Supplementary Methods

Detailed ChIP-Seq protocol

Cross-linked cell line pellets were thawed and lysed in cold cytoplasmic lysis buffer (20 mM Tris-HCl ph8.0, 85 mM KCl, 0.5% IGEPAL CA-630 + PI). Nuclei were pelleted at 3000g, resuspended in cold SDS lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.1 + PI) for 10 minutes, and sonicated to an average fragment size of 200-400 bp on a Branson sonifier. Sonication settings were as follows: 0.7 seconds on, 1.3 seconds off for 5-8 minutes of total sonication time, amplitude ~45%, adjusted as needed for a power output of 10-12 watts. Samples were diluted 1:10 in ChIP dilution buffer (0.01% SDS, 0.25% Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl, pH 8.1, 167mM NaCl +PI), and rotated at 4°C overnight with 2-5 ug of antibody. Antibodies used were H3K27ac (39133, Active Motif), p300 (A300-358A, Bethyl), MEF2B (ab33540, Abcam and HPA004734, Atlas), Pu.1 (sc-22805, Santa Cruz), Bach2 (ARP39513, AVIVA), PAX5 (sc-1975x, Santa Cruz), and HA (sc-7392x, Santa Cruz). Antigen-antibody complexes were collected with protein G Dynabeads (Life technologies) for 4 hours at 4°C, and sequentially washed 2-6x each with RIPA buffer (0.1% sodium deoxycholate, 0.1% SDS, 1% Triton x-100, 10mM Tris-HCl pH 8.0, 1mM EDTA, 140 Mm NaCl), RIPA/high salt buffer (0.1% DOC, 0.1% SDS, 1% Triton x-100, 10mM Tris-HCl pH 8.0, 1mM EDTA, 360 mM NaCl), LiCl wash buffer (250mM LiCl, 0.5% NP40, 0.5% deoxycholate, 1mM EDTA, 10mM Tris-HCl, pH 8.0), and TE Buffer pH 8.0 (10mM Tris-HCl pH 8.0, 1mM EDTA). Beads were then resuspended in Low-SDS ChIP elution buffer (10mM Tris-HCl pH 8.0, 0.5M EDTA, 300mM NaCl, 0.1% SDS, 5mM DTT) and incubated for 6 hours at 65°C to elute DNA and reverse crosslinking. Supernatants were treated with RNase and proteinase K, and ChIP DNA was then purified with AMPure beads (Beckman-Coulter). For sequencing library preparation, ChIP DNA was end-repaired (End-It, Epicentre), A-tailed (Klenow

fragment 3'-->5' exo-, New England Biolabs), and ligated to barcoded illumina adaptors (Quick T4 DNA ligase, NEB). Each reaction was followed by clean-up with AMPure beads. Ligation products were amplified by PCR for 14-18 cycles with illumina indexing primers and PFU Ultra II HS PCR mix (Agilent). Library size selection to 300-600 bp was performed by gel purification (E-Gel SizeSelect 2%, Life technologies) or two-step AMPure bead selection.

For primary tissues, pellets of cross-linked material were processed as above, except that the cytoplasmic lysis step was omitted and pellets were lysed directly in 200-500 ml of SDS lysis buffer, diluted with ChIP dilution buffer to 900 mcl for sonication, and subsequently brought to a final dilution of 1:10. Prior to immunoprecipitation, DNA was quantified from sonicated chromatin, and an equivalent of 2-5 million cells were used per ChIP.

PEAR-ChIP Rearrangement filtering

Since dRanger (1) and BreakPointer were designed to work on WGS of tumor and normal pairs, technical changes to the original dRanger post-processing and filtering criteria were required to accommodate highly non-uniform H3K27ac profiles with short reads and no matching normals. For rearrangements supported by more than 5 read pairs, we assigned a quality score of 1 to events meeting the following criteria: min(zstdev1,zstdev2)>-2, max(zstdev1,zstdev2)<6, min(nuwpT1,nuwpT2)<10, and max(nuwpT1,nuwpT2)<30. For rearrangements supported by 3-5 read pairs, a quality score of 1 was assigned for min(zstdev1, zstdev2)>-2, max(zstdev1, zstdev2)>-1, max(zstdev1, zstdev2)<4, (1-fmapqzT1)*stdev1>2, (1-fmapqzT2)*X.stdev2>2, fmapqzT1<0.9, fmapqzT2<0.9, min(nuwpT1,nuwpT2)<5, and max(nuwpT1,nuwpT2)<20, and zero quality to all others. Rearrangements supported by only 2 read pairs were discarded, except when found in

association with another detected rearrangement (as is often the case in balanced translocations). To remove alignment artifacts and common germline translocations, we filtered out rearrangements detected by WGS in a panel of 12 normal DNA samples from prior cancer sequencing studies (2), and known germline variants from the DGV database (3). We removed 24 rearrangements occurring within the T cell receptor loci, which may have been derived from infiltrating reactive T cells. Additionally, 12 events involving the HLA loci were removed, as the combination of high inter-gene homology and sequence variation made rearrangement detection unreliable. Finally, we manually reviewed the results and removed 53 rearrangements (over all samples) that showed features suggestive of an artifact, typically due to a combination of repetitive or low-complexity DNA sequence, high overlap of all supporting reads on one side of the putative event, overlapping read pairs mapping to multiple distant sites in the genome, and lack of adjacent properly-paired reads indicative of true local acetylation signal.

