

METHODS

Maintenance of mouse lines

Constitutive H1c/e-deficient ($H1c^{+/-}/e^{+/-}$) mice were previously generated by the Skoultchi lab (Albert Einstein Medical College¹). By crossing heterozygous mice, we generated homozygous $H1c^{-/-}/e^{-/-}$ and wild-type littermate controls. Mice were used for assessment of GC formation induced by immunization with SRBCs or affinity maturation by immunization with NP-KLH and NP-CGG. We also used CD45.1 mice (stock 002014 purchased from the Jackson laboratory) for mixed bone marrow chimera experiments. By crossing $H1c^{-/-}/e^{-/-}$ with the transgenic *VavP-Bcl2*, we generated *VavP-Bcl2; H1c^{-/-}/e^{-/-}*, *VavP-Bcl2; H1c^{+/-}/e^{+/-}*, and littermate controls. These mice were used as bone marrow donors for transplantation into C57BL6 recipients for lymphomagenesis and survival studies. Experiments used 8-week old age- and background-matched C57BL/6J and female NOD/SCID mice (NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ, stock no. 005557, Jackson Laboratory). No experiment was blinded nor randomized.

Animals were maintained in a pathogen-free animal facility on a 12h:12h light:dark cycle at an ambient temperature of 23°C, 50% humidity. Water and food were provided ad libitum. All experiments were conducted with sex, age and strain matched mice. All mice were followed until any of several criteria for euthanizing were met, including severe lethargy, more than 10% body weight loss, and palpable splenomegaly that extended across the midline, in accordance with our Weill Cornell Medicine IACUC-approved animal protocols. Animal care and all experiments were performed in strict compliance with the institutional guidelines and protocols of Weill Cornell Medicine Institutional Animal Care and Use Committee.

Germinal center formation experiments and flow cytometry

Age- and sex-matched C57BL/6 mice (8-12 weeks old) were immunized intraperitoneally with 0.5 ml of a 2% SRBC suspension in PNBS (Cocalico Biologicals) and euthanized after 7-16d. Single cell suspensions from mouse spleens were separated by Ficoll gradient centrifugation and stained using fluorescent labeled anti mouse antibodies (**Supplementary Table 1**). DAPI was used for the exclusion of dead cells. For EdU cell cycle analysis, mice were injected intravenously with EdU (1mg) one hour prior to euthanasia (Invitrogen, Click-it Plus EdU Alexa Fluor 488 Flow Cytometry Assay Kit). Cells were stained, fixed and permeabilized according to kit instructions, and stained with DAPI to determine DNA content. For RNA-seq studies of sorted GC B cell populations, mononuclear splenocytes were stained with APC anti-B220 (553092, dilution 1:400), PE-Cy7 anti-FAS (55763, BD Biosciences, dilution 1:500), and FITC anti-GL7 (BD Biosciences, 553666, dilution 1:500) and DAPI, and sorted on a BD FACSAria III instrument.

Bone marrow transplantation

For murine bone marrow (BM) transplantation assays, BM cells from 6- to 8-week-old male and female donors were harvested. 1×10^6 BM cells of each type were retroorbital vein injected to female C57BL/6 mice lethally irradiated with two doses of 450 rad. For mixed BMT assays, 50:50 syngeneic bone-marrow was transplanted and

experiments were initiated after two months of engraftment. All mice were monitored for survival until any one of several criteria for euthanizing were met, including severe lethargy, palpable mass, severe body weight loss and hunched body. In our survival analysis, we censored one mice that died within less than 1.5 months after transplantation and lacked distinct pathological features due to failed BM engraftment.

Flow cytometry and sorting

Data was acquired on BD FCAS Canto II flow cytometer analyzers, and analyzed using FLOWJo software package (v10.5.3 Beckton Dickinson). When B cell populations were sorted, single cell suspensions of splenocytes were pre-enriched in B cells using CD45R (B220) magnetic microbeads (Miltenyi Biotech, 130-049-501). Cell sorting was performed using BD Aria II sorter with 5 lasers (355, 405, 488, 561, 640) and BD Influx sorter with 6 lasers (355, 405, 455, 488, 561, 640).

Immunohistochemistry and quantification

Mice organs were fixed in 4% formaldehyde and embedded in paraffin. Tissue processing and staining were performed by the Laboratory of Comparative Pathology (MSKCC). Briefly, five micron-sections were deparaffinized and heat antigen retrieved in citrate buffer pH=6.4, and endogenous peroxidase (HRP) activity was blocked by treating the sections with 3% hydrogen peroxide in methanol. Indirect immunohistochemistry was performed with anti-species specific biotinylated secondary antibodies followed by avidin-horseradish peroxidase or avidin-AP, and developed by Vector Blue or DAB color substrates (Vector Laboratories, Burlingame, CA USA). Sections were counterstained with hematoxylin. The following primary antibodies were used: biotin-conjugated anti-B220 (550286; BD Biosciences), anti-CD3 (ab16669; Abcam; Cambridge, UK), anti-PNA (B1075; Vector Laboratories), anti-KI67 (12202; Cell Signaling Technology; Danvers, MA, USA). Gamma-H2A.X staining (05-636, Millipore, clone) and Cleaved Caspase-3 (9661m Cell Signaling) was performed for immunohistochemistry. Immunohistochemistry quantification for H3K36me2 intensity (high, mid, low, negative) as well as B220+ lesion area was performed on digitally scanned slides on Halo software (v3.0.311.201).

ELISA and ELISPOT

For analysis of T cell-dependent antibody production, mice were immunized intraperitoneally with NP-KLH (200ug) in alum and were boost-immunized with NP-CGG28-30 (200ug) 21d after the primary immunization. Serum samples were collected at 14d, 21d, 35d, at 61d post primary immunization, and the abundance of NP hapten-specific immunoglobulin IgG1 titers analysed by ELISA. Sera were tested for binding of NP-specific antibodies to low haptenated BSA (NP9-BSA)- versus high haptenated BSA (NP30-BSA)-coated plates using HRP-conjugated antibodies (SouthernBiotech, dilution 1:500). Optical density (OD) at 450nm was measured in a plate reader (BioTek), and the absorbance ratio was calculated by dividing the mean OD of NP8-BSA by the mean OD of NP30-BSA-coated wells. Mice were euthanized 61d post primary immunization, and bone marrow was collected for enzyme-linked immunospot (ELISPOT) assay. Cells were incubated for 20hours at 37°C on NP₂₆-BSA-coated or NP₄-BSA-coated 96-well MultiScreen-HA filter plates (Millipore). NP-

specific spots were visualized with goat antibody to mouse IgG1 (1034-05) or IgM (1021-05) conjugated to horseradish peroxidase (HRP; Southern Biotechnology), and color was visualized by the addition of 3,3',5,5'-tetramethylbenzidine (Southern Biotechnology). Plates were evaluated using an automated Zeiss ELISPOT reader system (ZellNet Consulting, Inc.).

Secondary tumor transplantation

Lymph node tumors from *wild-type H1;VavP-Bcl2* or *H1c^{+/-}/e^{+/-};VavP-Bcl2* were collected from moribund animals with enlarged lymph node. Single cell suspension was prepared with digestion buffer (Collagenase I, Dipasell). Tumor cells were counted in PBS, resuspended 2x in Matrix Matrigel (Thermo Fischer #CB40234A), and injected subcutaneously in NOD-SCID recipient female mice (1M tumor cells). Subcutaneous mass growth was monitored weekly. Palpable masses were measured with caliper, and tumor volume was calculated as $\text{Volume (mm}^3\text{)} = (\text{length} \times \text{width}^2) / 2$. Mice were euthanized six-weeks post engraftment. In accordance with our IACUC-approved protocol, in no experiments the limit volume 1500 mm³ was exceeded.

Mouse embryonic fibroblast reprogramming to induced pluripotent stem cells

Fibroblast cells were harvested from E14 embryos from pregnant *H1c^{+/-}/e^{+/-}* mice bred with *H1c^{+/-}/e^{+/-}* males. Confirmed genotype *H1c^{-/-}/e^{-/-}* and littermate control mouse embryonic fibroblasts (MEFs) were placed into culture and passaged a maximum of three times before reprogramming. MEFs were lentivirally transduced in six-well plates with OKSM-mCherry plasmids expressing Yamanaka factors (Oct4, KLF4, Sox2, Myc) and mCherry reporter upon Doxycycline induction. Infected MEFs were sorted two days post Doxycycline induction for mCherry⁺ cells and 3000 cells/well cultured on a fibroblast monolayer. Sorted cells were cultured in leukemia inhibitory factor (LIF) iPS media for 18 days with Doxycycline and imaged with alkaline phosphatase (Vector Red AP Substrate Kit Cat: SK-5100) for iPS clones (red) on Day 21.

PCR clonality

RT-PCR to evaluate IgVH rearrangements was performed on genomic DNA (gDNA) of B220 enriched splenocytes with a set of forward primers that anneal to the framework region of the most abundantly used IgVH gene families² (Vλ1, GCCATTTCCCCAGGCTGTTGTGACTCAGG), and reverse primer located in the JH1-4 gene segments: (Jλ1,3, ACTCACCTAGGACAGTCAGCTTGGTTCC). PCR electrophoresis analysis was ran on QIAxcel Advanced System (Qiagen)

WGS Driver Analysis

Whole genome sequencing (WGS) alignments for 101 DLBCL and matched normal sample pairs were retrieved as hg19-aligned .bam files from the European Genome Archive (<https://ega-archive.org/>, EGAD00001002123) and stripped of alignment information via picard tools revertSam, according to GATK best practices (<https://software.broadinstitute.org/gatk/best-practices>) to generate unmapped .bam files.

