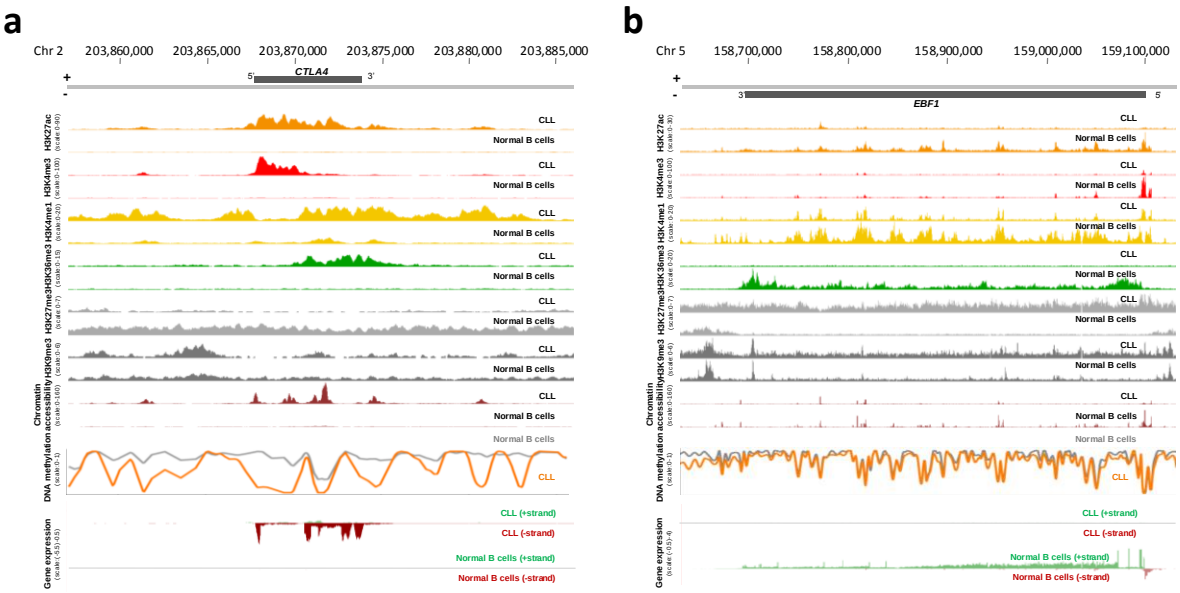
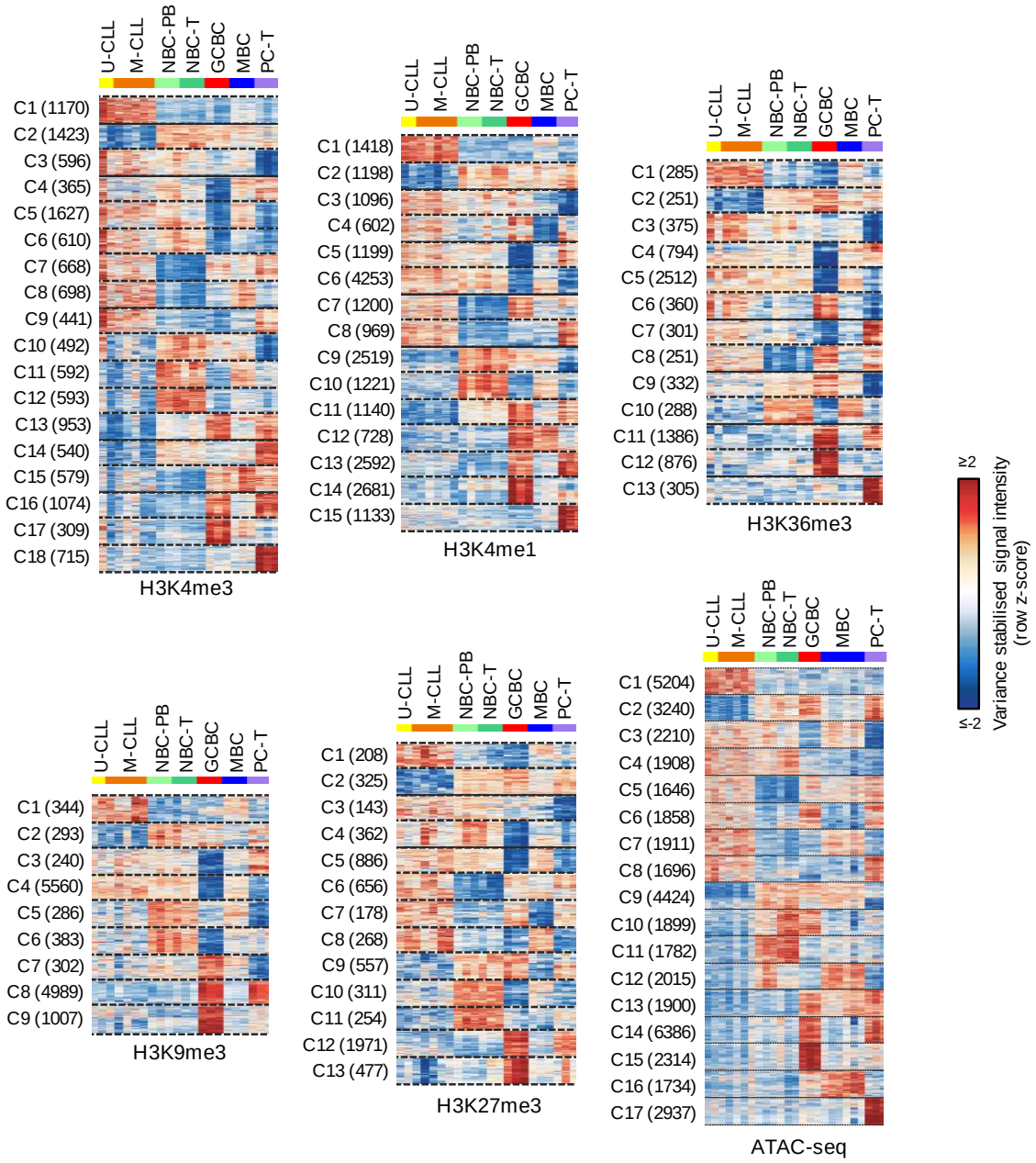
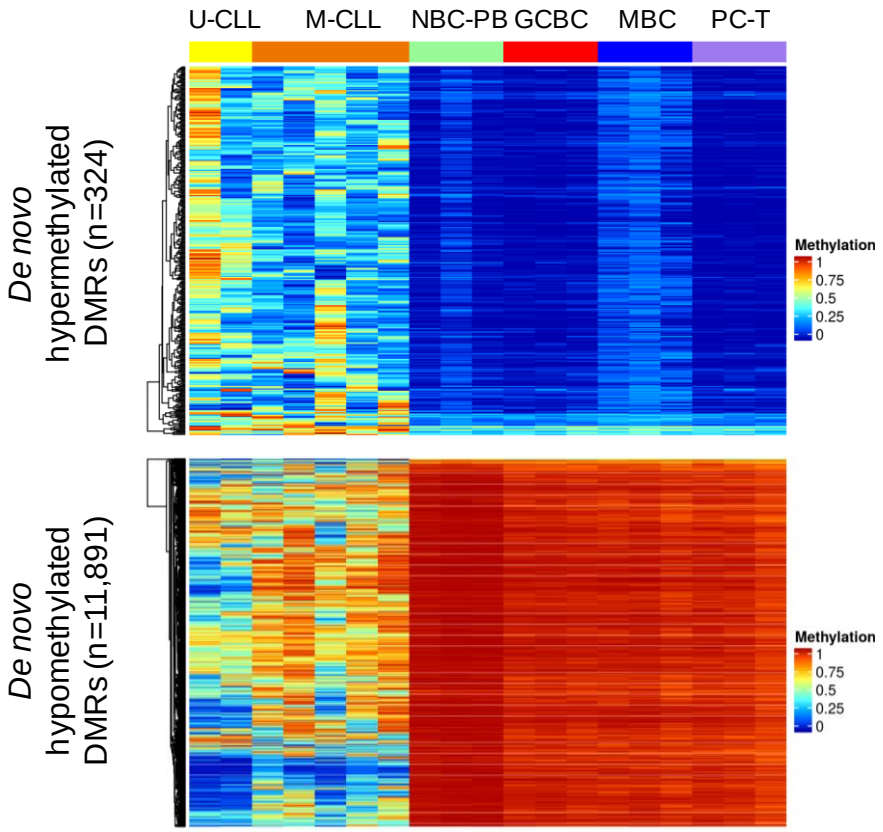
Supplementary Figures



