Enhancer Sequence Variants and Transcription-Factor Deregulation Synergize to Construct Pathogenic Regulatory Circuits in B-Cell Lymphoma

Graphical Abstract

Highlights

- Complete gene-regulatory circuits for human B-cell lymphoma and normal GC B subsets
- Two cohorts of transcription factors target pirated versus lost enhancers in lymphoma
- Pathogenic enhancers enriched for SNPs and somatic mutations affect factor binding
- Epigenome-centric analysis identifies two new subtypes of follicular lymphoma

Authors

Olivia I. Koues, Rodney A. Kowalewski, ..., Jacqueline E. Payton, Eugene M. Oltz

Correspondence

jpayton@wustl.edu (J.E.P.), eoltz@wustl.edu (E.M.O.)

In Brief

A fundamental focus of lymphoma biology remains the identification of cooperative mechanisms that promote oncogenesis. Oltz and colleagues define the pathogenic circuitry of human follicular lymphoma (FL), revealing distinct genetic and epigenetic etiologies for germinal-center B-cell transformation.

Koues et al., 2015, Immunity 42, 186–198

January 20, 2015 ©2015 Elsevier Inc.

http://dx.doi.org/10.1016/j.immuni.2014.12.021
Enhancer Sequence Variants and Transcription-Factor Deregulation Synergize to Construct Pathogenic Regulatory Circuits in B-Cell Lymphoma

Olivia I. Koues,1 Rodney A. Kowalewski,1 Li-Wei Chang,1 Sarah C. Pyfrom,1 Jennifer A. Schmidt,1 Hong Luo,1 Luis E. Sandoval,1 Tyler B. Hughes,1 Jeffrey J. Bednarski,2 Amanda F. Cashen,3 Jacqueline E. Payton,1,* and Eugene M. Oltz1,*

1Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA
2Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA
3Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA
*Correspondence: jpayton@wustl.edu (J.E.P.), eoltz@wustl.edu (E.M.O.)
http://dx.doi.org/10.1016/j.immuni.2014.12.021

SUMMARY

Most B-cell lymphomas arise in the germinal center (GC), where humoral immune responses evolve from potentially oncogenic cycles of mutation, proliferation, and clonal selection. Although lymphoma gene expression diverges significantly from GC B cells, underlying mechanisms that alter the activities of corresponding regulatory elements (REs) remain elusive. Here we define the complete pathogenic circuitry of human follicular lymphoma (FL), which activates or decommissions REs from normal GC B cells and commandeers enhancers from other lineages. Moreover, independent sets of transcription factors, whose expression was deregulated in FL, targeted commandeered versus decommissioned REs. Our approach revealed two distinct subtypes of low-grade FL, whose pathogenic circuitries resembled GC B or activated B cells. FL-altered enhancers also were enriched for sequence variants, including somatic mutations, which disrupt transcription-factor binding and expression of circuit-linked genes. Thus, the pathogenic regulatory circuitry of FL reveals distinct genetic and epigenetic etiologies for GC B-cell transformation.

INTRODUCTION

B-cell lymphoma is one of the most common human cancers. Its prevalence is a byproduct of the complex cellular and molecular processes that tailor humoral immune responses to antigens in the germinal center (GC). In secondary lymphoid organs, GC B cells undergo clonal expansion while activation-induced cytidine deaminase (AICDA) targets their genomes for DNA damage-associated alterations. In the GC dark zone, proliferative centroblasts (CBs) perform somatic hypermutation (SHM) to fine-tune the affinity of immunoglobulin (Ig) variable regions for their cognate antigens. Some of the CBs then migrate to the light zone, where they become non-cycling centrocytes (non-cycling CCs), are selected for antigen affinity, and undergo class-switch recombination (Allen et al., 2007). Recent studies indicate that, rather than representing two distinct stages of differentiation, CCs and CBs are alternate activation states of GC B cells (Victora et al., 2012).

Although essential for optimization of humoral responses, the GC reaction involves coincident genome damage and rapid proliferation that increase the risk of the oncogenic lesions that underlie most types of B-cell lymphoma (Rui et al., 2011). For example, follicular lymphoma (FL), a common form of non-Hodgkin lymphoma (NHL), is thought to arise from CCs (Victora et al., 2012). The pathologic hallmark of FL is an AICDA-mediated t(14;18) [IgH-BCL2] translocation and causes overexpression of the anti-apoptotic BCL2 protein (Leich et al., 2011). However, this primary genetic lesion is insufficient to initiate FL. Indeed, BCL2 transgenic mice infrequently develop FL, and t(14;18)-bearing B cells are found in healthy individuals who do not develop lymphoma (Cheung et al., 2009; Dööken et al., 1996). A fundamental focus of lymphoma biology remains the identification of cooperative mechanisms that promote oncogenesis.

