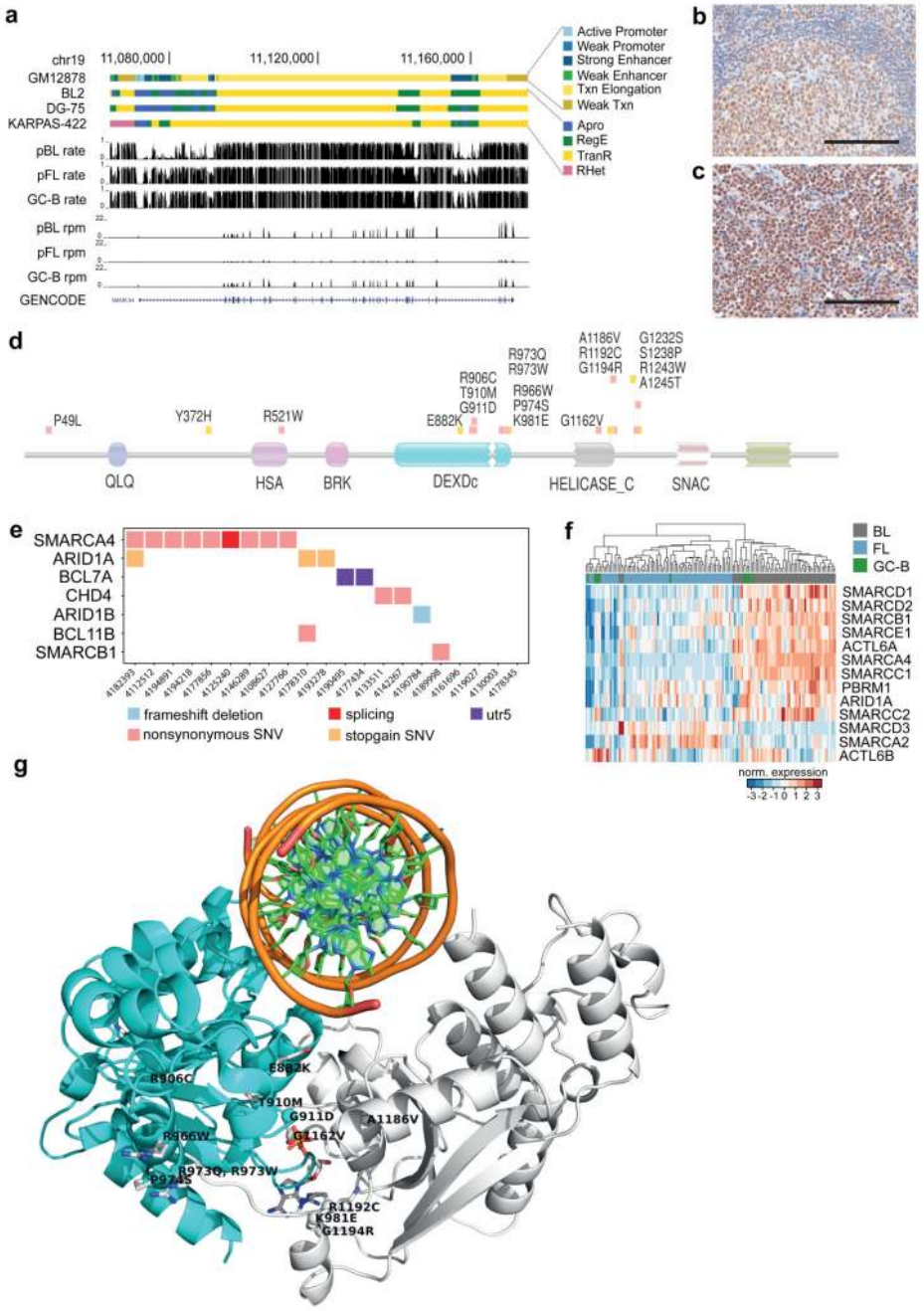

Figure 5. *SMARCA4* genome architecture and protein expression.

(a) Genome browser view of the *SMARCA4* locus. Chromatin segmentations of GM12878, BL2, DG-75 and KARPAS-422 cell lines (top), and average CpG methylation (middle) and RNA expression (bottom) of Burkitt lymphoma, follicular lymphoma and germinal center B-cell samples. **(b,c)** Staining of SMARCA4 (BRG1) by immunohistochemistry in **(b)** a normal reactive follicle (tonsil), and **(c)** a Burkitt lymphoma with SMARCA4 R973W mutation (infiltrated lymph node; original magnification 400x; scale bar = 200 μm). **(d)** Accumulation of missense mutations in the *SMARCA4* DEXDc helicase (cyan) and helicase

Europe PMC Funders Author Manuscripts

conserved C-terminal (grey) domains in Burkitt lymphoma samples. Heights of lollipops indicate relative numbers of mutations at that position. Red squares indicate severe changes in physico-chemical properties (score<0); yellow: 0<score<4. **(e)** Distribution of somatic mutations (SNVs or indels) of the SWI/SNF complex in Burkitt lymphomasuggests mutually exclusive mechanisms of inactivation. **(f)** Unsupervised analysis of SWI/SNF complex RNA-expression data (Affymetrix hgu133a) segregates Burkitt lymphoma and follicular lymphoma. **(g)** Model of *SMARCA4* interaction with a DNA helix. Amino acid residues mutated in Burkitt lymphoma are indicated. but the model suggests them to 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; grey, C-terminal helicase domain. ATP is shown with sticks colored by atom types. GC-B, germinal center B-cells; BL, Burkitt lymphoma; FL; follicular lymphoma.