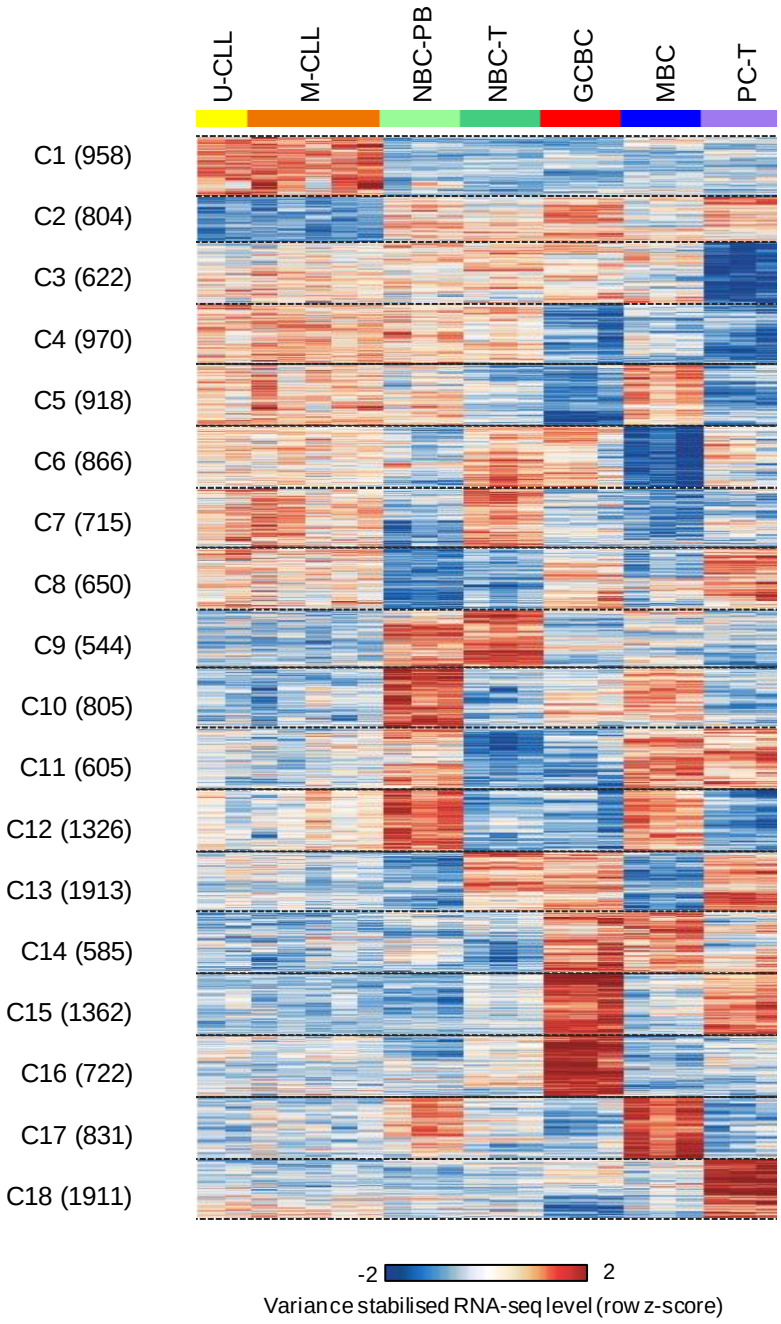
Supplementary Figure 1. Example of a *de novo* active and inactive gene in CLLs. All nine layers of the reference epigenome (H3K27ac, H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3, chromatin accessibility, DNA methylation and gene expression) are shown for a gene *de novo* active in CLL (*CTLA4*, panel a) and a gene *de novo* inactive in CLL (*EBF1*, panel b). Interestingly, *de novo* gene activation (panel a) is associated with a loss of DNA methylation in promoters and enhancers whereas *de novo* inactivation (panel b) is not related to any clear change in DNA methylation.



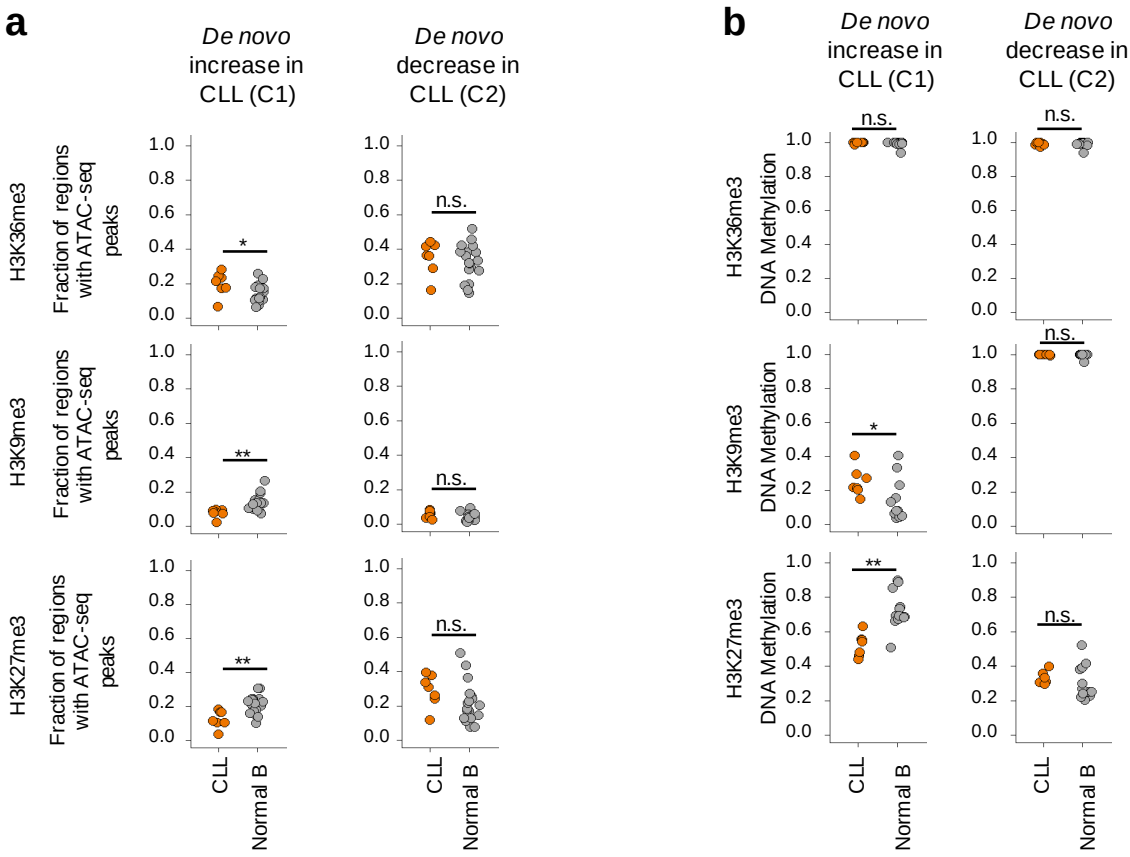
Supplementary Figure 2. K-means cluster analysis of histone mark and ATAC-seq data. K-means clustering of genomic regions showing differences in the dynamics of histone mark enrichment and DNA accessibility in CLL and normal B cells. For each cluster, the number of independent genomic regions is indicated between brackets. C1 and C2 represent respectively regions with *de novo* increase and *de novo* decrease in CLL. U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal centre B cell; MBC, memory B cell; PC-T, plasma cell from tonsil.