One clear outcome of these cooperative mechanisms is a substantial revision of normal GC expression programs during transformation of GC cells to NHL (Klein and Dalla-Favera, 2008), which culminates from a combination of genetic and epigenetic changes. First, compared with GC B cells, NHL samples have altered expression of at least some transcription factors (TFs) that are essential to the GC reaction (Alizadeh et al., 2000). Second, NHL samples have mutations in histone modifiers, leading to altered chromatin landscapes (Morin et al., 2011; Yap et al., 2011). These epigenomic perturbations might coordinately silence large cohorts of genes (e.g., tumor suppressors) while activating genes involved in proliferation (Baylin and Jones, 2013). Third, many diseases are commonly linked to genetic or epigenetic differences in non-coding sequences that encompass transcriptional regulatory elements (REs) (Akhtar-Zaidi et al., 2012; Maurano et al., 2012). This shared genetic–epigenetic etiology may be especially relevant to NHL because of its inherent genome instability and dysregulation of histone modifiers, both of which could impact RE function.

To understand the oncogenic processes that sculpt NHL transcriptomes, it is critical to identify altered distal REs (DREs) and
assign them to target genes, thereby defining pathogenic circuitry that drives cellular transformation. Analyses of chromatin modification patterns have linked some DREs with their target promoters in normal cells, revealing aspects of lineage-specific regulatory circuits (Gerstein et al., 2012; Maurano et al., 2012). To decipher oncogenic changes to regulatory circuits, one should compare purified malignant cells from primary tumors with normal counterparts rather than using unsorted biopsy samples containing a mixture of cell types, or cancer lines, which accumulate widespread revisions to the genome and epigenome during long-term culture (Masters, 2000; Victora et al., 2012). To circumvent these obstacles, we focused on FL, an indolent B-cell malignancy that is incurable and often transforms to a more aggressive lymphoma (Lenz and Staudt, 2010).

We report an integrative analysis of -omics data from purified human FL B cells and their normal GC B counterparts, exposing the oncogenic regulatory circuitry of this cancer. The FL circuitry is a composite of enhancers that are engaged normally in CC or CB subsets but are chronically activated or attenuated in FL, as well as DREs usurped from a wide range of other, non-B lineages. Importantly, augmented or attenuated DREs are each targeted by a distinct set of TFs that exhibit altered expression in FL. Our approach also revealed two previously unappreciated subtypes of low-grade FL; these subtypes have distinct expression profiles and corresponding sets of aberrant DREs. The FL-altered DREs are also enriched for disease-associated single-nucleotide polymorphisms (SNPs) and, remarkably, for somatic mutations. Indeed, several of the variants, located in attenuated DREs, impair binding of TFs that play key roles in GC reactions and correspond to reductions in target gene expression. These discoveries establish the epigenetic and genetic etiologies that mediate transformation of normal GC B cells into a common cancer, and they pave the way for personalized epigenetic therapies that target subtype-specific control elements.

**RESULTS**

**The FL Regulome Reflects Its Centricyte Origins**

Rewiring of gene expression in cancer evolves from a series of genetic and epigenetic changes that impinge on the activities of transcriptional promoters and their DREs. Deciphering how these processes drive cancer requires characterization of regulomes in purified malignant cells and comparison of these cells with their normal counterparts. Thus, we purified malignant B cells from lymph-node biopsies of 18 FL patients (Table S1). Control B cells were purified from the peripheral blood of healthy individuals (these cells are referred to as PBPs), FL patients, and excised tonsils (the latter are referred to as TsBs). Chromatin, RNA, and genomic DNA harvested from these samples were analyzed by ChiP-seq, SNP arrays, and either RNA-seq or expression arrays, respectively. We used FAIRE-seq to identify putative cis elements in each sample (Giresi and Lieb, 2009) and analyzed hallmark chromatin features (Bernstein et al., 2012), including H3K27ac and H3ac, to assess relative RE activities among B-cell populations. To standardize comparisons between samples, we normalized the data for differences in read depth and merged overlapping peaks to generate a consolidated list of putative REs.

On the basis of gene expression profiles, cells of FL and diffuse large B-cell lymphoma (DLBCL) most closely resemble GC B cells (Victora et al., 2012). We purified CCs and CBs from human tonsils by using flow cytometry and the following surface markers: CD19^CD10^CD44^CXCR4^ (CBs) and CD19^CD10^CD44^CXCR4^ (CCs). Successful separation was confirmed by comparison of the expression of genes known to discriminate CCs from CBs (Figure S1A). Importantly, the close relationship between CCs and CBs, compared with other hematopoietic lineages, is evident from clustering analyses of open chromatin regions (Figure S1B), underscoring the strong functional links between regulomes and transcriptomes.

To determine the most appropriate control cell type for defining pathogenic FL circuitry, we first compared expression of genes that distinguish CCs from CBs (Victora et al., 2012). Consistent with a previous study, expression of these genes in FL is similar to CC expression and diverges almost completely from CB expression (Figure 1A). In contrast, cultured NHL cell lines resemble the highly proliferative CB subset. Epigenomic analyses complement these transcriptome comparisons; the majority of REs in FL are shared with GC B cells, though a greater proportion overlap with CCs than with CBs (Figure 1B). Together, these data support the notion that FL arises from CCs (Victora et al., 2012) and that comparison to CCs should be used for identification of REs with altered function in FL.