These unmapped.bam files subsequently underwent alignment to hg38 and somatic variant calling for SNVs, indels, and structural variants (SV's) via the Sarek pipeline

version 2.5.1 (<https://github.com/nf-core/sarek>). Briefly, the Sarek pipeline³ applies BWA mem alignment, duplicate removal, base recalibration, somatic SNV / indel calling via Strelka2⁴, somatic SV calling via Manta⁵, and variant annotation with VEP⁶. In summary, this pipeline yielded 1.36 million SNV across 101 DLBCL cases, which were used for downstream analyses.

Following the Sarek pipeline .vcf files containing SNV's were converted to MAF files via vc2maf (<https://github.com/mskcc/vcf2maf>) and used to characterize mutational signatures with deconstructSigs (<https://github.com/raerose01/deconstructSigs>) using the version 2 COSMIC signatures reference (v2, https://cancer.sanger.ac.uk/cosmic/signatures_v2) comprising 30 annotated SNV signatures. Briefly, deconstructSigs⁷ applied non-negative least squares to deconvolve each tumor sample 96-dimensional vector of somatic strand-collapsed trinucleotide context SNV counts into a vector of signature-specific SNV counts or *signature burdens* across the set of 30 COSMIC signatures S . We then computed a posterior probability for each signature $s \in S$ and variant v with strand collapsed trinucleotide context $c(v)$ as

$$P(s | c(v)) = \frac{P(c(v) | s) \pi(s)}{\sum_{\hat{s} \in S} P(c(v) | \hat{s})}$$

where the conditional probability $P(c(v) | s)$ is directly specified by the COSMIC v2 signature definition. We then used this posterior probability to assign a *maximum a posteriori* (MAP) signature label $signature(v) = argmax_{s \in S} P(s | context(v))$ for each somatic variant v in that sample. We then defined ncAID (non-canonical activity induced deamination) variants as those with COSMIC Signature 9 as their MAP assignment.

In addition to using deconstructSigs, we defined non-canonical AID variants (nc-AID) and canonical-AID (cAID) variants using *de novo* mutation signature analysis as described^{8,9}. Nearest mutation distance (NMD) was computed for SNVs within the same tumor-normal pair and was used to partition SNVs in two groups of clustered (NMD > 1kb) and non-clustered mutations. SNVs falling in protein coding sequences were classified into 96 categories according to 6 base substitutions. Clustered SNVs were counted separately for each sample, resulting in a 96 x 101 count matrix. A *de novo* signature extraction for the 101 DLBCLs was performed using Bayesian NMF (BayesNMF), with the 96x101 matrix as input^{8,10}. All 100 BayesNMF runs converged to a 4 signature solution and identified the c-AID signature, with dominant representation of G[C>T]T and G[C>G]T variants. We then labeled individual mutations genome-wide according to the 4 *de novo* signatures using deconstructSigs as above, and c-AID SNVs were retained for further analysis.

FishHook (<https://github.com/mskilab/fishHook>) was used to model background mutational processes and nominate DLBCL mutational hotspots across the cohort of 101 DLBCL cases (see above). FishHook takes as input mutation calls (e.g. VCF), a set of hypothesis intervals (.bed files), an eligible territory (.bed file), and a set of genomic covariates (.bed, .bw files), and outputs a model of mutation density and hypothesis intervals associated with a P value and enrichment score. Briefly,

fishHook⁸, models genome-wide somatic mutation density in tumor genomes across the intersection of an eligible territory (e.g. coding sequences, mappable genomic regions) and set of hypothesis intervals (e.g. genes) as a function of genomic covariates, which can represent sequence context composition, chromatin features, or the fractional overlap with reference genomic annotations. The maximum likelihood fit of a fishHook model, implemented as a gamma-Poisson regression, assigns weights to covariates and an expected mutation density to each hypothesis interval. An enrichment value is computed at each hypothesis interval as the ratio of observed to expected mutation density. The model is further used to define a cumulative distribution function (CDF) for mutation density at each interval. Each interval is then assigned a one-sided P value as the probability that the mutation density is greater or equal than the observed density.

FishHook was applied to analyze the distribution of 1.36M DLBCL SNVs across 24,498 protein coding genes (GRCm38.p6) using coding sequences (CDS) as the eligible territory and two sets of covariates: (1) the expected density of c-AID and nc-AID mutations (2) B-cell specific transcriptional and chromatin state. To generate the first set of covariates, two additional genome-wide fishHook models were first fit to nc-AID and c-AID mutation calls (as defined above) across 60,643 50 Kbp non-overlapping genome-wide interval tiles using non-CDS (including intergenic) regions as the eligible territory and the following covariates: overlap of ATAC-seq peaks from human purified GC B-cells (unpublished data); overlap of human GC B-cell super-enhancers. Super-enhancers were called from previously reported H3K27ac ChIP-seq data from human purified GC B-cells¹¹ using the ROSE method¹². The expected nc-AID and c-AID density from the model was then used to annotate .bed files of the genomic tiles and used as covariates for the downstream genic fishHook analysis.

To generate the second set of covariates, we annotated .bed files of protein coding genes annotated with the number of ATAC-seq peaks from human purified GC B-cells (unpublished data) within 10kb, the number of GC B-cell H3K27ac ChIP-seq defined super-enhancers within 100kb (see above), and the number of TSS within 10kb for genes expressed > 1 TPM in human GC B-cells (RNA-seq). Combining (1) and (2) yielded a fishHook model with 4 covariates, which was fit to annotate genes with P values and enrichment scores. We labeled genes with (Benjamini-Hochberg) FDR<0.01 as significant. We generated Q-Q plots by pairing observed $-\log_{10}$ transformed quantiles of observed P values (y-axis) with their corresponding $-\log_{10}$ transformed quantiles from the uniform distribution (x-axis). A genomic inflation factor λ was computed from as the slope of a least-squares regression line fitting these data while intercepting the origin. The value of λ (<1.05) was consistent with minimal statistical inflation¹³.

Bulk RNA-seq analysis

Sequencing results were aligned to mm10 using STAR v2.3.0 and annotated to RefSeq assembly accession GCF_000001635.20 using the R subread package v1.28.1. Differentially expressed genes within murine GCB were identified using the EdgeR package v3.20.9 GLM¹⁴ with thresholds of fold-change >1.5 and $p < 0.01$,

adjusted for multiple testing using Benjamini-Hochberg correction. Hierarchical clustering was performed on top variable genes (top 90th percentile) using correlation distance and Ward's minimum variance method. Differentially expressed genes between H1C^{-/-} E^{-/-} and H1 WT DLBCL cases were identified using the EdgeR package GLM with thresholds of fold-change >1.5 and p<0.05, adjusted for multiple testing using Benjamini-Hochberg correction. Gene set enrichment analysis was performed using the GSEA v2.0.13 algorithm, as described¹⁵, reporting the normalized enrichment scores (NES). Pathway analysis was performed using PAGE algorithm using default parameters. Pathway heat maps were prepared using hypergeometric mean distribution test.

ATAC-seq sample preparation and analysis

GC B-cells were isolated from wildtype and H1c^{-/-}/e^{-/-} mice by FACS. ATAC was performed as per the optimized Omni-ATAC protocol with digitonin as described¹⁶. ATAC-seq DNA libraries were prepared as described previously, and library concentration was checked by qPCR using the KAPA Library Quantification Kit^{16/15}. DNA libraries were sequenced on a Hi-Seq 2000 (Illumina) at the Weill Cornell Epigenomics Core Facility.

Paired-end 50 base pair ATAC-seq reads were trimmed to remove adapter sequences using NGmerge with the options “-z -u 41 -a”. Trimmed read pairs were aligned to version 38 of the mouse reference genome (GRCm38) using bowtie2 with the following options: “-X2000 --local --mm -k 4”. Aligned reads were sorted and filtered to exclude reads mapping to mitochondrial DNA and black-listed regions, and duplicate read pairs were removed using the “MarkDuplicates” program in picard tools (<http://broadinstitute.github.io/picard>), resulting in a final aligned, sorted, and filtered BAM file that was used for all subsequent analysis.

ATAC-seq peaks were called from Tn-5 corrected insertions using MACS2 callpeak with option “-g hs --nomodel --shift -75 --extsize 150 --keep-dup all --call-summits”. A B cell chromatin accessibility atlas containing 500 bp disjoint genomic intervals (DNA elements) was constructed from called peak summits across all primary cells using an iterative peak-ranking method as previously described¹⁶. To quantify accessibility across samples, the number of single-base Tn5-corrected insertions that fell within each 500 bp interval was counted from ATAC-seq bam files using the command “pyatac counts” in the nucleoATAC package¹⁷. Differential accessibility between conditions was computed using DESeq2. Normalization factors included in the call to DESeq2 were computed by quantile normalization with GC sequence content bias correction using the R package EDASeq. The chromatin accessibility log2 fold changes reported and used in all analysis were computed and shrunken using the “lfcShrink” function with option “type=ape” in the DESeq2 R package. Differentially accessible DNA elements were identified at FDR<0.05. All gene-based annotation was performed using Gencode Release M23 (GRCm38.p6, https://www.gencodegenes.org/mouse/release_M23.html). DNA elements were assigned to genes according to the nearest TSS of a protein coding gene.

ATAC-seq peaks Peak calling was performed by pooling all the samples to control false discovery using the MACS2 callpeak command with parameters ‘—shift-75 –extsize 150 –pval 0.01 –keep-dup all –call-summits’. A matrix of nonnucleosomal Tn5 insertion counts was generated for the common set of peaks. The count matrix was used to identify differentially accessible peaks with DESeq2. Reads within the final set of peaks were taken as library size for normalization. Differential peaks at FDR < 0.05 were retained for analysis. The peaks were annotated to RefSeq mouse genes using upstream promoter and regulatory regions (basal region plus upstream extension to the nearest gene up to 1 Mb).

MicroChIP-Rx and library preparation.