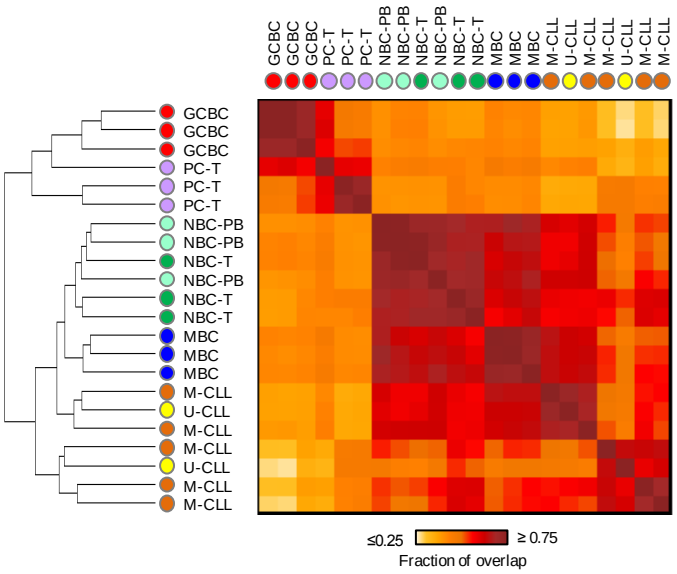
Supplementary Figure 3. Differential DNA methylation in CLL. Heatmaps showing DNA methylation levels of differentially methylated regions (DMRs) in CLL as compared to normal B cells. In brackets the number of independent DMRs are indicated. U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal centre B cell; MBC, memory B cell; PC-T, plasma cell from tonsil.



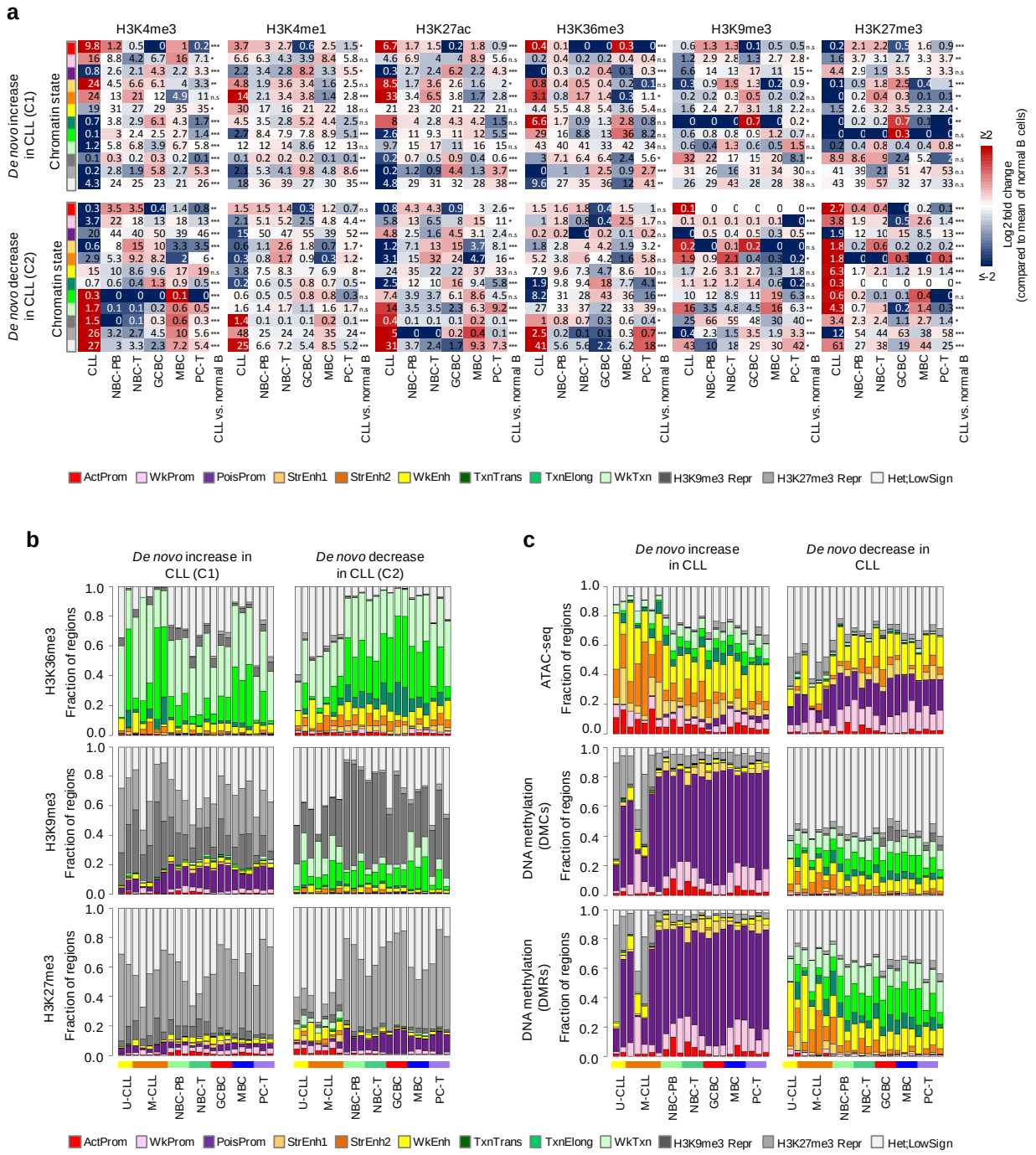
Supplementary Figure 4. K-means cluster analysis of RNAseq data. K-means clustering of genes showing differences in the dynamics of gene expression levels in CLL and normal B cells. For each cluster, the number of independent genes is indicated between brackets. C1 and C2 respectively represent regions with *de novo* increase and *de novo* decrease in CLL. U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal centre B cell; MBC, memory B cell; PC-T, plasma cell from tonsil.



Supplementary Figure 5. DNA accessibility and methylation in *de novo* active and inactive histone mark clusters in CLL. (a) Fraction of regions in CLL (n=7 biologically independent samples) and normal B cells (n=15 biologically independent samples) harbouring ATAC-seq peaks in regions with *de novo* increase (C1) or *de novo* decrease (C2) in CLL of H3K36me3 (respective P-values 4.7 x 10^{-2} and 4.9 x 10^{-1}), H3K9me3 (respective P-values 1.6 x 10^{-3} and 3.3 x 10^{-1}) and H3K27me3 (respective P-values 7.6 x 10^{-3} and 8.5 x 10^{-2}). P-values were calculated using a Wilcoxon signed-rank test (two-sided). (b) Median DNA methylation levels in CLL (n=7 biologically independent samples) and normal B cells (n=15 biologically independent samples) of regions with *de novo* increase (C1) or *de novo* decrease (C2) in CLL of H3K36me3 (respective P-values 3.3 x 10^{-1} and 8.6 x 10^{-1}), H3K9me3 (respective P-values 4.2 x 10^{-2} and 8.1 x 10^{-1}) and H3K27me3 (respective P-values 1.5 x 10^{-3} and 2.3 x 10^{-1}). P-values were calculated using a Wilcoxon signed-rank test (two-sided).