**Defining the FL-Altered Regulome**

Pathogenic gene expression in many diseases, including cancer, is caused by changes in DRE activity. These changes in turn alter promoter function (Akhtar-Zaidi et al., 2012). Accordingly, we set out to identify DREs whose activities in FL samples were augmented or attenuated in comparison to those of normal CCs. Although FAIRE-seq data mark DRE locations in the genome, both poised and active cis elements emerge as peaks. To measure the relative activity of DREs, we incorporated ChiP-seq data for H3K27ac and H3ac, which mark active DREs. For example, nearly identical FAIRE peaks are detected in both CC and FL samples for a cis element located near CXCR4 (Figure 1C), which encodes a chemokine receptor required for light- (CC) and dark (CB)-zone organization (Allen et al., 2007). This region is substantially enriched for H3K27ac in FL samples, indicating augmented activity. Indeed, FL samples also exhibited higher CXCR4 expression than CC controls. A global view of H3K27ac levels at DREs in FL and GC B cells segregates these regions into multiple categories on the basis of relative activity, including FL hyper-activation or attenuation of DREs characteristic of CCs, CBs, or neither of these closely related subsets (Figure 1D).

To define the FL dysregulome, we assigned variable DREs as regions for which FAIRE, H3K27ac, or H3ac intensity for an individual lymphoma differ more than two-fold from averaged values for the same regions in CC samples. As expected from similarities between FL cells and CCs, the majority of REs remain unchanged in lymphoma B cells, and the remainder exhibit altered signal intensity (15K–30K per FL sample). In nearly all FL samples, more DREs are attenuated, though a substantial proportion exhibit enhanced activity (25%–55%) (Figure 1E). Epigenetic alteration of DREs cannot be attributed exclusively to mutations in chromatin modifiers because these occurred in only a subset of samples (Table S2). As shown in Figure S1C, H3K27ac peak
intensity in CC control samples is highly reproducible; less than 5% of DREs are variable in CC samples, whereas 45%–50% of these elements are variable in FL samples (Figure S1D). Augmented or attenuated DRE intensities were also similar between FL and patient-matched PBBs or any of the CC controls (Figure S1E). Underscoring their relevance to common pathways of lymphomagenesis, the majority of variable DREs are recurrent, and more than 80% were identified in at least two FL samples (Figure 1F).

The genomic location of variable DREs—in distal, non-promoter regions—along with their epigenetic profiles (FAIRE+ H3K27ac+) indicates enhancer function. Indeed, luciferase reporter assays revealed enhancer activity for 7/10 DREs in cultured lymphoma cells (Figure S1F). A subtype of DRE, termed super-enhancers (SEs), are composed of large RE clusters and drive the expression of lineage-specifying genes (Hnisz et al., 2013; Whyte et al., 2013). In addition, SEs might contribute to pathologic gene expression in DLBCL (Chapuy et al., 2013). To evaluate the variability of SEs in FL samples versus CCs, we identified these elements by their unusually high H3K27ac density (Figure S1G and Table S3) (Whyte et al., 2013). Consistent with previous reports, genes within 500 kb of SEs were expressed more robustly than those near conventional enhancers (Figure S1H). Using our criteria, one can categorize a subset

Figure 1. The Centrocyte Origins and Alterations to the FL Regulome

(A) Expression profiles of FL and NHL cell lines for a panel of genes differentially expressed in CBs versus CCs.
(B) Bar graph showing unique and shared FAIRE peaks in FL and GC B-cell populations.
(C) UCSC Genome Browser views of FAIRE-seq, H3K27ac ChIP-seq and RNA-seq data from FL and CC samples, illustrating a collection of DREs located near CXCR4. FAIRE and ChIP data are presented as the number of reads per million mapped reads and are plotted on an axis of 1–25 (FAIRE) and 1–90 (H3K27ac). RNA data are presented as the number of aligned, in silico extended reads per 10 bp on a scale of 1–400 reads.
(D) H3K27ac ChIP-seq intensities for DREs in representative CB, CC, and FL samples. Data are presented as k-means clustering of tag densities per 200 bp within a window of 10 kb around the DREs.
(E) Percent of variable DREs with a 2-fold or greater increase (augmented) or decrease (attenuated) in FAIRE-seq, H3ac ChIP-seq, or H3K27ac ChIP-seq signal for FL samples relative to CC samples.
(F) Recurrence rates of variable DREs are depicted by the proportion detected in a certain number of FL samples.
of stringency: concordant changes in the expression of genes predicted to be targets of DREs by chromatin profiling (Figure S2A). Only 53% of putative target genes were verified by this expression filter (Table S4), indicating a 47% false-positive rate for predicting DRE-promoter connections solely via chromatin patterns. Approximately 40% of the circuits with altered activity in FL samples are normally employed in the GC reaction; these include CC- or CB-specific connections, as well as those used by both subsets (Figure S2C). However, the majority of altered FL circuits are not engaged in either subset of GC B lymphocytes, indicating a widespread change in the B-cell regulome.