MicroChIP was performed using TrueMicroChip kit (Diagenode, C01010130) according to manufacturer instructions with following modifications. For each replicate, $\sim 2 \times 10^5$ sorted GC B-cells were pooled together with 1×10^6 *Drosophila* Kc167 cells in 1 ml total volume and processed using Covaris E220E sonicator at peak power 140, duty factor 5, 5 cycles/burst, 20 min per sample. The sonicated samples were then cleared with 10 min centrifugation and split into aliquots containing chromatin from 3×10^4 B cells, mixed with 20 μ g antibody (anti-H3 K27me3, C36B11 - Cell Signaling 9733; anti-H3 K36me2, C75H12 - Cell Signaling 2901), and incubated overnight with rotation at 4C. Complexes were pulled down using Protein A Dynabeads (Thermo, 10002D), washed and de-crosslinked using kit buffers, and DNA isolated using standard phenol-chloroform procedure. Libraries were prepared using MicroPlex Library preparation kit (Diagenode, C05010013) with 8 cycles of amplification, and cleaned up using AMPureXP beads (Beckman, A63881). Libraries were validated using Agilent High Sensitivity D1000 ScreenTape and pooled for sequencing using Illumina NextSeq 500. 1x75 bp Resulting .fastq files were aligned to mouse mm10 and *Drosophila* dm6 genomes using bowtie2.

ChIP-seq analysis

Resulting FASTQ files were aligned to mouse mm10 and *Drosophila* dm6 genomes using bwa-mem function of the BWA suite. ChIPseq data was normalized to dm6 spike-in reads using CompChIPseq algorithm. H3K27me3 and H3K36me2 ChIPseq peaks were called using the SICERpy algorithm¹⁸ on pooled BAM files from all respective replicates using pooled input from replicates (using parameters -F 1540 -gs 0.74 -w 200 -rt 1). H3K27Ac ChIPseq peaks were called using MACS2 narrow peak calling on pool BAM files from all respective replicates using pooled input from replicates (FC>2, q-value<0.001). Loci showing differences in ChIPseq abundance were determined by calculating the Comp-ChIPseq normalized read count within the union of peaks from both genotypes using the multiBigwigSummary function of the deepTools package, identifying loci showing FC>1.5. Difference in fraction of peak coverage between genotypes (sum of width of respective peaks divided by size of compartment bin) for H3K27me3 and H3K36me2 was used to determine shifting compartment grouping according to hierarchical clustering based on Euclidean distance into k=5 groups. Pathway heat maps were prepared using hypergeometric mean distribution test. To determine the prevalence of ChIPseq marks according to

genomic compartment, we calculated the fraction of each 100kb compartment that was covered by ChIPSeq peaks.

CUTANA CUT&RUN, Illumina sequencing, and data analysis

CUT&RUN was performed with H1wt and H1c^{-/-}e^{-/-} germinal center B-cells using CUTANA[®] protocol v1.5.1 [www.epicypher.com] which is an optimized version of that previously described¹⁹. For each sample, nuclei from two biological replicates were extracted by incubating cells on ice for 10 min in Nuclei Extraction buffer (NE: 20 mM HEPES–KOH, pH 7.9; 10 mM KCl; 0.1% Triton X-100; 20% Glycerol; 0.5mM spermidine; 1x complete protease inhibitor [Roche # 11836170001]), collecting by centrifugation (600 g, 3 min, 4°C), discarding the supernatant, and resuspending at [100 µl / 500K nuclei] sample in NE buffer. For each target 500K nuclei were immobilized onto Concanavalin-A beads (*EpiCypher* #21-1401) and incubated overnight (4°C with gentle rocking) with 0.5 µg of antibody (**Supplementary table 3**: IgG, H3K4me3, H3K27me3, H3K9me2 and H3K9me3 [all validated to SNAP-ChIP nucleosome standards²⁰]. CUT&RUN enriched DNA was purified using the *Monarch DNA Cleanup* kit (*New England Biolabs* #T1030S), and 10 ng used to prepare sequencing libraries with the *Ultra II DNA Library Prep* kit (*New England Biolabs* #E7645S).

Libraries were sequenced on the Illumina NextSeq 550, obtaining ~6 million paired-end reads on average (**Supplementary table 3**). Paired-end fastq files were aligned to the mm10 reference genome using the Bowtie2 algorithm²¹. Only uniquely aligned reads were retained, blacklist regions²² filtered out, biological replicates assessed for reproducibility (R>0.8), and then merged for subsequent analyses. H3K9me2 and H3K9me3 Cut&Run peaks were called without input using the SICER2 algorithm (-w 200 -rt 1 -s mm10). Differential peak calling for H3K9me2 and H3K9me3 was performed using SICER2 sicer_df command (-w 200 -rt 1 -s mm10). H3K4me3 Cut&Run peaks were called using MACS2 narrow peak calling without input (FC>2, q-value<0.001). Peak coverage within compartments was calculated as the sum of width from respective peaks within respective compartment bin.

Hi-C prep and sequencing

GC B-cells were isolated from wildtype and H1c^{-/-}/e^{-/-} mice FACS. Cells were resuspended in 1M cells/mL suspension and crosslinked with 1% formaldehyde for 10min with mixing at room temperature. The reaction was quenched with 0.2M glycine for 5min with mixing at room temperature. The Arima-HiC kit streamlined workflow was followed as per manufacturer's protocol in conjunction with Swift Biosciences Library kit preparation. Sequencing for Hi-C 4 samples was performed on half of an S2 100 NovaSeq flow cell.

Hi-C data processing

All study and public Hi-C data used have been pre-processed with the hic-bench pipeline²³. In short, sequencing reads have been aligned against the mouse reference sequence GRCm38/mm10 using bowtie2 version 2.3.1²¹ (special parameters: --very-sensitive-local --local). Next, reads have been filtered with GenomicTools tools-hic

filter command (special parameters: MAPQ > 20; read-pair distance > 5kb), yielding ~40% usable reads across all samples, with an absolute number of accepted intra-chromosomal reads > 220 million per replicate. Hi-C contact matrices have been generated at 10kb (for visualization only) and 100kb resolutions and normalized using an approach called “iterative correction and eigenvector decomposition” (ICE)²⁴.

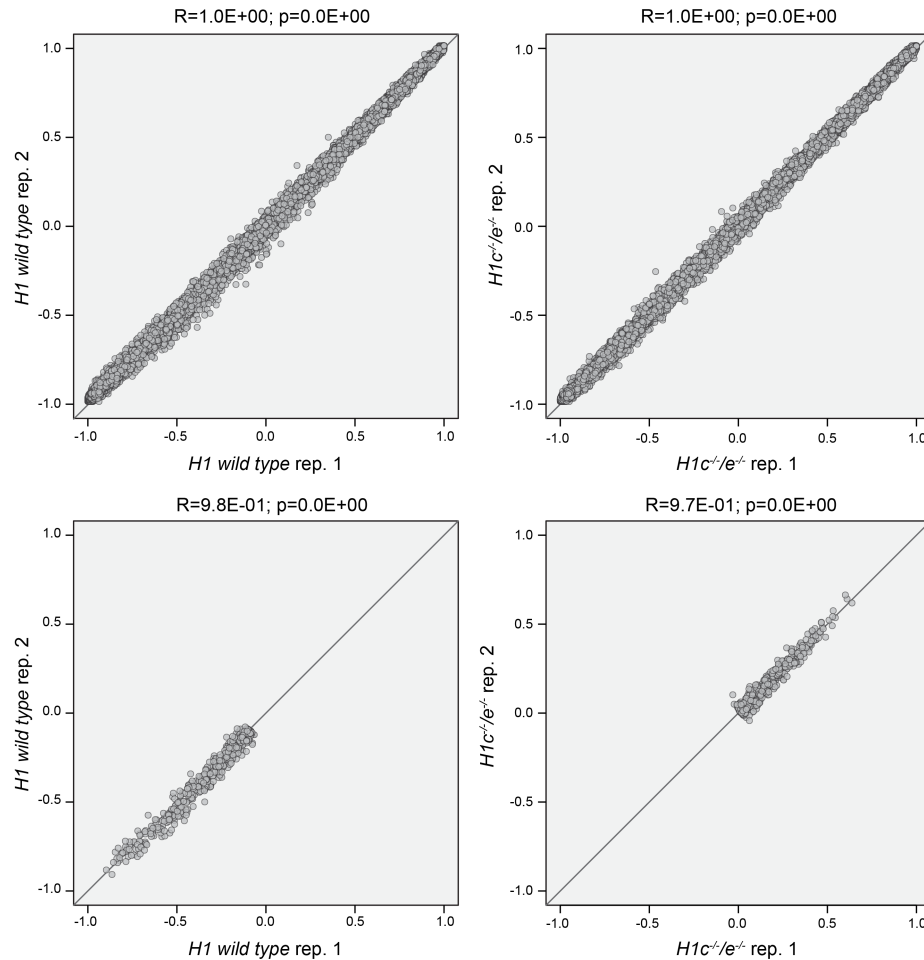
Genome-wide reproducibility analysis

In order to determine the genome-wide correlation and thus reproducibility between replicates of the same genotype and across genotypes, we employed HiCRep²⁵ to calculate the “stratum-adjusted correlation coefficient” (SCC) scores between all pairwise combinations of WT and H1DKO Hi-C replicates. To this end, we have used the ic-normalized matrices to first compute the SCC per chromosome for each pairwise comparison, and averaged the scores across all chromosomes to receive a single correlation score for each comparison (as shown in Extended Data Fig.5a).