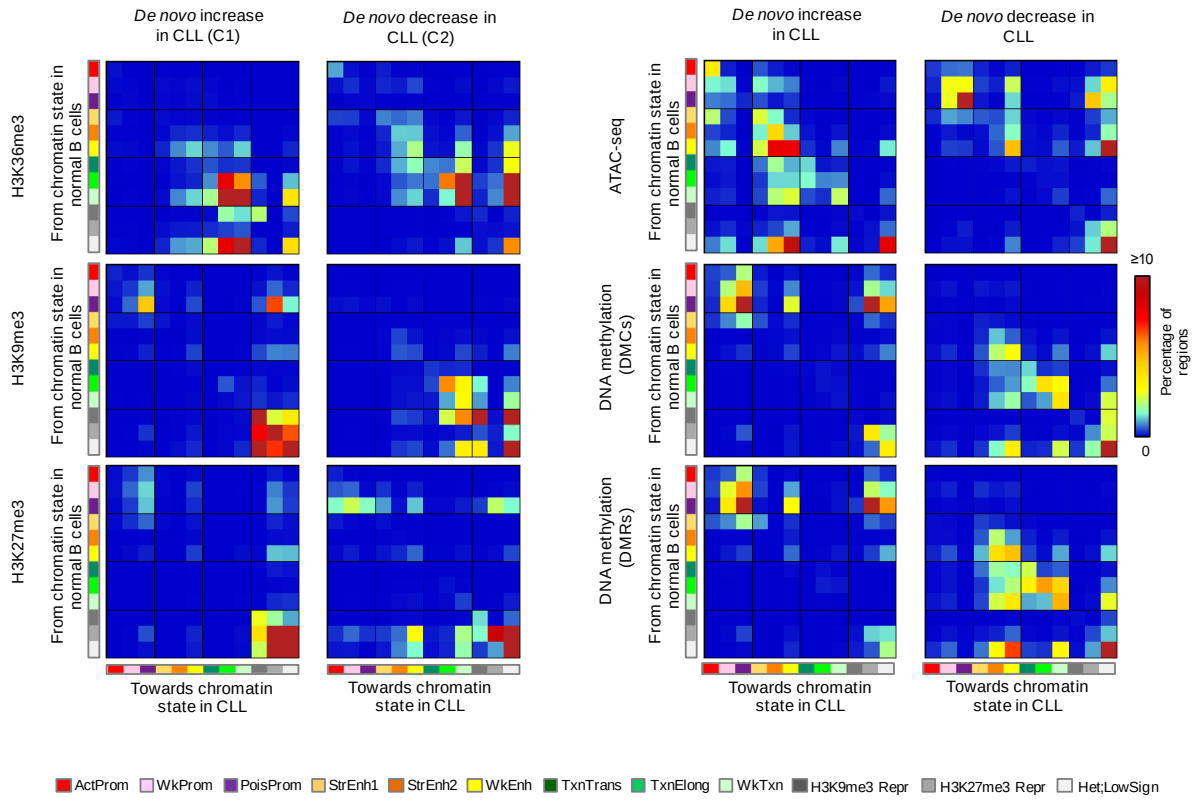


Supplementary Figure 6. Overlap of chromatin states among reference epigenome samples. Heatmap indicating the fraction of overlap of chromatin states in the regions differential among CLL and normal B cells (n=2,167,103 independent genomic regions of 200 base pairs). Clustering is performed using the dissimilarity matrix (1-fraction of overlap). U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal centre B cell; MBC, memory B cell; PC-T, plasma cell from tonsil.

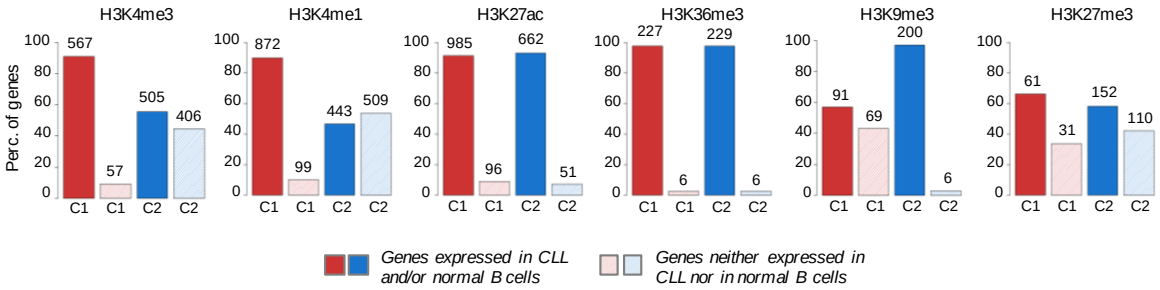


Supplementary Figure 7. Chromatin states in regions with *de novo* increase or decrease of histone marks, DNA accessibility or DNA methylation in CLL. (a) Occupation of each chromatin states (in percentage) in regions that show *de novo* increase and *de novo* decrease of the indicated histone marks in CLL. Represented are the means of CLL (n=7 biologically independent samples) and the means of the normal B-cell subpopulations (n=3 biologically independent samples per subpopulation). P-values from top to bottom for H3K4me3 were 1.4×10^{-4}, 4.7×10^{-2}, 2.3×10^{-5}, 3.8×10^{-3}, 1.4×10^{-1}, 3.2×10^{-2}, 1.4×10^{-4}, 2.5×10^{-4}, 1.2×10^{-5}, 9.2×10^{-4}, 1.2×10^{-5}, 1.2×10^{-5}, 2.9×10^{-3}, 4.7×10^{-5}, 3.5×10^{-4}, 2.2×10^{-4}, 4.7×10^{-2}, 8.9×10^{-1}, 8.4×10^{-1}, 8.4×10^{-2}, 3.5×10^{-4}, 2.2×10^{-4}, 8.2×10^{-5}, 1.2×10^{-5}; for H3K4me1 were 4.7×10^{-2}, 3.7×10^{-1},1.7×10^{-2}, 6.3×10^{-1}, 2.2×10^{-4}, 9.1×10^{-2}, 9.5×10^{-1}, 2.2×10^{-4}, 9.5×10^{-1}, 6.6×10^{-3}, 4.7×10^{-5}, 1.4×10^{-4}, 1.6×10^{-1}, 1.5×10^{-3}, 8.2×10^{-5}, 2.6×10^{-2}, 1.4×10^{-2}, 2.9×10^{-3}, 2.9×10^{-3}, 7.8×10^{-1}, 6.8×10^{-1}, 1.2×10^{-5}, 6.6×10^{-3},