The validity of connections identified by this strategy is highlighted by the cytokine receptor locus that includes IFNAR1, IFNAR2, IL10RB, and IFNGR2 (Figure 2C). Activation of these receptors alters B-cell responses, including viability and proliferation, which are enhanced. However, to our knowledge, overexpression of these genes has not been linked to FL. The pathogenic circuits for each of the two augmented DREs at the 3’ end of this cluster include predicted regulatory interactions with multiple TSSs of the receptor genes. Published 5C data for the B lymphoblastoid cell line GM12878 confirms interactions between these DREs and restriction fragments near the target-gene TSSs (Sanyal et al., 2012). Thus, integrative analysis of transcriptome and epigenome data reveals a core pathogenic FL circuitry that incorporates regulome components from both GC subsets. The newly defined circuitry also identifies variable DRE connections to genes whose roles in FL pathogenesis were previously unknown.

Figure 2. Pathogenic Circuitry of FL
(A) Number of TSSs for which chromatin alterations are concordant with nearby variable DREs (within 500 kb).
(B) Mean transcript abundance as determined by RNA-Seq for genes linked to augmented, unchanged, or attenuated DREs and located within the distances shown from the DREs. Statistical significance (Mann-Whitney test): *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.0001. Results represent the mean ± SEM of ten FL samples.
(C) UCSC Genome Browser views of H3K27ac ChIP-seq data, showing augmented DREs (highlighted in blue). RNA-seq data depict the corresponding upregulation of mRNA from DRE target genes. The bottom track shows significant spatial interactions in GM12878 B cells (ENCOD, 5C data) between restriction fragments encompassing the DREs and restriction fragments near the target-gene TSSs (Sanyal et al., 2012).

Pathogenic FL Regulatory Circuits
To decipher the pathogenic regulatory circuitry of FL, one must connect variable DREs to their target genes. Chromatin accessibility at enhancers and nearby promoters often correlate, providing a pattern-based strategy by which to assign regulatory circuits (Thurman et al., 2012). Using this approach, we connected variable DREs to genes within 500 kb that exhibit concordant changes in chromatin at their transcription start sites (TSSs) (Figure S2A). Approximately one-third of the variable DREs were linked to a single TSS, whereas the majority had potential connections to two or more target genes (Figure 2A). Moreover, the average expression of genes with concordant TSS chromatin is significantly elevated or attenuated in comparison to that of genes near unchanged DREs (Figure 2B) or to that of neighboring genes (Figure S2B).

Nevertheless, chromatin-based correlations do not allow evaluation of the functional output of putative DRE-gene connections, i.e., expression levels. Therefore, we applied a new level
Epigenome-Centric Analysis Segregates FL into Two Distinct Subtypes

Despite significant heterogeneity in clinical behavior and therapeutic responses, low-grade FL is uniformly classified for prognosis and treatment (Izutsu, 2014). One source of FL heterogeneity could be utilization of distinct B-cell regulatory circuits during oncogenesis. To test this, we performed unbiased hierarchical clustering of chromatin features for FL and normal B-cell populations, including CCs, CBs, resting PBBs, and in-vitro-activated PBBs, which resemble plasmablasts (Tarte et al., 2002). Unexpectedly, this epigenomic analysis revealed two distinct FL DRE subtypes that differ significantly (p < 0.01). Whereas the subtype 1 enhancer profile parallels that of CC, the subtype 2 profile includes DREs with activities similar to those of in-vitro-activated PBBs (Figure 3A). Thus, enhancer profiling identifies two distinct subtypes of FL that resemble different stages of B-cell activation and differentiation, much like previous expression-profiling studies have segregated subtypes of DLBCL (Alizadeh et al., 2000).

To gain insight into the genes regulated by subtype 1 and 2 DREs, we focused on a cohort shown previously to distinguish GC B-cell populations from other mature B-cell populations (Longo et al., 2009). Strikingly, subtype-specific DREs form regulatory circuits with >95% of these genes. Figure 3B shows expression data for genes linked to subtype-specific DREs, confirming that components of CC and plasmablast transcriptional programs are differentially engaged by enhancers specific to subtypes 1 and 2, respectively. Importantly, when we focused on these genes defined by our epigenome-centric approach, the divergent expression patterns were also evident in microarray data from larger FL sample sets (Figure S3A) (Compagno et al., 2009; Dave et al., 2004). For example, a subtype 2-specific circuit connects a variable DRE with enhanced expression of CCR7. CCR7 encodes the CCL19 and CCL21 chemokine receptor (Figure S3B), which is involved in NHL retention within the lymph node and thus provides transformed cells with a survival advantage (Rehm et al., 2011). The two FL subtypes also have remarkably different expression patterns for genes involved in a range of relevant biological processes, including DNA repair and NF-κB signaling pathways (Figure S3C).

Although sample numbers are limited, we observed no association of either subtype with patient age, gender, tumor purity, grade, or presence of BCL2 translocations (Tables S1 and S5). However, patients in subtype 2 were more likely to have received chemotherapy for lymphoma than were patients in subtype 1 (3/6 versus 2/10, respectively). Our epigenome-centric approach uncovered two subtypes of low-grade FL characterized by distinct enhancer profiles and linked transcriptional circuitries, suggesting divergent modes of pathogenesis.