Compartment analysis

We have used the c-score tool²⁶ to define compartments from all Hi-C samples at both 10kb (for visualization only) and 100kb resolutions. C-score tool uses a probability-based approach that allows for the quantitative comparison of compartment-scores (c-scores) across conditions. Quantitative differences between WT and *H1c^{-/-}/e^{-/-}* have been determined as delta c-score, which is defined as average H1-DKO c-score – average WT c-score, with average being the average c-scores across replicates. In order to identify significant differences between any two conditions (e.g. WT vs *H1c^{-/-}/e^{-/-}*), we first applied a t-test for each 100kb bin between replicates of the two conditions. Next, using a permutation-based approach by relabeling samples, we recalculated p-values using a t-test between permuted samples in order to create a background distribution of p-values²⁷. Then, we estimate a false-discovery rate per 100kb bin as the fraction of the number of 100kb bins with permuted p-values below the respective permuted p-value of the bin (estimate of false-positives) per the number of 100kb bins with real p-values below the respective p-value of the bin (estimate of true positives). We filtered for B to A shifts as FDR < 0.1, delta c-score > 0.1 and WT c-score < -0.1 and H1DKO c-score > 0. A to A stable / B to B stable have been filtered as absolute delta c-score < 0.1 and c-scores > 0.1 (A to A) or < -0.1 (B to B). B to B with A shift / A to A with A shift have been determined as FDR < 0.1 and delta c-score > 0.1 with WT c-score < -0.1 (B to B) and WT c-score > 0.1 (A to A). The same filters have been applied for comparisons of naïve B-cell reprogramming stages, comparing each time-point of reprogramming with naïve B-cells. Islands of B to A shifts have been defined as one or more continuous B to A shifts surrounded by B compartment in WT condition both upstream and downstream. The threshold within a 100kb analysis was set to ≥ 2 directly adjacent 100kb bins must have WT c-score < 0, both upstream and downstream of the B to A shift. Extension of B to A shifts have been defined similarly, with at least two of the adjacent three 100kb bins either upstream or downstream bins having a WT c-score > 0, and extensions from B upstream to A downstream have been inverted and merged with A upstream to B downstream extensions.

In order to estimate per-replicate reproducibility of both genome-wide compartment analysis as well as differential compartment analysis, we calculated the per-replicate Pearson correlation scores genome-wide as well as for B to A shifts (**Supplementary Figure 3**).



Supplementary Figure 3. Genome-wide correlation of compartment score between replicates (top) and correlation of compartment scores only undergoing B to A shifts (bottom) between replicates. Per-replicate reproducibility tested with Pearson correlation score.

TAD analysis

TADs were called with the integrated TAD caller of hic-bench making use of the hic-ratio insulation score method (pipeline-step is called “hicratio”), using ic-normalized matrices at 40kb resolution as input. Differential intra-TAD activity was determined using only TADs that had common TAD boundaries between WT and *H1c^{-/-}/e^{-/-}* (using a bias of three bins at each boundary between WT and H1-DKO). All such TADs have been evaluated in their TAD activity, which is defined as the average of all interaction scores within the common TAD area in either WT or H1-DKO. Differential intra-TAD activity is defined as the log₂ fold-change between the TAD activity of *H1c^{-/-}/e^{-/-}* and WT. Statistical significance was estimated by performing a paired two-sided t-test pairing interaction bins of the intra-TAD area between H1-DKO and WT, followed by

multiple testing correction. Compartment score integration was performed by overlapping 100kb compartment bins with TADs only if they fully lie within the called TAD boundaries. Gene integration was performed using gene promoter information from Ensembl Genes V85

Additionally, we compared TAD border insulation scores between WT and *H1c^{-/-}/e^{-/-}* samples. To this end, we used 40kb hicratio-scores created by hic-bench and calculated differential fold-changes between averages of hicratio scores across replicates supported by a two-sided unpaired *t*-test. There were no significant differences in called TAD boundaries in *H1c^{-/-}/e^{-/-}* compared to WT GCB-cells (FDR<0.05).

Interactivity analysis

We have estimated the local “interactivity” of chromatin around each 100kb in the genome as the average of interaction scores (from ice-normalized matrices) of a “500kb local domain”. This 500kb local domain was defined as the respective 100kb itself and additionally 200kb up- / downstream of the respective 100kb bin. The interactivity scores from all interaction scores within the 500kb local domain are calculated as a built-in function of hic-bench (boundary-scores step “activity 500kb”).

Compartment integration analysis with ChIP-seq

Gene integration was performed using gene promoter information from Refseq (TSS +/-2kb). Enrichment analysis with other features (TADs, ATAC peaks) was performed with the GSEA algorithm. Association of chromatin marks within shifting compartments was assessed by calculating the fraction of peak coverage between genotypes (sum of width of respective peaks divided by size of compartment bin). Differences in peak coverage fraction between genotypes for H3K27me3 and H3K36me2 was used to determine shifting compartment grouping according to hierarchical clustering based on Euclidean distance into k=5 groups. Curated lists of genes characteristic for respective shifting compartment characterization were defined as follows: group 1 type genes had gene promoters overlapping Group1 shifting compartments possessing WT and H1DKO c-scores < -0.1 in addition to either H3K9me2 differentially decreased loci or H3K9me3 differentially decreased loci; group 2 type genes had gene promoters overlapping Group2 shifting compartments in addition to H3K27me3 increased loci; group 3_4 type genes had gene promoters overlapping either Group3 or Group4 shifting compartments in addition to H3K36me2 increased loci; group 5 type genes had gene promoters overlapping Group5 shifting compartments possessing WT and H1DKO c-scores > 0.1 in addition to H3K27me3 decreased loci.

Virtual 4C analysis

To illustrate 3D chromosomal interactions of certain viewpoints with the rest of the genome, we employed an approach called “virtual 4C” to our Hi-C data as follows. First, a viewpoint is defined as a 1bp position in the genome, for example the annotated transcription start site of a gene of interest. Next, using this fixed point, we extended it by 10kb on each side, and find all Hi-C read-pairs of which one read aligns within the extended viewpoint area. Then, we created a read-counts table of the mate-reads of all such read pairs found using a genome binning approach at 20kb resolution

with adjacent windows overlapping by 2kb in order to create a smoothed signal (see^{28,29}; each read is thus counted in a total of 10 adjacent windows). The per-replicate read-count files were then normalized to the total number of reads per replicate times 1000 (similar to a counts-per-million normalization in RNA-Seq), and the average normalized read-counts were plotted for WT and H1-DKO for the viewpoint. Statistical differences in interaction strengths were calculated using functions glmQLFit and glmQLFTest from the edgeR package version 3.16.5. We have not performed a multiple testing correction on p-values, because each data-point is dependent due to overlapping windows, and would thus potentially over-correct the p-value.

Motif analysis

In order to identify potentially enriched sequence motifs in genomic regions, we employed the homer³⁰ findMotifsGenome.pl algorithm (version 4.10; default parameters). We used ATAC-Seq peaks that are defined as differential (see above) as genomic regions input. We further categorized the ATAC-Seq peaks by overlapping them with our compartment categories as defined above using bedtools. Results were subsequently merged by motif-related transcription factors as called by homer. We are reporting uncorrected p-values due to the low statistical power of using < 200 input sequences for motif search.

Single-Cell RNA-seq and Analysis

Splenocytes from 2 *H1c^{-/-}e^{-/-}* and 2 WT mice were sorted for GCB-cells (B220⁺Fas⁺GL7⁺) 7 days after SRBC immunization. Ten thousand sorted cells from each spleen were subjected to single cell RNA-seq using 10X Genomics Chromium platform. Library preparation for single cell 3' RNA-seq v3, and sequencing were performed at the Epigenomics Core at Weill Cornell Medicine. Libraries were prepared according to 10X Genomics specifications. Briefly, the four independent cellular suspensions were loaded onto the 10X Genomics Chromium platform to generate barcoded single-cell GEMs (Gel Bead-In Emulsions). After RT reaction, GEMs were broken and the single-strand cDNA was cleaned up with DynaBeads MyOne Silane Beads (ThermoFisher Scientific). cDNA was amplified for 12 cycles. Quality of the cDNA was assessed by Agilent Technologies 2100 BioAnalyzer, obtaining a product of about 1588 bp. This cDNA was enzymatically fragmented, end repaired, A-tailed, subjected to a double-sided size selection with SPRIselect beads (Beckman Coulter) and ligated to adaptors provided in the kit. A unique sample index for each library was introduced through 14 cycles of PCR amplification. Indexed libraries were subjected to a second double-sided size selection, and libraries were then quantified using Qubit fluorometric quantification (ThermoFisher Scientific). The quality was assessed on an Agilent Technologies 2100 Bioanalyzer, obtaining an average library size of 437 bp. Libraries were diluted to 10 nM and clustered on a HiSeq4000 at 1 nM on a pair end read flow cell and sequenced for 28 cycles on R1 (10x barcode and the UMIs), followed by 8 cycles of 17 Index (sample Index), and 98 bases on R2 (transcript). Libraries were sequenced to an average of 250 million reads per sample with average of 3500 cells per sample and an average depth of 60,000 mean reads per cell (resulting in 70% average sequencing saturation).

Sequencing data was processed with the 10X recommended pipeline. Briefly, raw base call files were demultiplexed by experiment from the sequencing run into FASTQ files. Cell Ranger was then run on the FASTQ files from each sample for alignment and cell assignment with the appropriate reference genome. This generated a sparse matrix file of features by barcodes. This sparse matrix data was then read into R using the R package Seurat 3.0.2, and standard quality control was run to remove cells with few genes or an over representation of mitochondria reads. X and Y chromosome genes were also removed from this analysis. Data was then scaled and normalized. Linear dimensional reduction was performed by calculation of PCA from the most variable genes. Cells were then clustered using a resolution value of 1.0 and visualized by UMAP. Module scores were calculated using the AddModuleScore function with a control value of 5. Individual genes and module scores were projected and used to identify appropriate classification of clusters.