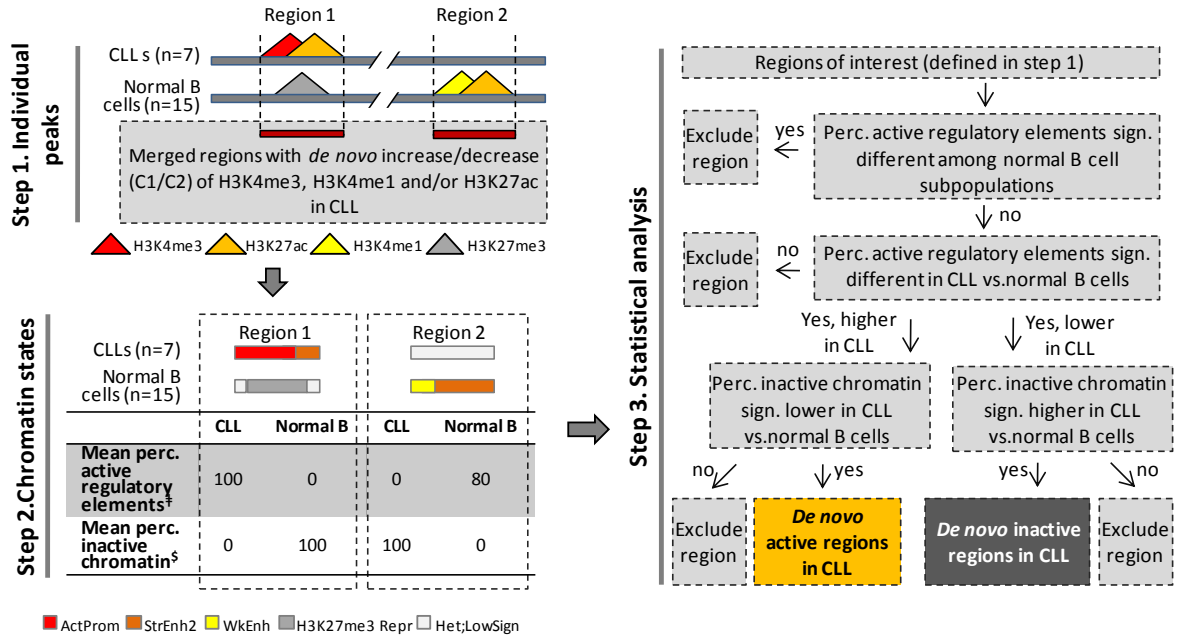
1.2×10^{-5}; for H3K27ac were 2.2×10^{-4}, 3.0×10^{-1}, 1.2×10^{-5}, 2.6×10^{-2}, 2.3×10^{-5}, 4.1×10^{-1}, 2.1×10^{-1}, 4.7×10^{-5}, 7.8×10^{-2}, 8.2×10^{-5}, 1.2×10^{-5}, 1.2×10^{-5}, 5.1×10^{-3}, 1.7×10^{-2}, 9.1×10^{-2}, 3.5×10^{-4}, 2.1×10^{-3}, 4.1×10^{-1}, 1.4×10^{-4}, 7.3×10^{-1}, 4.7×10^{-5}, 7.7×10^{-4}, 2.5×10^{-4}, 1.2×10^{-5}; for H3K36me3 were 7.7×10^{-4}, 3.0×10^{-1}, 2.1×10^{-4}, 3.3×10^{-1}, 1.1×10^{-2}, 2.1×10^{-1}, 9.1×10^{-2}, 4.5×10^{-1}, 6.3×10^{-1}, 1.4×10^{-2}, 2.4×10^{-4}, 5.1×10^{-3}, 4.1×10^{-1}, 5.3×10^{-1}, 2.9×10^{-1}, 3.2×10^{-2}, 5.3×10^{-1}, 5.3×10^{-1}, 3.5×10^{-4}, 3.5×10^{-4}, 7.8×10^{-2}, 2.6×10^{-2}, 5.3×10^{-4}, 1.2×10^{-5}; for H3K9me3 were 5.3×10^{-1}, 1.1×10^{-2}, 1.1×10^{-3}, 2.1 x 10-2, 4.4×10^{-1}, 7.8×10^{-2}, 4.1×10^{-2}, 5.6×10^{-2}, 2.1×10^{-1}, 2.9×10^{-3}, 5.3×10^{-1}, 1.2×10^{-1}, 1.7×10^{-4}, 5.3×10^{-4}, 2.4×10^{-4}, 1.2×10^{-1}, 2.1×10^{-2}, 9.1×10^{-2}, 8.9×10^{-1}, 5.8×10^{-1}, 2.9×10^{-3}, 2.1×10^{-3}, 1.2×10^{-5}, 1.1×10^{-2}; and for H3K27me3 were 5.3×10^{-4}, 2.1×10^{-3}, 6.6×10^{-2}, 1.4×10^{-4}, 6.5×10^{-3}, 1.4×10^{-2}, 1.0×10^{-3}, 6.6×10^{-2}, 9.1×10^{-3}, 1.6×10^{-1}, 7.8×10^{-1}, 6.8×10^{-1}, 4.2×10^{-4}, 1.4×10^{-4}, 1.2×10^{-5}, 2.2×10^{-4}, 2.3×10^{-4}, 3.5×10^{-4}, 2.4×10^{-3}, 6.1×10^{-2}, 1.2×10^{-5}, 2.6×10^{-2}, 1.2×10^{-5}, 2.3×10^{-5}. P-values were calculated using a using a Wilcoxon signed-rank test (two-sided) comparing CLL (n=7 biologically independent samples) and all normal B cells (n=15 biologically independent samples). (b) Distribution of the different chromatin states in all analysed samples separately (seven CLLs and 15 normal B cells) at regions with *de novo* increase (C1) or *de novo* decrease (C2) of H3K36me3, H3K9me3 and H3K27me3 in CLL. (c) Distribution of the different chromatin states in all analysed samples separately (seven CLLs and 15 normal B cells) at regions with *de novo* increase (C1) or *de novo* decrease (C2) of DNA accessibility (ATAC-seq) or in *de novo* differentially methylated CpGs and regions in CLL. U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal center B cell; MBC, memory B cell; PC-T, plasma cell from tonsil. ActProm, Active Promoter; WkProm, Weak Promoter; PoisProm, poised Promoter; StrEnh1, Strong Enhancer 1; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; Txn_Trans, Transcription Transition; Txn_Elong, Transcription Elongation; Wk_Txn, Weak Transcription; H3K9me3_Repr, H3K9me3 Repressed; H3K27me3_Repr, H3K27me3 Repressed; Het;LowSign, Heterochromatin;Low Signal. DMCs, differentially methylated CpGs; DMRs, differentially methylated regions.