FL Discards Non-Essential Components of GC Circuitry

Humoral immune responses in the GC are evolutionary processes that incorporate molecular and cellular mechanisms unique to B lymphocytes; such mechanisms include Ig isotype switching, somatic mutation, and B-cell receptor signaling (Vic- tora and Nussenzweig, 2012). Many of these antigen-induced responses may be dispensable or deleterious for FL cell survival. Indeed, more than half of the DREs identified in CCs are attenuated in at least one FL sample (Figure 1E). When genes connected to attenuated DREs were subjected to ontology analysis, we found enriched pathways that are most likely detrimental to lymphomagenesis; such pathways include cell-cycle checkpoints and apoptosis (Figure 4A). Many pathways critical for the GC reaction, but presumably dispensable for generating FL, including somatic hypermutation of Ig genes and B-cell-mediated immunity, also were downregulated.
One strategy for decommissioning large blocks of DREs could involve suppression of their cognate TFs. To test this possibility, we identified TF binding motifs enriched in attenuated DREs and found that most, including *POU2F2*, *SPIB*, and *TCF3*, serve key functions in lymphocyte development and the GC reaction (Figure 4B) (Hagman and Lukin, 2006). Indeed, mRNA expression of these TFs is considerably decreased in all FL samples compared with CC samples, and there is relatively greater suppression of *TCF3* in subtype 1 than in subtype 2 (Figure 4B). To test whether diminished amounts of one such TF contribute to the decommissioning of DREs, we expressed a validated shRNA specific for *SPIB* in GM12878 B cells (Figure 4C). We selected for focused analysis four DREs that are known to bind the related ETS family factor PU.1 in these cells (Neph et al., 2012), which are attenuated in FL expressing low amounts of *SPIB*. Amounts of H3K27ac were reduced significantly at three of the four DREs in cells that express the *SPIB*-specific but not the control shRNA (Figure 4D). The loss of H3K27ac at these three DREs coincided with attenuated expression of at least one putative target gene, whereas the putative target for the fourth DRE was unaffected (Figure 4E). Our findings indicate that coordinated suppression of B cell TFs and their target REs is a key component of the pathologic circuitry for FL transformation and survival.

**FL Usurps Regulatory Circuits from Other Cell Lineages**

In addition to discarding non-essential GC circuitry, FL might evolve by co-opting circuits from other cell lineages that promote growth and survival. Indeed, ~20% (11,318) of the variable DREs identified in FL are completely absent in all tested control B cell populations, including CCs, CBs, TsBs, resting or activated PBBs, and the B cell line GM12878. Nearly one-third of these “de novo” DREs overlap a region of open chromatin found in at least one of 27 primary cell types (Figure S4A) (Bernstein et al., 2012; Thurman et al., 2012). Many of the remaining de
Figure 5. Novel DREs in Pathogenic FL Regulatory Circuits
(A) Distribution of 706 de novo DREs (orange) that overlap REs in other cancer types (purple).
(B) Heatmap representation of enriched GO terms and KEGG pathways identified from upregulated genes within 500 kb of de novo DREs.
(C) RNA expression profiles (microarray) for TFs predicted to bind de novo DREs that are consistently upregulated in FL samples. Refer to Table S6 for expression values.
(D) RNA expression profiles (microarray data) for TFs predicted to bind de novo DREs that are differentially upregulated in either subtype 1 or subtype 2 FL samples. Refer to Table S6 for expression values.
(E) The relative expression of TFs, ranked by number of corresponding TF motifs within private DREs, in individual FL samples is compared to the average expression in all FL samples. Additional data are shown in Figure S4C.

Novo DREs are shared with one or more non-B-cell cancers (Figure 5A). The pathologic relevance of these elements is highlighted by a de novo DRE overlapping the promoter of an antisense non-coding RNA in the HOXA10 locus. This de novo DRE coincides with enhanced expression of HOXA10 in FL samples compared to CC samples (Figure S4B). HOXA genes are master TFs of embryonic development and are aberrantly expressed in many cancers (Shah and Sukumar, 2010). Indeed, overexpression of HOXA10 in bone marrow is associated with a block in B cell differentiation and induction of leukemia in mice (Argiropoulos and Humphries, 2007). Thus, FL commanders DREs that are irrelevant for normal B-cell identity or activation, suggesting that they control the expression of genes involved in cellular transformation.

To investigate this possibility, we performed ontology analysis on genes connected to de novo REs in our FL circuitry.
Consistent with cell-type distribution of the de novo DREs, many of the enriched pathways govern biological or oncogenic processes, such as vasculature development and gliomagenesis, in other lineages (Figure 5B). Other pathways correspond to more general oncogenesis mechanisms, including regulation of cell signaling and growth. Similar to attenuated DREs, the de novo DREs are enriched in motifs for TFs that exhibit higher expression levels across FL samples (Figure 5C). Dysregulated TFs include those involved in general cellular transformation (MAX, GFI1B) and the development of non-B cells (MEF2A, RUNX2). A subset of the enriched motifs corresponds to TFs whose augmented expression distinguishes subtype 1 from subtype 2 FL (Figure 5D). TFs with higher expression in subtype 1 FL samples, which resemble CC samples, include BACH2, a transcriptional repressor that cooperates with BCL6 to regulate gene expression in GC B cells (Huang et al., 2014), as well as TFs with diverse roles in hematopoiesis (FOXO1, GATA6). TFs with relatively higher expression in subtype 2 FL cells, which resemble activated B cells, include both the RELA subunit of NF-kB and IRF1, a component of the MYD88 signaling cascade (Kushima et al., 2013). We next considered whether variable DREs unique to individual FL samples (i.e., “private” DREs) arise from deregulation of additional TFs. We find that when an FL sample expresses substantially higher levels of a unique TF than other FL samples, its private DREs are enriched for the binding site for that TF (Figures 5E and S4C). We conclude that FL commandeers oncogenic regulatory circuits, in part, by activating expression of general or lineage-inappropriate TFs.