To compare PEAR-ChIP detection to WGS in 4 HGB cell lines, WGS data was aligned to hg19 with BWA-MEM, and rearrangements were called by dRanger and BreakPointer, using the default parameters and same normal variant filtering strategy listed above, and taking all rearrangements with somatic score of 4 or higher. We considered all large-scale rearrangements (inter-chromosomal and intra-chromosomal with span > 2 Mb) detected in WGS data with breakpoints falling within 250 kb of the transcriptional start site of a recurrently rearranged gene in diffuse large B-cell lymphoma as listed in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer.

Supplementary figure legends

Supplementary Figure S1: H3K27ac ChIP-Seq tracks and position of rearrangements of possible oncogenic significance detected by PEAR-ChIP. Red bars join discordant intra-chromosomal read pairs. **A.** Inter-chromosomal rearrangement (black arrows) between the *IGH* locus and a gene desert near *ID2* in HGB-07. **B.** Positions of an inter-chromosomal rearrangement (HGB-03) and a deletion (HGB-04) disrupting the coding sequence of *FOXN3*. **C.** Positions of a deletion (HGB-04) and a complex inversion / deletion (HGB-02) involving acetylated elements in the first intron of *MEF2C*. **D.** An inversion encompassing several exons of *ETV6* in HGB-01. **E.** A large intra-chromosomal deletion on chromosome 14, resulting in rearrangement between *ZFP36L1* and the IGH locus. **F.** ChIP-Seq tracks for H3K27ac in HGB-04 and IRF-1 in interferon-stimulated K562 cells (ENCODE data) at the site of a PEAR-ChIP detected tandem duplication upstream of *NOS2* in HGB-04. The red bar marks the region reported to contain an interferon-responsive *NOS2* enhancer .

Supplementary Figure S2: Quantitative RT-PCR evaluation of lymphoma samples. **A.** Classification scheme and heatmap showing normalized signature gene expression levels measured by qRT-PCR in 17 high-grade B cell lymphoma samples. Cases evaluated by PEAR-ChIP are indicated in bold. PMBL samples were initially identified on the basis of mediastinal presentation ("site"). The "qPCR MvD" score was calculated as the average of the log2 fold change of four PMBL signature genes over the sample mean. All mediastinal lymphomas showed >2-fold mean elevation of PMBL signature genes and were classified as PMBL. Non-mediastinal HGB lymphomas (DLBCL), were initially classified as GCB or non-GCB subtype on the basis of the Hans immunohistochemical classifier. The "qPCR GvA" score was calculated as the difference of the average log2 fold change in GCB-DLBCL signature genes minus the average log2 fold change in ABC-DLBCL signature genes. All Hans classifier-identified GCB-DLBCL cases had a positive GvA score, while all but one Hans non-

GCB case had a strongly negative score. The discordant case (HGB-13) was considered unclassified, and all other cases were classified accordingly as GCB-DLBCL or ABC-DLCBL. **B.** Results of RT-PCR performed with primers designed to detect the predicted fusion transcript resulting from the *NCOA3-PDCD1LG2* rearrangement detected in HGB-01 by PEAR-ChIP. The expected product is detected only in RNA from HGB-01. **C.** Normalized gene expression by RT-qPCR in 17 HGB lymphomas for *NOS2* and *MEF2C*. Samples are color coded by HGB subtype: green = GCB-DLBCL, red = ABC-DLBCL, blue = PMBL. Solid bars indicate samples evaluated by PEAR-ChIP. Samples containing genomic lesions are marked with a symbol on the plot for the corresponding gene: "+" denotes an enhancer tandem duplication, while "±" denotes intronic deletions affecting acetylated elements.

Supplementary Figure S3: Enhancer looping interactions and TF binding at *BCL6* enhancers. **A.** 3C evaluation of interactions between the *BCL6* promoter and *BCL6* locus enhancers in HGB cell lines. Super-enhancer regions are marked as in Figure 3A. **B**. ChIP-seq tracks, generated with antibodies against the indicated sequence-specific TFs, p300, or H3K27ac, across the *BCL6* locus and upstream regions in HGB cell lines. H3K27ac data for normal B cell populations shown for comparison. Super-enhancer regions are marked as in Figure 3A. MEF2B binding sites (black arrows) and selected other TF binding sites (open arrows) are marked as in Figures 3A, 3C, and 3E. **C.** Protein blot using antibodies against MEF2B and MEF2C in representative HGB and MCL cell lines. TATA-Binding protein (TBP) is shown as a loading control.