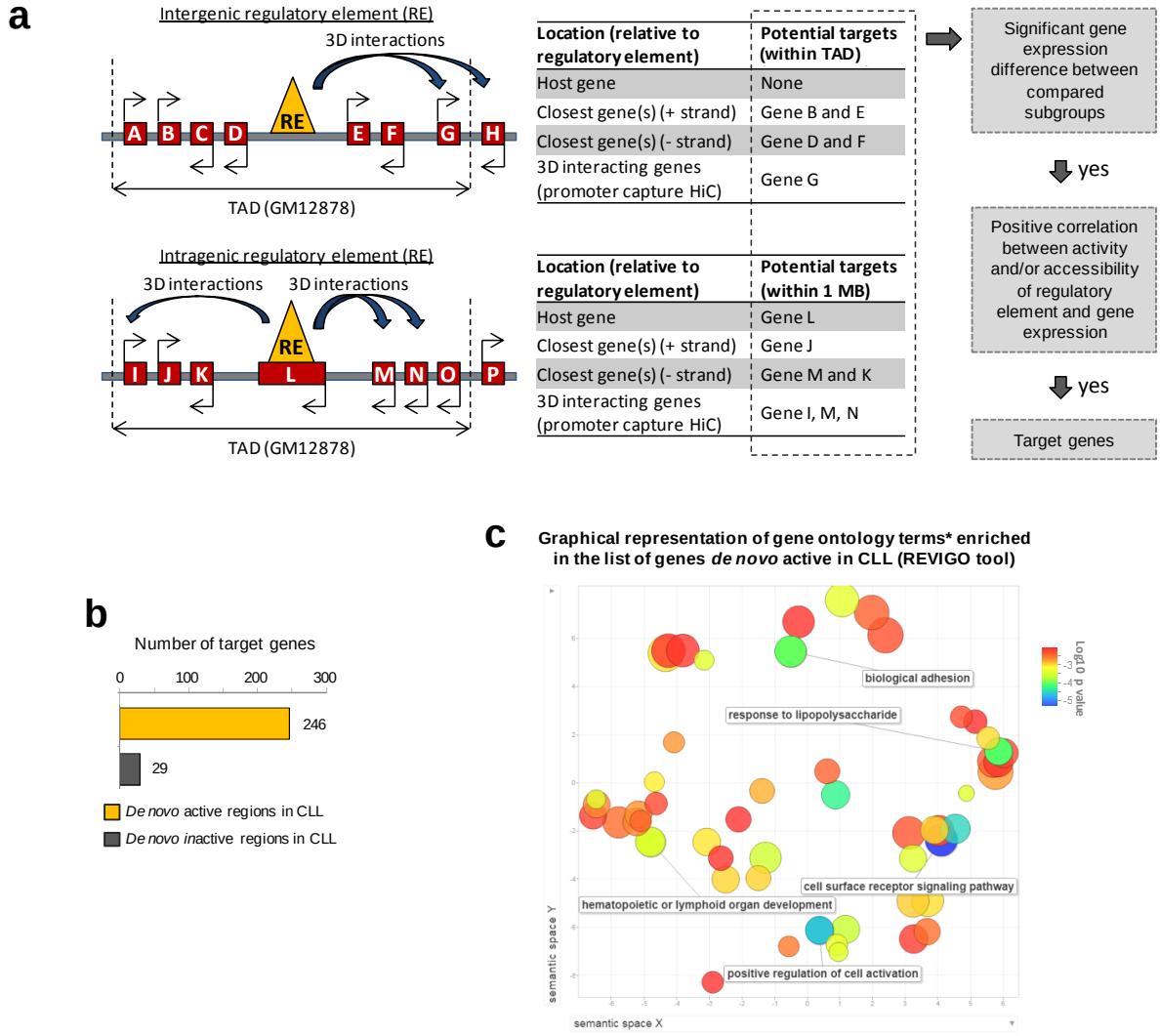
Supplementary Figure 8. Chromatin state transitions in regions with *de novo* increase or decrease of histone marks, DNA accessibility or DNA methylation in CLL. Percentages of regions with *de novo* increase (C1) or *de novo* decrease (C2) of H3K36me3, H3K9me3, H3K27me3, ATAC-seq and DNA methylation in CLL that harbour a specific chromatin state in normal B cells (rows, n=15 biologically independent samples) and the same (diagonal, no change of chromatin state) or another state (chromatin state switch) in CLL (columns, n=7 biologically independent samples). The total matrix represents 100 percent of the regions. U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal center B cell; MBC, memory B cell; PC-T, plasma cell from tonsil. ActProm, Active Promoter; WkProm, Weak Promoter; PoisProm, poised Promoter; StrEnh1, Strong Enhancer 1; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; Txn_Trans, Transcription Transition; Txn_Elong, Transcription Elongation; Wk_Txn, Weak Transcription; H3K9me3_Repr, H3K9me3 Repressed; H3K27me3_Repr, H3K27me3 Repressed; Het;LowSign, Heterochromatin;Low Signal; DMCs, differentially methylated CpGs; DMRs, differentially methylated regions.



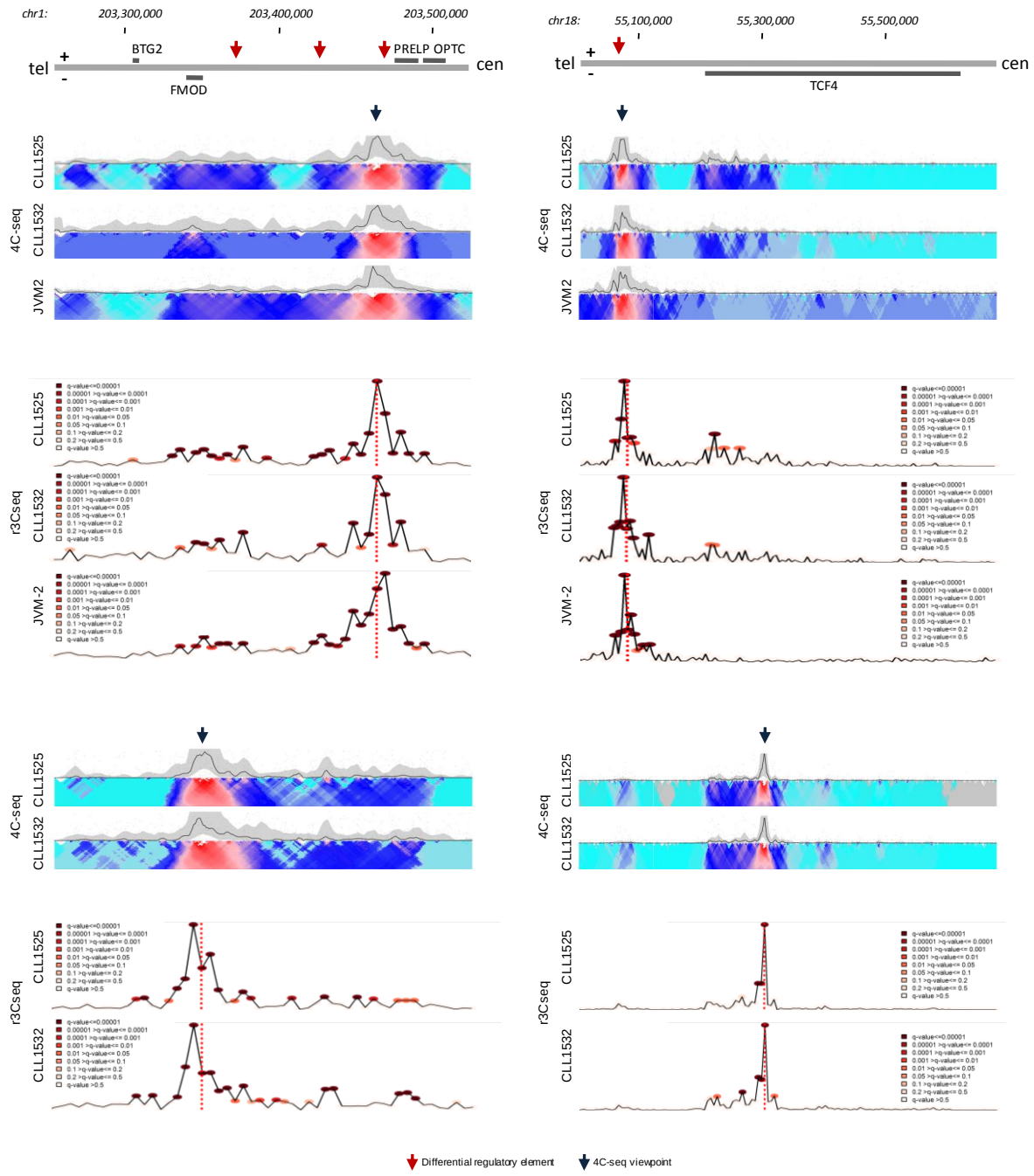
Supplementary Figure 9. Number of genes expressed in *de novo* active and inactive histone mark clusters in CLL. Number of genes overlapping with regions with *de novo* increase (C1, red) or *de novo* decrease (C2, blue) of the indicated histone marks that are expressed in CLL and/or normal B cells (filled bars) versus the ones that are neither expressed in CLL nor in normal B cells (striped bars), with expressed being defined as a mean expression of log10(fpkm) greater than -1.