Single-Cell Trajectory Analysis

RNA trajectory analysis was performed using the R package Slingshot³¹ version 1.2.0. This package was used to create a “pseudotime” based on a combination of PCA 1 and 2 calculated by Seurat, using the cells identified as Dark Zone as the anchor point. Linage plots were generated by projecting pseudotime onto cells mapped by PCAs 1 and 2. Pseudotime density plots were generated by genotype and zone using the ggplot2 version 3.2.1 geom_density function. Pseudotime module plots were generated by plotting pseudotime values against module scores with the ggplot2 geom_point function. The geom_smooth function was used to display the average lines. Cells were binned into ten deciles with equal numbers of cells, then a Wilcoxon rank sum test was used to compare module scores between genotypes to determine significance. The p-values were corrected for multiple hypothesis testing through Benjamini Hochberg correction. The differences between the splines for each module score across pseudotime was calculated, and colored according to the adjusted p-value. These plots were arranged with ggpubr version 0.2.2 grid.arrange.

Cell were binned by rounding pseudotime values, then a Wilcoxon rank sum test was used to compare module scores between genotypes to determine significance, and the $-\log$ p-value was graphed by barplot and shown over the corresponding pseudotime. These plots were arranged with ggpubr version 0.2.2 grid.arrange.

Immunofluorescence on frozen sections and quantification of CD45.1/CD45.2 positivity within germinal centers

Spleens (n=3 per genotype) were fixed with 4% paraformaldehyde overnight, cryopreserved in sucrose gradient (10%, 20%, 30%) and embedded in O.C.T. and frozen down for cryostat sectioning. Cryostat sections (7micron) were used for antibody (1:500 dilution, CD45.1 (AF594), CD45.2 (AF647), PNA-Cy3) staining overnight. Tissue sections were imaged at 40x using Zeiss LSM 880 Airyscan confocal microscope with 7 laser excitations at 405, 458, 488, 514, 594, 561, and 633.

To quantify the abundance of CD45.1 and CD45.2 positivity within germinal centers (GCs), a mask was generated for each image based on manual delineations of PNA positive regions within the tissue section. These masks were then used as regions of interest (ROIs) and projected onto the CD45.1 and CD45.2 channels to classify the GC regions, with every pixel outside the defined GC region set to a value of zero. The ROI was then linearized (2D image matrix into a 1D image vector), and the intensity levels per pixel of the CD45.1 and CD45.2 channels were recorded for each corresponding pixel within the ROI. We then plotted the intensity distributions and appropriate thresholds were set to determine the number of positive pixels, which was then visually confirmed. Briefly, thresholds were set to ensure that each pixel classified as CD45.1 positive was also CD45.2 negative, and vice versa. Based on these thresholds per channel, we determined the number of positive cells for CD45.1 and CD45.2, then calculated the positivity fraction as the ratio within each GC for every recorded image.

Western blot

For histone immunoblots, murine GC B cells were MACS-sorted (PNA enrichment, PNA MicroBead Kit, Miltenyi Biotec, 130110479) and lysed in 2x Laemmli buffer (BioRad #1610737). An equivalent of 50,000 cells was run per lane, using NuPage 4-20% Tris-Glycine gels, transferred to 0.2um PVDF membrane, blocked with 5% non-fat milk in TBS-Tween20 buffer and incubated overnight with primary antibodies at 1:1000 dilution (D4J5Q, anti-H1E, Cell-Signaling 41328; AE-4, anti-H1, Millipore 05-457; C36B11 anti-H3 K27me3, Cell Signaling 9733; C75H12 anti-H3 K36me2, Cell Signaling 2901; AC22 anti-Ezh2, Cell Signaling 3147; 29D1 anti-NSD2, Millipore Mabe191), washed and incubated with appropriate HRP-conjugated secondary at 1:10,000 dilution, developed using Immobilon substrate (Millipore, P90720) and imaged using Amersham 600 imager.

For KLF5 immunoblots, murine B220+ tumor cells were isolated from two independent groups of age-matched terminal *VavP-Bcl2; H1c^{+/-}/e^{+/-}* and *VavP-Bcl2; H1wild-type* mice. Cells were lysed in 2x laemmli buffer (BioRad #1610737). An equivalent of 500,000 cells was run per lane, using NuPage 4-20% Tris-Glycine gels, transferred to 0.2um PVDF membrane, blocked with 5% non-fat milk in TBS-Tween20 buffer and incubated overnight with primary antibody at 1:1000 dilution (anti-KLF5, rabbit polyclonal, Cat: 21017-1-AP) and 1:1000 (anti-GAPDH, mouse monoclonal 6C5, Abcam ab8245), washed and incubated with appropriate HRP-conjugated secondary at 1:10,000 dilution, developed using Immobilon substrate (Millipore, P90720) and imaged using Amersham 600 imager. Uncropped western blots can be found at Supplementary Information (**Supplementary Figure 1**).

Fluorescence recovery after photobleaching

Fluorescence recovery after photobleaching (FRAP) was performed in NIH 3T3 cells, grown in 35 mm glass bottom plates (Ibidi 81158), and transiently transfected with wild type or mutant human H1 C-terminally tagged with mEGFP (mEGFP-N1, a gift from Michael Davidson, Addgene plasmid #54767) using Lipofectamine LTX (Invitrogen 15338100) according to manufacturer instructions. FRAP was performed at 48 hours

post-transfection, using LSM 780 AxioObserver equipped with 63x/1.40 PlanApochromat objective and heated stage with 5% CO₂. After acquiring 2-5 pre-bleach images, an area was bleached with 30-50 laser bursts at 100% intensity, followed by 2-5 min of continuous imaging at sub-1 sec intervals. Collected images were processed and analyzed using FIJI and FRAP Profiler plugin to generate the recovery curves. Between 10 and 20 individual cells were imaged for each construct.

General Materials and Methods and Linker Histone Purification

Biochemicals reagents were purchased from Fisher Scientific or Sigma-Aldrich Corporation unless otherwise stated. T4 DNA ligase, DNA polymerase and restriction enzymes were obtained from New England BioLabs. Primer synthesis and DNA sequencing were performed by Integrated DNA Technologies and Genewiz, respectively. PCR amplifications were performed on a Bio-Rad T100™ Thermal Cycler. Centrifugal filtration units and MINI dialysis units purchased from Fisher Scientific. Size exclusion chromatography was performed on an AKTA FPLC system from GE Healthcare equipped with a P-920 pump and UPC-900 monitor. Sephacryl S-200 columns were obtained from GE Healthcare. Biolayer interferometry measurements were taken on an Octet Red96e system (PALL/ForteBio). UV spectrometry was performed on NanoDrop 2000c (Thermo Scientific). Semi-preparative reversed-phase HPLC (RP-HPLC) was performed on an Agilent 1200 series instrument with an Agilent C18 column (12 μm, 10 mm × 250 mm), employing 0.1% TFA in water (HPLC solvent A), and 90% acetonitrile, 0.1% TFA in water (HPLC solvent B), as the mobile phases. Gradients were 0-70% HPLC buffer B over 65 minutes at a flow rate of 2 mL/min, unless stated otherwise. Preparative scale purifications were conducted on an Agilent LC system using an Agilent C18 preparative column (15-20 μm, 20 × 250 mm) employed at a flow rate of 20 mL/min. HPLC Electrospray ionization MS (HPLC-ESI-MS) analysis was performed on an Agilent 6120 Quadrupole LC/MS spectrometer (Agilent Technologies). All wildtype and mutant linker histones were purified as previously³² with minor adjustments. Briefly, the His-SUMO-H1-GyrA-His construct was grown in Rosetta DE3 cells at 37 °C under ampicillin and chloramphenicol selections until OD600 reached 0.6, followed by the induction of protein expression with 0.5 mM IPTG overnight at 16 °C. Bacteria were harvested by centrifugation and lysed by sonication following resuspension. Lysate was clarified by centrifugation and incubated with Ni-NTA beads for 1 hour. DTT was added to a final concentration of 1 mM, and the elution was treated with 1:100 v/v of Ulp-1 enzyme for one hour. Next, beta-mercaptoethanol (β-ME) was added to a final concentration of 500 mM for an additional four-hour treatment at room temperature to facilitate intein cleavage. Solid urea was added to the cleavage mixture, and dialysis was performed to remove the β-ME, followed by incubation with Ni-NTA beads to remove the cleaved tags. After adjustment of the flow-through to pH 9, cation exchange chromatography was performed under denaturing conditions on a HiTrap HP column with a NaCl gradient from 200 mM to 750 mM. Fractions containing full-length H1 were pooled and purified by semi-preparative RP-HPLC. Final linker histone concentration was determined with the absorbance at 214 nm and extinction coefficient³³.

Recombinant histone purification

Wildtype human histones were purified from *E. coli* as previously described³⁴. Briefly, BL-21 cells transformed with the relevant plasmid were grown under ampicillin selection at 37 °C until OD₆₀₀ = 0.6 and then induced with 0.5 mM final IPTG at 37 °C for 3 hr. Following centrifugation, bacterial pellets were resuspended in and lysed by sonication. Lysates were cleared by a 20-minute spin at 17,000 x *g* at 4° C. Supernatants were discarded and proteins were extracted from the inclusion bodies with extraction buffer (20 mM Tris pH 7.5, 6 M guanidine hydrochloride, 1 mM EDTA, 100 mM NaCl) under agitation for one hour at room temperature. The suspension was centrifuged at 30,000 xg for 30 minutes followed by filtration through a 0.45 µm filter. The histones were purified on a preparative C-18 RP-HPLC on a gradient of 30 - 70 % HPLC solvent B. Pure fractions were analyzed by RP-HPLC and their masses were verified by ESI-MS.