Supplementary Figure S4: Genome-wide correlation and motif analysis of MEF2B and p300 ChIP-Seq in HGB cell lines. **A.** Hierarchical clustering of ChIP-Seq signal at TF and p300

ChIP-seq peaks. **B.** *De novo* motif analysis at peaks defined by MEF2B ChIP-seq in HGB cell lines. **C.** *De novo* motif analysis at peaks defined by p300 ChIP-seq in HGB cell lines.

Supplementary Figure S5: Additional details regarding candidate *MYC* enhancers identified in B cell lymphomas. **A.** H3K27ac ChIP-seq track from Karpas-422, with position of *MYC* locus ChIP-PCR primers used in part B. **B.** Results of H3K27ac ChIP-PCR performed in Karpas-422 cells at day 7 post-transduction with pLX-304-GFP or pLX-304-MYC. Values for each primer pair are expressed relative to the value for pLX-304-GFP. Significance was calculated by two-tailed t-test. **C.** H3K27ac ChIP-Seq tracks (coverage range 0-2.5 fpm) and position of a GWAS-identified risk SNP for CLL/SLL (rs2456449, red) and SNPs in linkage disequilibrium (r^2>0.8) in 1000 genomes project data (blue, one linked SNP lies outside displayed range). GM12878 lymphoblastoid cell line H3K27ac ChIP-Seq data is from ENCODE. Black dashes indicate the position of DNAse hypersensitivity peaks in merged ENCODE B cell populations, while green marks indicate the position of GWAS-identified or linked SNPs within DNAase peaks. The long red line marks the position of the SNP detailed below the diagram, which is labeled in bold. Linked SNP rs2445610 is at the center of an enhancer that shows acetylation in SLL, but not in DLBCL. Scoring against optimized motif position weight matrices from the HOMER motif library (logos displayed) showed that the variant allele for rs2445610 generates a significant motif for predicted binding by ETS family transcription factors, while the reference allele does not (BT=below threshold for motif match). **D.** H3K27ac ChIP-Seq tracks (coverage range 0-2.5 fpm) and position of a GWAS-identified risk SNP for DLBCL (rs13255292, red). Linked SNPs (r^2>0.8), ENCODE DNAase HS peaks, and DNAase HS-associated SNPs are indicated as in (C). rs13255292 itself is in the center of an enhancer that shows acetylation in normal centroblasts and DLBCL/PMBL, but not in SLL. The variant in this case lies 10 bp from a strong E-box motif,

supported as functional by ENCODE ChIP-Seq data for the E-box-binding TF BHLHE40 in GM12878 cells. **E.** H3K27ac ChIP-Seq tracks (coverage range 0-2.5 fpm) and position of a GWAS-identified risk SNP for Hodgkin lymphoma (rs2019960, red). Linked SNPs (r^2>0.8), ENCODE DNAase HS peaks, and DNAase HS-associated SNPs are indicated as in (C). Seven linked SNPs lie outside the displayed range. Linked SNP rs7826019 is in the center of an enhancer that shows acetylation in normal centroblasts and DLBCL/PMBL, but not in SLL, and lies 2 bp from a strong binding motif for the TF Pu.1. ChIP-seq demonstrated strong Pu.1 binding at this site in HGB lymphoma cell lines. Linked SNP rs59602790 is in a DNAase site within an enhancer that is acetylated in HGB cell lines. The reference allele supports a significant match for an NF-κB motif, while the variant allele does not (BT=below threshold for motif match). ENCODE ChIP-Seq data shows a weak but significant binding peak for the NF-κB subunit RELA at this site in GM12878. Both rs7826019 and rs59602790 show perfect correlation with the GWAS-identified SNP in the 1000 genomes GBR and CEU populations (r^2=1).

Supplementary Table S1: Clinical and diagnostic information on lymphoma cases

Supplementary Table S2: Rearrangements detected by PEAR-ChIP in 14 lymphoma primary samples and 8 lymphoma cell lines.

Supplementary Table S3: Detection of large-scale rearrangements at known oncogene loci in four HGB lymphoma cell lines by whole genome sequencing data, in comparison to PEAR-ChIP.

Supplementary references

1. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, et al. The genomic complexity of primary human prostate cancer. Nature. 2011;470:214–20.

2. Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499:214–8.

3. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: A curated collection of structural variation in the human genome. Nucleic Acids Res. 2014;42:986–92.