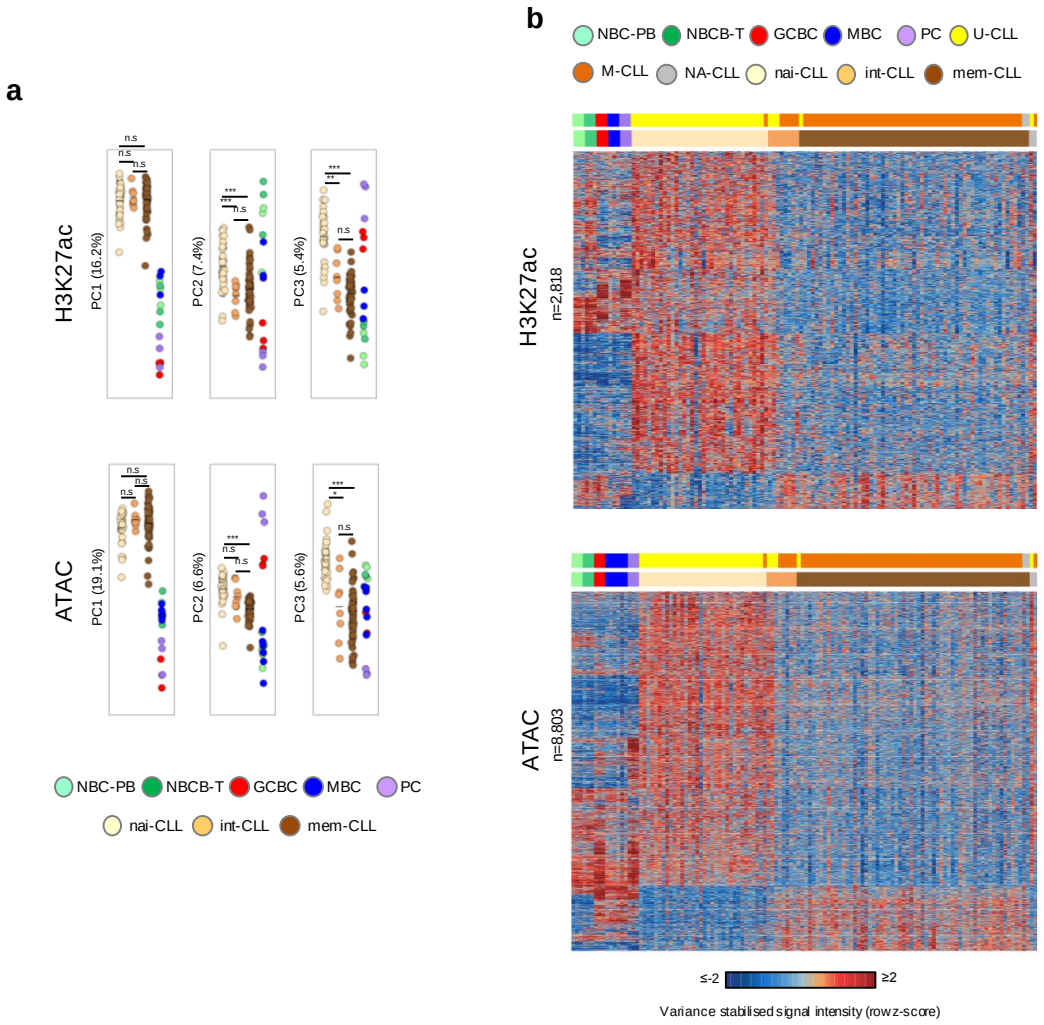
Supplementary Figure 10. Selection of *de novo* active and inactive regions in CLL. Strategy used to define *de novo* active and inactive regulatory elements in CLL. For the statistical differences among normal B cells, a minimal difference in mean percentage of 10% in at least one pairwise comparison (comparing the 5 different normal B-cell subpopulations, n=3 biologically independent samples each) and a P-value of 1.0×10^{-1} were used (Kruskal-Wallis test) as selection criteria. For the statistical differences between CLL (n=7 biologically independent samples) and normal B cells (n=15 biologically independent samples) a minimal difference in mean percentage of 25% and an FDR of 1.0×10^{-2} were used (Wilcoxon signed-rank test, two-sided) as selection criteria. ‡Active regulatory elements = Active Promoter, Strong Enhancer 1 and Strong Enhancer 2; $Inactive chromatin = Poised promoter, H3K9me3 repressed, H3K27me3 repressed and Heterochromatin;Low Signal. ActProm, Active Promoter; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; H3K27me3_Repr, H3K27me3 Repressed; Het;LowSign, Heterochromatin;Low Signal; RE, regulatory element.



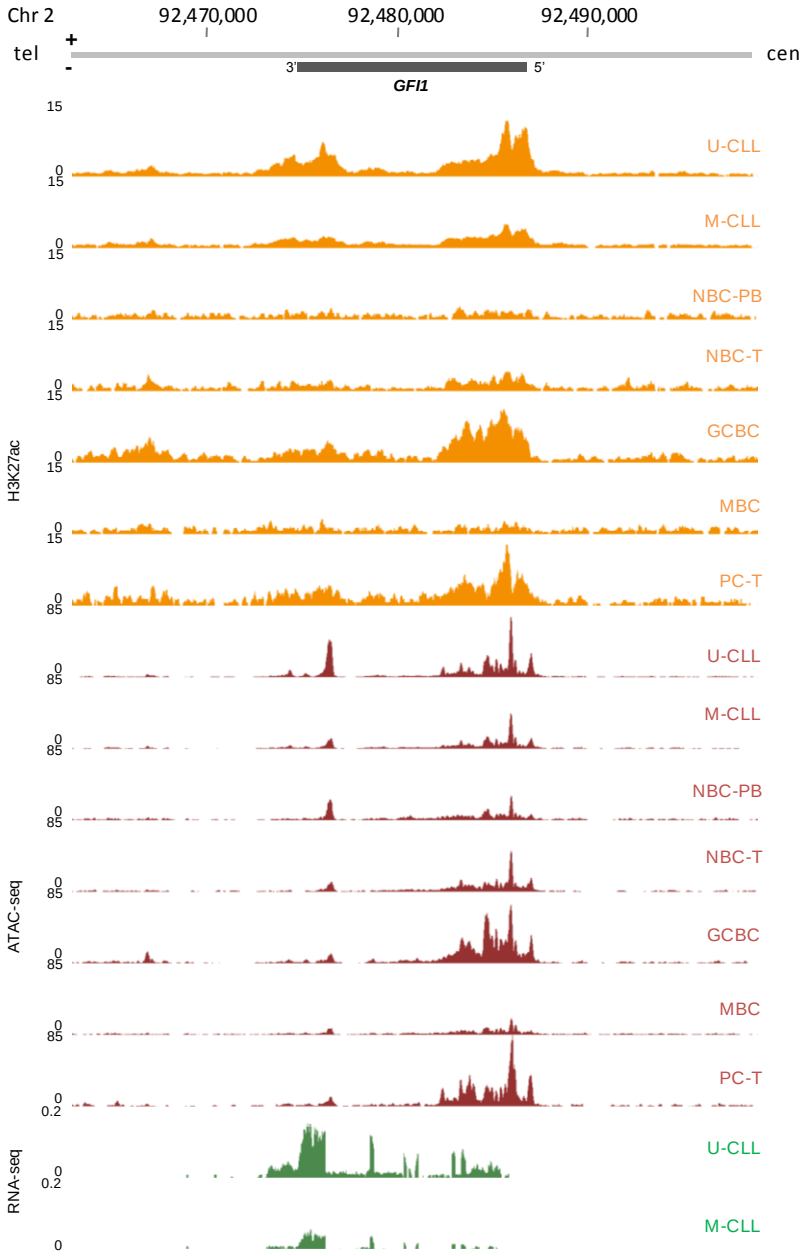
Supplementary Figure 11. Assignment of target genes and gene ontologies. (a) Selection criteria to select target genes of regulatory elements. Potential target genes, based on linear and 3-dimensional proximity, were only assigned if they were located within the topologically associated domain (TAD) of the regulatory element (as determined in GM12878). The final target genes were assigned when a significant difference (DE-seq2 package, nbinomWaldTest, mean(log10(fpkm +0.01)) in at least one of the compared subgroups (CLL/CLL subgroups/normal B cells) > -1.0) was observed between the compared samples, but only if a positive correlation was observed between activity/accessibility and the gene expression levels. (b) Number of target genes assigned to the *de novo* active and inactive regions in CLL. (c) Scatter plot (REVIGO online software, http://revigo.irb.hr) of the 100 most significant gene ontology terms associated with the 246 independent genes showing *de novo* gain of regulatory elements in CLL. * A full list of GO terms and associated genes can be found in Supplementary Table 5c.