Histone octamer formation

Assembly of histones into a protein octamer was performed as described³⁴. Each purified lyophilized histone was dissolved in unfolding buffer (20 mM Tris pH 7.5, 6 M guanidine hydrochloride, 1 mM DTT) and A280 measurements were taken to determine concentration. The histones were combined in equimolar ratios with a 5 % excess of H2A and H2B in order to prevent H3-H4 dimer formation. The final combined histone concentration was adjusted to 1 mg/ml and the mixture was dialyzed against refolding buffer (10 mM Tris pH 7.5, 2 M NaCl, 1 mM EDTA, 1 mM DTT) for 8 hours. The dialysis was exchanged two subsequent times. After recovery, the sample was spun down at 17,000 xg for 10 minutes at 4 °C. The supernatant was concentrated to 250 µL on a 10 kDa concentrator, followed by injection onto a S200 10/300 size exclusion column on an AKTA FPLC. The octamer containing fractions were pooled and concentrated on a 10 kDa MWCO centrifugal filter unit. Glycerol was added to a final concentration of 50 % (v/v), and the octamer concentration was determined by A280.

Linker DNA preparation

A 30 bp linker DNA fragment was added to the 3' end of the nucleosome positioning sequence ('601')³⁵ by PCR amplification, as described previously³². The DNA was purified and a DNA-purification column, lyophilized, resuspended in water and quantified by A260. A biotinylated primer was used to generate 5' linker DNA for biolayer interferometry experiments.

Nucleosome assembly

Nucleosome assembly was performed as described before with minor changes³⁶. The assembly component were mixed at 2 µM concentration in 10 mM scale in 10 µL Tris pH 7.5, 1 mM EDTA, 1 mM DTT, 2 M NaCl. Assemblies were placed at 37 °C for 15 minutes after which they were transferred to 30 °C. Dilution buffer (10 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT, 10 mM NaCl) was then added at 15-minute intervals in the following volumes (µL): 6.7, 5, 3.6, 4.7, 6.7, 10, 30, 20, 100. The correct DNA:octamer ratio for each octamer preparation was determined empirically. Assembly quality was

determined by running reactions on a 5 % acrylamide native gel. Suitable assemblies were pooled and quantified by A260.

Nucleosome-array reconstitution

12-mer nucleosome arrays were assembled as described before³⁷. Briefly, 12-mer DNA (containing 12 repeats of the 601 sequences with 30 bp of linker DNA between them), octamers, mouse mammary tumor virus buffer DNA (MMTV), and linker histones were combined in a high salt buffer (10 mM Tris pH 7.5, 0.1 mM EDTA, 2 M NaCl, 1 mM DTT). Reconstitutions were performed at 0.5 μ M 601 sites in 50 μ L. As with nucleosomes, DNA:octamer ratios were determined empirically. Linker histones were used in a 1.5 molar excess relative to 601 sites in order to ensure full saturation. The samples were dialyzed at 4 °C into 1.4 M TEN buffer (10 mM Tris pH 7.6, 0.1 mM EDTA, 1.4 M NaCl, 1 mM DTT) for one hour. Using a peristaltic pump at a flow rate of 2.5 mL/min, the arrays were diluted into 0.5 M TEN buffer (10 mM Tris pH 7.6, 0.1 mM EDTA, 0.5 M NaCl, 1 mM DTT) over the course of seven hours. Dialysis into the final low salt buffer (10 mM Tris pH 7.6, 0.1 mM EDTA, 10 mM NaCl, 1 mM DTT) was performed overnight. Assembled arrays were purified from unassembled materials as described previously⁶. The concentration of arrays was determined by A260.

Magnesium Precipitation

Magnesium precipitation was used to determine the compaction state of the nucleosomes arrays as previously reported³⁷. Briefly, purified arrays were treated with MgCl₂ in 0.25 mM increments starting at 0.25 mM. They were then centrifuged at 17,000 xg for 10 minutes at 4 °C. The A260 of the supernatant of the supernatant was measured following each treatment. Each Mg²⁺ concentration treatment was performed in triplicate. We generated 12-mer nucleosome arrays, exposed them to WT or C-terminal mutant H1 P118S, and then performed Mg²⁺ precipitation. This assay functions as a readout for fiber compaction, which is reflected in their propensity to oligomerize in the presence of divalent cations.

Biolayer Interferometry

Biolayer interferometry was performed as described earlier³². Briefly, 5' biotinylated mononucleosomes with 30bp linker DNA, nucleosome core particle (NCPs), were diluted to a concentration of 1 μ g/m and immobilized on streptavidin-functionalized optical sensors. Linker histones were diluted in a 2-fold series starting at 10 nM. Assays were performed at room temperature with a 180 s association phase and a 1200 s disassociation phase. All experimental samples were referenced against streptavidin sensors in buffer without linker histone added. Data was analyzed with the manufacturer's data analysis software version (Octet 11.0) with estimation of kinetic parameters made using curve fitting to a 1:1 model. The kinetic data reported was derived from a global fitting of replicates at five protein concentrations, with standard error determined by the analysis software. Residual plots for the fitted curves, as well as the R² and χ^2 values of each fit were used to determine the goodness of fit. Mononucleosomes represent a homogeneous substrate and each nucleosome contains only one H1 binding site.

We used bilayer interferometry to evaluate the affinity of H1 mutants for mononucleosomes.

Atomic Force Microscopy

Atomic Force Microscopy imaging of 12-mer nucleosomal arrays with wildtype or mutant H1 was adapted from a previous method³². Briefly, harvested 12-mer arrays were diluted to a concentration of 0.6 ng/ul in 10 mM TEN buffer. 1 molar equivalent of H1 was added and incubated at room temperature for 15 minutes. 20 uL of the mixture was placed on an AP-mica slide then incubated for 10 minutes, washed with water, and dried. Atomic Force Microscopy measurements were taken on an Asylum Research MFP-3D-BIO instrument with Zeiss Axio Observer Inverted Optical Microscope. Cantilever calibration and height profiles were generated using the Asylum Research software package Version IX (AR Software). The AFM images display compact fibers in the presence of WT H1 and looser fibers in the presence of mutant H1.

Histone preparation for mass spectrometry.

Cells were sorted as described above and 10^5 events were collected and prepared according to Camarillo et al³⁸. Briefly, sorted cells were collected into 2 N H_2SO_4 , allowing for direct extraction of histones. Cellular debris was removed by centrifugation at 4,000 x g for 5 min and histones were precipitated from the supernatant with trichloroacetic acid at a final concentration of 20% (v/v) overnight at 4 °C. Histones were pelleted with centrifugation at 10,000 x g for 5 min, washed once with 0.1% HCl in acetone, then 100% acetone with centrifugation at 15,000 x g for 5 min. Pellets were dried briefly in a fume hood and stored at -80 °C until further processing. Derivatization and digestion were modified from Garcia et al³⁹. Dried histones were resuspended in 50 mM ammonium bicarbonate (10 μ L). Sodium hydroxide (5 μ L) then propionic anhydride (20 μ L, 1:3 dilution in isopropanol) was added to the histone solution and adjusted to pH 8 with additional sodium hydroxide. Samples were incubated at 52°C for 1 h before drying to completion in a SpeedVac concentrator. Propionylated histones were resuspended in 50 mM ammonium bicarbonate and digested for 16 h with 0.5 μ g trypsin. Digests were dried in a Speedvac concentrator and subjected to a final propionylation as described above.

Targeted mass spectrometry and quantification of histone modifications

LC-MS solvents and sulfuric acid were acquired from FisherScientific (Hampton, NH). Trichloroacetic acid, ammonium bicarbonate, propionic anhydride, and isopropanol were purchased from Sigma-Aldrich (St. Louis MO). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Samples were resuspended in water with 0.1% TFA and analyzed by nano-LC (Dionex, Sunnyvale, CA) on a triple quadrupole mass spectrometer (TSQ Quantiva, ThermoFisher Scientific). Peptides were loaded on a trapping column (3 cm \times 150 μ m, packed with ProntoSIL C18-AQ, 3 μ m, 200Å resin (New Objective, Woburn, MA)) with water with 0.1% TFA for 10 min at 2.5 μ L/min. The peptides were eluted at 0.30 μ L/min from the trapping and PicoChip analytical column (10 cm \times 75 μ m packed with ProntoSIL C18-AQ, 3 μ m, 200 Å resin

(New Objective)) over a 45 min gradient from 1 to 35% Nano Pump Solvent B (95% acetonitrile with 0.1% formic acid; Nano Pump Solvent A, water with 0.1% formic acid). Ions were produced by electrospray from a 10 μm emitter tip and introduced into the mass spectrometer with the following settings: collision gas pressure of 1.5 mTorr; Q1 peak width of 0.7 (FWHM); cycle time of 3 s; skimmer offset of 10 V; electrospray voltage of 2.5 kV. All injections were performed in technical triplicate. Targeted analysis of unmodified and various modified histone peptides was performed with transitions specific to each peptide species as described previously. Raw MS files were analyzed with Skyline (v4.1) using Savitzky-Golay smoothing and peak area assignments were manually assessed. The percent relative abundance of each histone PTM was calculated from the total peak areas exported from Skyline.

Mesoscale model of chromatin fibers

To explore the effect of linker histone (LH) density on chromatin organization at the nucleosome level, we perform Monte Carlo simulations of 50-nucleosome fibers with a living-system-like DNA linker length distribution, different LH densities, and two LHs, H1E and the shorter H1C. In particular, the chromatin fibers contain the following nucleosome repeat length (NRL) distribution: 30% 173 bp, 18% 182 bp, 14% 191 bp, 12% 200 bp, 8% 209 bp, 8% 218 bp, and 10% 226 bp, which is based on mouse embryonic stem cells data obtained by chemical mapping⁴⁰. The NRLs are randomly distributed along the fiber. LH densities correspond to 0, 0.25, 0.50, 0.75, and 1 LH per nucleosome, fulfilled by distributing LHs randomly along the fiber.