**Variable DREs Are Enriched for Inherited and Acquired Sequence Alterations**

Prior studies have shown that sequence variants co-localize with regulatory regions and disrupt TF binding motifs; they thereby provide a potential genetic source for perturbations in DRE activity and, consequently, target gene expression (Corradin et al., 2014; Huang et al., 2014; Kushara et al., 2013; Maurano et al., 2012). To explore whether this mechanism is active in FL, we selected high-quality DRE sequence variants that were present in multiple RNA-, FAIRE- or ChIP-Seq datasets in an individual FL sample. We assigned each variant either as a SNP on the basis of its annotation in the 1000 Genomes or dbSNP-All SNPs databases or as a putative somatic single-nucleotide variant (SNV). To validate this variant-calling method, we performed Sanger sequencing by using FL samples and patient-matched peripheral blood mononuclear cells (PBMCs) when available. Of 28 identified variants, 24 were confirmed (86%). Twelve were present in FL samples and matched PBMCs, suggesting that they are private SNPs; nine were present in the FL sample (no matched PBMCs), and three were confirmed to be of somatic origin (in FL samples but not PBMCs; Table S7). Thus, ourinformatics approach for identification of DRE sequence variants is highly accurate.

Consistent with a potential function for SNPs and SNVs in altering DRE activity, we found that both are significantly enriched (p < 0.0001) in variable DREs (Figure 6A). Analysis of disease- and trait-associated index SNPs from the GWAS catalog revealed that B-cell cancer SNPs in variable DREs were significantly enriched relative to the fraction of total SNPs in these regions (Figure 6B). In contrast, variable DREs are depleted for SNPs associated with other cancers and several unrelated traits. Although somatic mutation levels vary in subtypes of some cancers (Kandoth et al., 2013; Pasqualucci et al., 2014), there were no significant differences in SNV load for the variable DREs present in FL subtype 1 versus subtype 2 samples (Figure 6C). Similarly, we did not identify DRE sequence variants significantly associated with either subtype. Together, these findings suggest that the activity of some DREs is altered in FL via sequence variation and that a subset of these pathologic elements might arise from somatic mutation during lymphomagenesis.

To explore the impact of sequence variants on DRE function, we identified SNPs or mutations predicted to disrupt TF motifs by using TRANSFAC and FunSEQ (Kururana et al., 2013; Matys et al., 2006). A significantly greater fraction of motif-disrupting variants occurred in variable than in unchanged DREs, suggesting a functional role for sequence variation in altering enhancer activity (Figure 6D). Moreover, some of these variants overlap binding motifs for TFs important in GC B-cell biology (Figure 6E). These TFs include POU2F2, IKZF1, and TCF3. We selected three sequence variants, all located in attenuated DREs, for more in-depth analysis (Figure 7). Although not previously linked to FL, a SNP associated with familial chronic lymphocytic leukemia (rs674313) (Slager et al., 2011) is located in a binding motif for IKZF1, a TF that regulates cell-fate decisions during lymphopoiesis. The second variant most likely corresponds to a private SNP because it is also present in the patient’s PBMCs but is not found in available SNP databases (Table S5). The private SNP overlaps a binding site for SP1, a ubiquitous TF that governs many biological processes, including cell-cycle regulation and apoptosis (Archer, 2011). The third variant is a somatic mutation that was present in the FL sample but not in the matched PBMCs (Table S5). The mutation is located in a predicted binding site for TCF3, a TF important for many aspects of B-cell development and activation (Hagman and Lukin, 2006).

Binding of each TF to its predicted site in the attenuated DRE is supported by IKZF1, SP1, and TCF3 ChIP-seq data from GM12878 B cells (Figure 7A) (Neph et al., 2012). Importantly, expression of target genes predicted by FL pathogenic circuitry is also reduced in FLs harboring the altered DREs (Figure 7B). Each of the target genes has a demonstrated role in B-cell cancer: allelic variants of HLA-DQA1 have been associated with an increased risk of childhood acute lymphocytic leukemia (Uraya et al., 2013); DUSP6, a MAP kinase phosphatase specific for ERK1 and 2, is deregulated in multiple cancers (Bermudez et al., 2010); and IRF8, a TF that regulates BCL6 and AICDA in GC reactions, is mutated in several types of NHL (Morin et al., 2011; Pasqualucci et al., 2014; Wang and Morse, 2009). A functional impact for the variants is demonstrated by oligonucleotide precipitation assays. Each TF binds robustly to a sequence corresponding to the reference allele. In contrast, binding is substantially diminished when oligonucleotides contain the identified variants (Figure 7C). Finally, enhancer activity of each DRE, as measured by luciferase reporter assays, is significantly attenuated when the reference enhancer is mutated to its corresponding sequence variant (Figure 7D). To our knowledge, this is the first reported example of an acquired mutation that occurs in
an enhancer and attenuates TF binding and target-gene expression in cancer. Together, our data indicate that DRE sequence variants, whether inherited or acquired, contribute to altered gene-expression programs that drive lymphomagenesis.