Briefly, in our chromatin mesoscale model the nucleosome core with wrapped DNA is represented as a rigid body with ~ 300 Debye-Hückel charges; linker DNA is treated with a combined worm-like chain and bead model of ~ 9 bp resolution in which each bead has a salt-concentration dependent charge obtained by the Stigter method⁴¹; flexible histone tails are coarse grained as 5 amino acids per bead with the Levitt-Warshel united-atom bead model⁴² and charges computed by our DISCO algorithm⁴³, and linker histones H1E⁴⁴ and H1C⁴⁵ are coarse grained similarly with 6 beads for the globular head and 22 and 21 beads for the C-terminal domain, respectively, with charges also obtained by DISCO. Full details of the model components can be found in references⁴⁴⁻⁴⁷.

The total energy function of the model incorporates stretching terms for linker DNA, histone tails, and linker; bending terms for linker DNA, histone tails, and linker histone; twisting terms for linker DNA; electrostatic Debye-Hückel terms to represent all charge-charge interactions within chromatin; and excluded volume terms for all beads, described with Lennard-Jones potentials. Conformational sampling of the fibers is performed with Monte Carlo (MC) simulations with tailored local and global moves. These include local translation, local rotation, and global pivot rotations for linker DNA and nucleosomes, local translation for the C-terminal domain of the LHs, and regrowth for histone tails. Acceptance of the first four moves is based on the Metropolis⁴⁸ criteria

and acceptance of tails regrowth is based on Rosenbluth¹¹ criteria. Full details regarding the energy terms and MC sampling can be found in references^{46,49}.

Each system is simulated with 10 independent trajectories of 40 million MC steps, defined with different initial random number seeds and different DNA twist value of -12° , 0° , or 12° to mimic natural variations of the B-DNA twist⁵⁰. The last 10 million steps of each trajectory are used for analysis, corresponding to a 1000-configuration ensemble per system.

Internucleosome contact maps describe the fraction of MC steps that a nucleosome i is in “contact” with a nucleosome j . Contacts are defined by proximity, namely, if the core or tails of both cores become closer than 1.8 nm, it is assumed that a contact is established. Contacts are counted along the trajectory and normalized by the maximum number of contacts. Contacts located in the diagonal of the maps correspond to short-range interactions, typical of the canonical zigzag topology. Contacts near the diagonal correspond to short-range interactions (i.e., sequentially proximal nucleosomes are in close contact), and regions perpendicular to the main diagonal correspond to hairpin loops and kinks. Regions parallel to the main diagonal correspond to hierarchical loops⁵¹.

1. The internucleosome contact maps are decomposed into one-dimensional plots to determine the magnitude of $i, i \pm k$ interactions as:

$$I(k) = \frac{\sum_{i=1}^{N_C} I'(i, i \pm k)}{\sum_{j=1}^{N_C} I(j)},$$

2. Fiber packing ratio is calculated as the number of nucleosomes contained in 11 nm of fiber, as:

$$\text{Packing ratio} = \frac{11 * \text{number of cores}}{\text{fiber length}},$$

where fiber length is calculated using a cubic smoothing spline function native from Matlab.

3. Fiber volume is calculated assuming a cylindrical geometry for the fibers, as:

$$\text{Fiber volume} = \text{fiber radius}^2 * \pi * \text{fiber length},$$

where fiber radius is calculated as the distance between the nucleosome center and the fiber axis. The fiber axis is expressed as a 3D parametric curve:

$$\mathbf{r}^{\text{ax}}(s) = (r1^{\text{ax}}(s), r2^{\text{ax}}(s), r3^{\text{ax}}(s)),$$

where each component ($rj^{\text{ax}}(s)$, $j = 1, 2, 3$) is a nonlinear function fitted to the positions of nucleosome centers to one spatial dimension (x, y, or z).

The fiber axis is fitted into a polynomial form:

$$r_j^{ax}(s) \approx P_j(i) = p_1^n(s) + p_2^{n+1}(s) + \dots + p^n(s),$$

where the polynomial degree is chosen using the Matlab polyfit function so that the fiber axis fits a standard least-squares fitting procedure.

From the fiber axis, the local fiber radius is defined for a given nucleosome as the perpendicular distance between a nucleosome center and its closest linear fiber axis segment plus the nucleosome radius (5.5 nm). The fiber radius is obtained by averaging over all local fiber radii.

Genomic DNA and RNA Extraction

Murine gDNA was extracted using the Puregene Gentra cell kit (Qiagen) and eluted in water. The quality of purified DNA was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies). Total RNA was extracted from murine GC B cells or tumors using TRIzol (Life Technologies) and RNeasy Mini Kit (Qiagen) with DNase treatment. RNA concentration was determined using Qubit (Life Technologies), and integrity was verified using Agilent 2100 Bioanalyzer

Quantitative Real-Time PCR

cDNA synthesis from RNA was performed using the Verso cDNA Synthesis kit (Thermo Scientific). Hist1H1B-E expression was detected using the Green FastMix kit (Applied Biosystems) with human and mouse primers (**Supplementary table 2**) and ran on a QuantStudio6 Flex Real-Time PCR System (Applied Biosystems). Gene expression was normalized to housing keeping gene using the $\Delta C(t)$ method, and results were represented as mRNA expression.

Supplementary Table 1: Antibodies used in FACS experiments

Surface Marker	Fluorochrome	Clone	Manufacturer	Cat #
IgM	FITC	II-41	BD Biosciences	553437
GL7	FITC	GL7	BD Biosciences	553666
CD38	FITC	90	BD Biosciences	558813
CD21	FITC	eBio4E3	eBioscience	11-0121-81
IgD	PE	11-26c.2a	BD Biosciences	558597

CXCR4 (CD184)	PE	2B11	INVITROGEN	12-9991-82
CD23	PE	B3B4	eBioscience	12-0232-81
CD95 (FAS)	PerCP-Cy5.5	SA367H8	Biolegend	152610
CD45.1	PerCP-Cy5.5	A20	eBioscience	45-0453-82
CD45.2	PerCP-Cy5.5	104	Biolegend	109828
CD45.2	PE-Cy7	104	eBioscience	25-0454-82
CD86	PE-Cy7	GL1	Biolegend	105014
CD95 (FAS)	PE-Cy7	Jo2	BD Biosciences	557653
B220 (CD45R)	PE-Cy7	ra3-6b2	eBioscience	25-0452-82
IgG1	APC	A85-1	BD Biosciences	560089
CD86	APC	GL1	Biolegend	105012
Annexin	APC		BD Biosciences	550475
GL7	APC	GL7	BD Biosciences	561529
CD138	APC	281-2	BD Biosciences	558626
CD38	APC-CY7	90	Biolegend	102728
CD45.2	APC-CY7	104	Biolegend	109824
B220 (CD45R)	APC-CY7	ra3-6b2	Biolegend	103224
CD138	BV-421	281-2	Biolegend	142508
CD95 (FAS)	BV-421	Jo2	BD Biosciences	562633
B220 (CD45R)	eFluor 506	ra3-6b2	eBioscience	69-0452-82

Supplementary Table 2: QPCR primers used in this study

Human	Forward qPCR primer	Reverse qPCR primer
<i>HIST1H1B</i>	CAACCAAGAGTCCTGCCAAG	GGCTTTGTTGCGGTTTTTCAC
<i>HIST1H1C</i>	ACACCGAAGAAAGCGAAGAA	GCTTGACAACCTTGGGCTTA
<i>HIST1H1D</i>	TCAACAAGAAAGCGGCTTCC	CACTTTCTTGGTCCCAGCAG
<i>HIST1H1E</i>	CCCCAAAGAAGGCGAAGAAG	TTGGCGGTCTTTGGTTTAGC
<i>RPL13A</i>	AGAAAAAGCGGATGGTGGTT	CTCCGGTAGTGGATCTTGG

Mouse	Forward qPCR primer	Reverse qPCR primer
<i>Hist1h1B</i>	TGTAGAGAAGTCTCCCGCCA	GCGCTCCTTAGAGGCAGAAA
<i>Hist1h1C</i>	AGAAGGCGAAGGTCACCAAG	GGGGAGGCAGCCTACTTTTT
<i>Hist1h1D</i>	AAGCCTAAGAAGGCGACTGG	CTTGGCTGGACTCTTTGCTG
<i>Hist1h1E</i>	CTCTCTCCTCACACGCTTCG	GCCTTGGTGATGAGTTCGGA
<i>GAPDH</i>	CCAGCCTCGTCCCGTAGAC	GCCTTGACTGTGCCGTTG

Supplementary Table 3. CUT&RUN Sequencing Statistics

GCB-cell nuclei	Target	Antibody Vendor	Antibody Lot	Total PF clusters	Uniquely aligned PE pairs	Unique align rate
<i>wt_rep1</i>	IgG	EpiCypher 13-0042	20036001-52	4,981,248	4,379,833	88%
<i>wt_rep1</i>	H3K4me3	EpiCypher 13-0041	20010001	3,815,224	3,197,328	84%
<i>wt_rep1</i>	H3K27me3	EpiCypher 13-0030	18303001	5,838,200	5,210,429	89%
<i>wt_rep1</i>	H3K9me2	Millipore 05-1249	3170817	6,981,537	6,194,310	89%
<i>wt_rep1</i>	H3K9me3	Abcam 176916	GR3218257-2	7,617,674	6,793,822	89%
<i>wt_rep2</i>	IgG	EpiCypher 13-0042	20036001-52	5,572,458	5,071,683	91%
<i>wt_rep2</i>	H3K4me3	EpiCypher 13-0041	20010001	5,541,738	4,971,008	90%
<i>wt_rep2</i>	H3K27me3	EpiCypher 13-0030	18303001	5,756,281	5,157,595	90%
<i>wt_rep2</i>	H3K9me2	Millipore 05-1249	3170817	6,448,278	5,811,433	90%
<i>wt_rep2</i>	H3K9me3	Abcam 176916	GR3218257-2	9,044,632	8,003,716	88%
<i>H1c^{-/-}e^{-/-}_rep1</i>	IgG	EpiCypher 13-0042	20036001-52	5,190,903	4,743,897	91%
<i>H1c^{-/-}e^{-/-}_rep1</i>	H3K4me3	EpiCypher 13-0041	20010001	4,812,484	4,370,294	91%
<i>H1c^{-/-}e^{-/-}_rep1</i>	H3K27me3	EpiCypher 13-0030	18303001	6,948,155	6,188,713	89%
<i>H1c^{-/-}e^{-/-}_rep1</i>	H3K9me2	Millipore 05-1249	3170817	7,147,881	6,407,096	90%
<i>H1c^{-/-}e^{-/-}_rep1</i>	H3K9me3	Abcam 176916	GR3218257-2	7,585,753	6,757,301	89%
<i>H1c^{-/-}e^{-/-}_rep2</i>	IgG	EpiCypher 13-0042	20036001-52	5,400,284	4,738,758	88%
<i>H1c^{-/-}e^{-/-}_rep2</i>	H3K4me3	EpiCypher 13-0041	20010001	5,854,955	5,196,804	89%
<i>H1c^{-/-}e^{-/-}_rep2</i>	H3K27me3	EpiCypher 13-0030	18303001	7,685,593	6,725,283	88%
<i>H1c^{-/-}e^{-/-}_rep2</i>	H3K9me2	Millipore 05-1249	3170817	6,496,731	5,889,551	91%
<i>H1c^{-/-}e^{-/-}_rep2</i>	H3K9me3	Abcam 176916	GR3218257-2	8,276,380	7,262,918	88%
Average=				6,349,819	5,653,589	89%

METHODS REFERENCES

- 1 Fan, Y. *et al.* H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Mol Cell Biol* **23**, 4559-4572 (2003).
- 2 Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250-264, doi:10.1016/j.cell.2008.03.028 (2008).
- 3 Garcia, M. *et al.* Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Res* **9**, 63, doi:10.12688/f1000research.16665.1 (2020).
- 4 Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**, 591-594, doi:10.1038/s41592-018-0051-x (2018).
- 5 Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222, doi:10.1093/bioinformatics/btv710 (2016).
- 6 Shamsani, J. *et al.* A plugin for the Ensembl Variant Effect Predictor that uses MaxEntScan to predict variant spliceogenicity. *Bioinformatics* **35**, 2315-2317, doi:10.1093/bioinformatics/bty960 (2019).
- 7 Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* **17**, 31, doi:10.1186/s13059-016-0893-4 (2016).
- 8 Chapuy, B. *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* **24**, 679-690, doi:10.1038/s41591-018-0016-8 (2018).
- 9 Maura, F. *et al.* A practical guide for mutational signature analysis in hematological malignancies. *Nat Commun* **10**, 2969, doi:10.1038/s41467-019-11037-8 (2019).
- 10 Kasar, S. *et al.* Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. *Nat Commun* **6**, 8866, doi:10.1038/ncomms9866 (2015).
- 11 Jiang, Y. *et al.* CREBBP Inactivation Promotes the Development of HDAC3-Dependent Lymphomas. *Cancer Discov* **7**, 38-53, doi:10.1158/2159-8290.CD-16-0975 (2017).
- 12 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
- 13 Pearson, T. A. & Manolio, T. A. How to interpret a genome-wide association study. *JAMA* **299**, 1335-1344, doi:10.1001/jama.299.11.1335 (2008).
- 14 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 15 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297, doi:10.1093/nar/gks042 (2012).

- 16 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).
- 17 Schep, A. N. *et al.* Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Res* **25**, 1757-1770, doi:10.1101/gr.192294.115 (2015).
- 18 Xu, S., Grullon, S., Ge, K. & Peng, W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods Mol Biol* **1150**, 97-111, doi:10.1007/978-1-4939-0512-6_5 (2014).
- 19 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**, 1006-1019, doi:10.1038/nprot.2018.015 (2018).
- 20 Shah, R. N. *et al.* Examining the Roles of H3K4 Methylation States with Systematically Characterized Antibodies. *Mol Cell* **72**, 162-177 e167, doi:10.1016/j.molcel.2018.08.015 (2018).
- 21 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 22 Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354, doi:10.1038/s41598-019-45839-z (2019).
- 23 Lazaris, C., Kelly, S., Ntziachristos, P., Aifantis, I. & Tsirigos, A. HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. *BMC Genomics* **18**, 22, doi:10.1186/s12864-016-3387-6 (2017).
- 24 Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat Methods* **9**, 999-1003, doi:10.1038/nmeth.2148 (2012).
- 25 Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* **27**, 1939-1949, doi:10.1101/gr.220640.117 (2017).
- 26 Zheng, X. & Zheng, Y. CscoreTool: fast Hi-C compartment analysis at high resolution. *Bioinformatics* **34**, 1568-1570, doi:10.1093/bioinformatics/btx802 (2018).
- 27 Xie, Y., Pan, W. & Khodursky, A. B. A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics* **21**, 4280-4288, doi:10.1093/bioinformatics/bti685 (2005).
- 28 Raviram, R. *et al.* 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. *PLoS Comput Biol* **12**, e1004780, doi:10.1371/journal.pcbi.1004780 (2016).
- 29 Di Giammartino, D. C. *et al.* KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat Cell Biol* **21**, 1179-1190, doi:10.1038/s41556-019-0390-6 (2019).
- 30 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

- 31 Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477, doi:10.1186/s12864-018-4772-0 (2018).
- 32 Osunsade, A. *et al.* A Robust Method for the Purification and Characterization of Recombinant Human Histone H1 Variants. *Biochemistry* **58**, 171-176, doi:10.1021/acs.biochem.8b01060 (2019).
- 33 Kuipers, B. J. & Gruppen, H. Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis. *J Agric Food Chem* **55**, 5445-5451, doi:10.1021/jf070337l (2007).
- 34 Fierz, B. *et al.* Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat Chem Biol* **7**, 113-119, doi:10.1038/nchembio.501 (2011).
- 35 Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**, 19-42, doi:10.1006/jmbi.1997.1494 (1998).
- 36 White, A. E., Hieb, A. R. & Luger, K. A quantitative investigation of linker histone interactions with nucleosomes and chromatin. *Sci Rep* **6**, 19122, doi:10.1038/srep19122 (2016).
- 37 Debelouchina, G. T., Gerecht, K. & Muir, T. W. Ubiquitin utilizes an acidic surface patch to alter chromatin structure. *Nat Chem Biol* **13**, 105-110, doi:10.1038/nchembio.2235 (2017).
- 38 Camarillo, J. M. *et al.* Coupling Fluorescence-Activated Cell Sorting and Targeted Analysis of Histone Modification Profiles in Primary Human Leukocytes. *J Am Soc Mass Spectrom* **30**, 2526-2534, doi:10.1007/s13361-019-02255-x (2019).
- 39 Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat Protoc* **2**, 933-938, doi:10.1038/nprot.2007.106 (2007).
- 40 Voong, L. N. *et al.* Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell* **167**, 1555-1570 e1515, doi:10.1016/j.cell.2016.10.049 (2016).
- 41 Stigter, D. Interactions of highly charged colloidal cylinders with applications to double-stranded. *Biopolymers* **16**, 1435-1448, doi:10.1002/bip.1977.360160705 (1977).
- 42 Levitt, M. & Warshel, A. Computer simulation of protein folding. *Nature* **253**, 694-698, doi:10.1038/253694a0 (1975).
- 43 Zhang, Q., Beard, D. A. & Schlick, T. Constructing irregular surfaces to enclose macromolecular complexes for mesoscale modeling using the discrete surface charge optimization (DISCO) algorithm. *J Comput Chem* **24**, 2063-2074, doi:10.1002/jcc.10337 (2003).
- 44 Luque, A., Collepardo-Guevara, R., Grigoryev, S. & Schlick, T. Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. *Nucleic Acids Res* **42**, 7553-7560, doi:10.1093/nar/gku491 (2014).

- 45 Perisic, O., Portillo-Ledesma, S. & Schlick, T. Sensitive effect of linker histone binding mode and subtype on chromatin condensation. *Nucleic Acids Res* **47**, 4948-4957, doi:10.1093/nar/gkz234 (2019).
- 46 Bascom, G. D., Myers, C. G. & Schlick, T. Mesoscale modeling reveals formation of an epigenetically driven HOXC gene hub. *Proc Natl Acad Sci U S A* **116**, 4955-4962, doi:10.1073/pnas.1816424116 (2019).
- 47 Bascom, G. D. & Schlick, T. Chromatin Fiber Folding Directed by Cooperative Histone Tail Acetylation and Linker Histone Binding. *Biophys J* **114**, 2376-2385, doi:10.1016/j.bpj.2018.03.008 (2018).
- 48 Metropolis, N. & Ulam, S. The Monte Carlo method. *J Am Stat Assoc* **44**, 335-341, doi:10.1080/01621459.1949.10483310 (1949).
- 49 Arya, G. & Schlick, T. A tale of tails: how histone tails mediate chromatin compaction in different salt and linker histone environments. *J Phys Chem A* **113**, 4045-4059, doi:10.1021/jp810375d (2009).
- 50 Drew, H. R. & Travers, A. A. DNA bending and its relation to nucleosome positioning. *J Mol Biol* **186**, 773-790, doi:10.1016/0022-2836(85)90396-1 (1985).
- 51 Grigoryev, S. A. *et al.* Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. *Proc Natl Acad Sci U S A* **113**, 1238-1243, doi:10.1073/pnas.1518280113 (2016).