**DISCUSSION**

A key question regarding lymphomagenesis is how the normal gene-expression programs of GC B cells are dramatically altered during transformation. We now provide a comprehensive wiring scheme for pathogenic gene-expression circuits in a common B cell cancer; this wiring scheme is a composite of changes to normal GC B-cell circuitry and regulatory circuits commanded from other cell lineages. Importantly, the collection of regulatory elements incorporated into FL pathogenic circuits is significantly enriched for sequence variants, some of which disrupt TF binding and attenuate the expression of their target genes.

Our epigenome-centric approach revealed two distinct subtypes of low-grade FL, which is considered a single diagnostic entity despite clinical heterogeneity. Each subtype exhibits a characteristic pathogenic enhancer profile. The variable DREs and linked genes specific to subtype 1 most resemble the regulatory circuits seen in CCs, whereas subtype 2 FL cells acquire components of normal plasmablast circuitry. Notably, the expression of genes linked to subtype-specific DREs largely recapitulates patterns observed for subtypes of DLBCL, termed GC- and activated B cell (ABC)-DLBCL. Similar to findings in DLBCL, in FL the distinct variable DREs and target genes most likely reflect divergent modes of pathogenesis, whereby subtype 1 maintains survival pathways downstream of tonic BCR signaling, as occurs in GC-DLBCL. In contrast, subtype 2 variable DREs govern genes that are responsive to chronic BCR signaling (as occurs in ABC-DLBCL) (Rui et al., 2011). Our definition of FL regulatory circuits should inform mechanistic studies into common and distinct modes of lymphoma pathogenesis. Researchers could use variable DREs that distinguish subtype-specific circuits to develop new precision-medicine strategies by directly targeting these elements with sequence-specific chromatin modifiers to reverse their pathogenic function.

A key insight from our integrative analysis of FL regulatory circuits is the central role for TFs in driving pathogenic changes to DRE function. Specifically, we identified distinct TF cohorts...
associated with attenuated versus enhanced activity of DREs in FL samples. Compared with normal CC counterparts, FL cells decommission some circuitry by attenuating the expression of TFs that regulate GC processes that are either dispensable for lymphomagenesis, such as Ig class-switch recombination, or prohibitive for transformation, such as pro-apoptotic pathways. Our study provides experimental support for this general mechanism and demonstrates that depletion of one such factor, SPIB, reduces the activity of predicted enhancer targets and genes linked in the pathogenic FL circuitry. In contrast, a second set of TFs overexpressed in FL activate DREs that are normally silent in all subsets of mature B cells. The circuits targeted by these TFs include genes involved in general oncogenesis mechanisms, such as cell signaling, survival, and proliferation.

The central role of TFs in rewiring regulatory circuits was manifested at a genetic level by enrichment of DREs harboring the indicated sequence variants in FL samples and the reference sequence in CC samples (upper 2 tracks). The bottom track shows ChIP-seq data for the indicated TFs performed in the GM12878 B cell line (Neph et al., 2012). Arrows indicate the variant sequence and location in position-weight matrices for each TF. (B) Expression of the DRE target genes quantified by microarray analysis. (C) Oligonucleotide precipitation assays demonstrate reduced TF binding in variant-containing compared to reference sequences. Immunoblots were probed with antibodies specific to the TF of interest (representative of three experimental replicates). (D) Luciferase reporter assays performed in lymphoma cell lines demonstrate significantly reduced activity for the variant-containing compared to reference sequences in the attenuated DREs. Luciferase activity is presented as an n-fold change for the enhancer vector relative to a reporter containing only the SV40 promoter (luciferase activity for the promoter-only reporter is set to a value of 1.0). Results represent the mean ± SEM of three independent experiments. Statistical significance (paired t test) was set at *p < 0.05.

In summary, our epigenome studies provide a rich resource for deciphering aberrations in the transcriptional circuitry that fosters pathogenesis of B-cell lymphoma. In particular, we find that FL co-opts beneficial regulatory circuits and prunes potentially deleterious connections to construct pathogenic cistromes via diverse mechanisms, including inappropriate TF expression and the acquisition of somatic mutations in DREs.

EXPERIMENTAL PROCEDURES

Detailed methods for sample collection, ChIP-seq, FAIRE-seq, and informatics analyses can be found in the Supplemental Experimental Procedures.

B-Cell Isolation

Single-cell suspensions of CD19+ B lymphocytes from each FL biopsy were isolated by physical disruption and magnetic-assisted cell sorting (MACS, Human CD19 Microbeads, Miltenyi). PBBs were isolated from blood samples by negative sorting (Human B cell Isolation Kit II, Miltenyi) and immediately processed (for obtaining resting PBBs) or subjected to in vitro activation with B-cell activation, perhaps as off-target consequences of somatic hypermutation (Khodabakhshi et al., 2012).

IL4, anti-CD40, IgM, and IgD (for obtaining activated PBBs). Tonsilar tissues were mechanically disrupted and digested with collagenase for 1 hr. GC B cells were isolated by MACS (CD19+ sorting) and flow cytometry so that CD10+CD44–CXCR4+ (CB) and CD10+CD44+CXCR4– (CC) populations could be sorted from tonsil populations (Caron et al., 2009).

ChiP- and FAIRE-seq
ChiP and FAIRE assays were performed as described (Giresi and Lieb, 2009; Koues et al., 2008). At least 3 ng of FAIRE, ChIP, or input DNA was used for indexed-library preparation. Samples were pooled (nine samples) and subjected to 42 bp single-end sequencing.

Knockdown Experiments
Knockdown of SPiB was achieved by electroporation of control (target sequence: 5′-AAGCTGGAGTACACTAC-3′) or SPiB shRNA (target sequence: 5′-TACAGCTGAAGTGTGGCCCGTC-3′) plasmid containing a selectable marker (CD4). Transfected cells were purified 48 hr post transfection using magnetic bead isolation for CD4 (Stem Cell Technologies, Vancouver, Canada).

Gene-Expression Analysis
RNA was purified (QiAgen RNeasy), amplified (Nugen Ovation Picolo SL or Ovation Pico), labeled (Nugen Encore Biotin), and hybridized on Affymetrix Human Gene 1.0ST arrays. Expression was quantified with Expression Console software (v1.2.0.20) with probe-level RNA and default settings. RNA-depleted (Ribo-Zero, Epicenter) libraries were prepared with TruSeq RNA sample kits with indexed adapters (Illumina), pooled (three libraries), and subjected to 100 bp paired-end sequencing. RNA-seq data were aligned to the reference genome (build GRCh37/hg19) with TopHat (Trapnell et al., 2012). Fragments per kilobase of transcript per million fragments mapped (FPKM values) were obtained via Cufflinks with default parameters.

Mutation Analysis
SNVs were identified from RNA- and ChiP-seq files via SAMtools (Li et al., 2009). Variants were filtered so that non-coding SNVs common to multiple sequencing formats could be identified, and known SNPs were removed on the basis of comparison to all SNPs in dbSNP, build 138. SNVs predicted to disrupt TF binding motifs were identified with TRANSFAC or FunSEQ (Khurana et al., 2013; Matys et al., 2006). For 28 randomly selected SNVs, the regions flanking the SNV positions were amplified from tumor and PBMC genomic DNA by PCR, and the products were sequenced.

Luciferase Assays
DREs were amplified by PCR and inserted downstream of the luciferase coding sequence in the SV40 promoter-driven pGL3 plasmid (Promega, France). Reporter vectors were transfected by electroporation into three lymphoma cell lines: OCI-LY7 (DREs 1–10), Raji (DREs 15–16), and RL (DRE 17). Dual luciferase assays were performed in duplicate according to the manufacturer’s protocol (Promega).

Oligonucleotide Precipitation Analysis
Assays were performed as described (Busu et al., 2009), 293T cells (American Type Culture Collection) transfected with TCF3, IKZF1, or SP1 expression plasmids were lysed in HKMG buffer (10 mM HEPES [pH 7.9], 100 mM KCl, 5 mM MgCl2, 10% glycerol, 1 mM DTT, and 1% Nonidet P-40). Lysates were incubated overnight at 4°C with biotinylated double-stranded oligonucleotides spanning the reference or variant TF sequence in the presence of a 10-fold excess of poly(dI–dC). DNA-bound proteins were collected with NeutraVidin UltraLink Resin (Thermo Scientific, Waltham, MA) and assayed by immunoblot.

SUPPLEMENTAL INFORMATION
Supplemental Information includes four figures, seven tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.immuni.2014.12.021.

AUTHOR CONTRIBUTIONS
E.M.O., J.E.P., and O.I.K. conceptualized the study and designed experiments. E.M.O. and J.E.P. supervised all aspects of the project. Experiments and data analyses were performed by O.I.K., J.E.P., R.A.K., S.O.P., L.-W.C., J.A.S, H.L., L.E.S., J.I.B. and T.B.H. Specimens were processed by O.I.K., R.A.K., J.A.S. and A.F.C. The manuscript was written by E.M.O., J.E.P. and O.I.K. with input from A.F.C.

ACKNOWLEDGMENTS
This work was supported by NIH grants CA156690 (to J.E.P. and E.M.O.), CA090547 (to O.I.K.), TR000448 (to the Washington University Institute of Clinical and Translational Sciences), and CA91842 (to the Siteman Cancer Center). We thank B. Screckman, T. Fehniger, N. Mosammaparast, M. Artyomov, and T. Wang for comments; the Washington University School of Medicine Lymphoma Banking Program of the Division of Medical Oncology in the Department of Medicine, N. Bartlett, N. Wagner–Johnston, K. Carson, and T. Fehniger for lymphoma biopsies and patient histories; M. Colonna for tonsil samples; and the Genome Technology Access Center for assistance with –omics analyses.

Received: July 14, 2014
Accepted: November 17, 2014
Published: December 25, 2014

REFERENCES