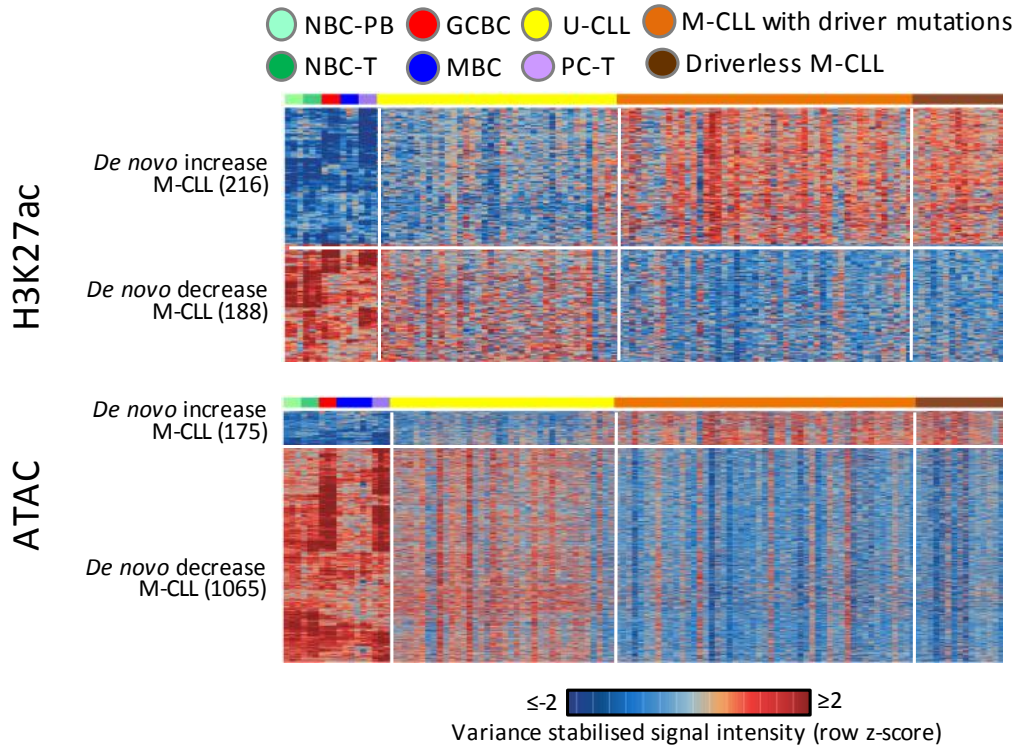
Supplementary Figure 12. Three-dimensional chromatin interactions measured by 4C-seq. *4C-seq*: The normalised levels of 3D chromatin interaction frequencies from the indicated viewpoint (dark blue arrow), as analysed by 4Cseq pipe in CLL1525, CLL1532 and JVM-2, with turquoise indicating the lowest and red the highest frequencies and the black line showing the median contact frequencies in a window of 5,000 bp. *r3Cseq*: Normalised reads per million, representing chromatin interaction frequencies from the indicated viewpoint (red dotted line), as analysed by r3Cseq in CLL1525, CLL1532 and JVM-2 using 5,000 bp windows. Red colored dots indicate significant 3D interactions with different q-values. The investigated enhancers are not active in JVM-2, which therefore serves as a negative control. The 3D interaction landscape in the *FMOD* locus was maintained in JVM-2 regardless of enhancer activity, in contrast, the *TCF4* locus only showed 3D interactions when the enhancer is active (CLL1525 and CLL1532) but not when it is inactive (JVM-2).



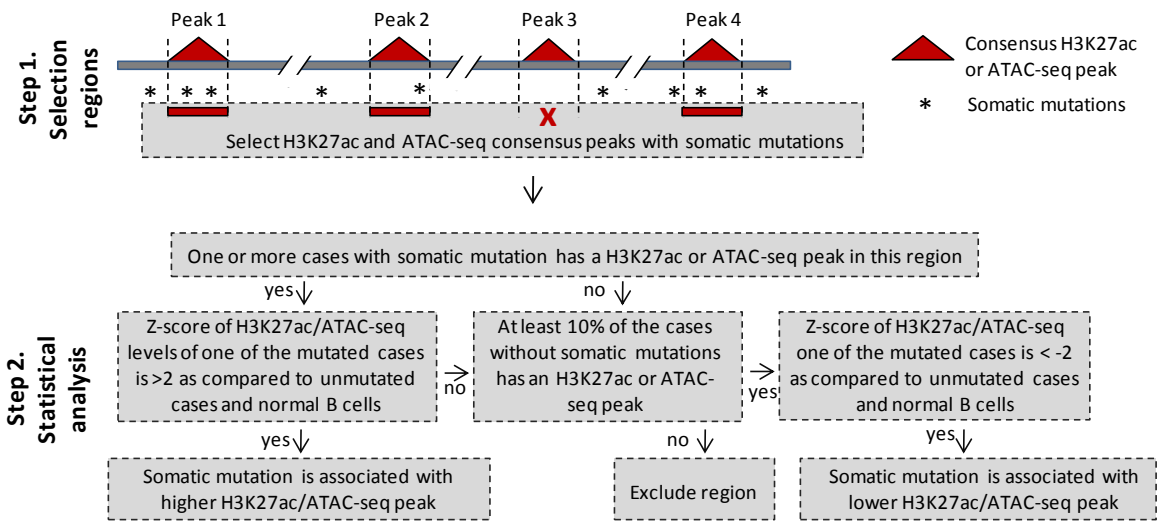
Supplementary Figure 13. Chromatin activity and accessibility in epigenetic CLL subgroups. (a) Unsupervised principal component analysis (first three components) of the extended CLL cohort for H3K27ac (respective corrected P-values of PC1, PC2 and PC3 between nai-CLL (n=35 biologically independent samples) and int-CLL (n=8 biologically independent samples) were 1.0, 3.7 x 10^{-4} and 2.7 x 10^{-3}; between nai-CLL (n=35 biologically independent samples) and mem-CLL (n=59 biologically independent samples) were 1.0, 2.8 x 10^{-5} and 6.2 x 10^{-16}; and between int-CLL (n=8 biologically independent samples) and mem-CLL (n=59 biologically independent samples) were 1.0, 8.1 x 10^{-1} and 2.7 x 10^{-1}) and for ATAC-seq (respective corrected P-values of PC1, PC2 and PC3 between nai-CLL (n=34 biologically independent samples) and int-CLL (n=8 biologically independent samples) were 1.8 x 10^{-1}, 5.0 x 10^{-1} and 1.9 x 10^{-2}; between nai-CLL (n=34 biologically independent samples) and mem-CLL (n=62 biologically independent samples) were 6.6 x 10^{-1}, 5.8 x 10^{-9} and 6.8 x 10^{-19}; and between int-CLL (n=8 biologically independent samples) and mem-CLL (n=62 biologically independent samples) were 1.0, 1.7 x 10^{-1} and 1.0). Number of datapoints analyzed to generate the PCAs: H3K27ac (n=58,790 independent genomic regions) and ATAC-seq (n=115,352 independent genomic regions). P-values were calculated using a Student's t-test (two-sided) and corrected using the Bonferroni correction. (b) Heatmap of signal intensities of H3K27ac and ATAC-seq in differential regions between U-CLL and M-CLL. Signal intensities are indicated as row z-scores. U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; nai-CLL, naive like CLL; int-CLL, intermediate CLL; mem-CLL, memory like CLL; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal center B cell; MBC, memory B cell; PC-T, plasma cell from tonsil.



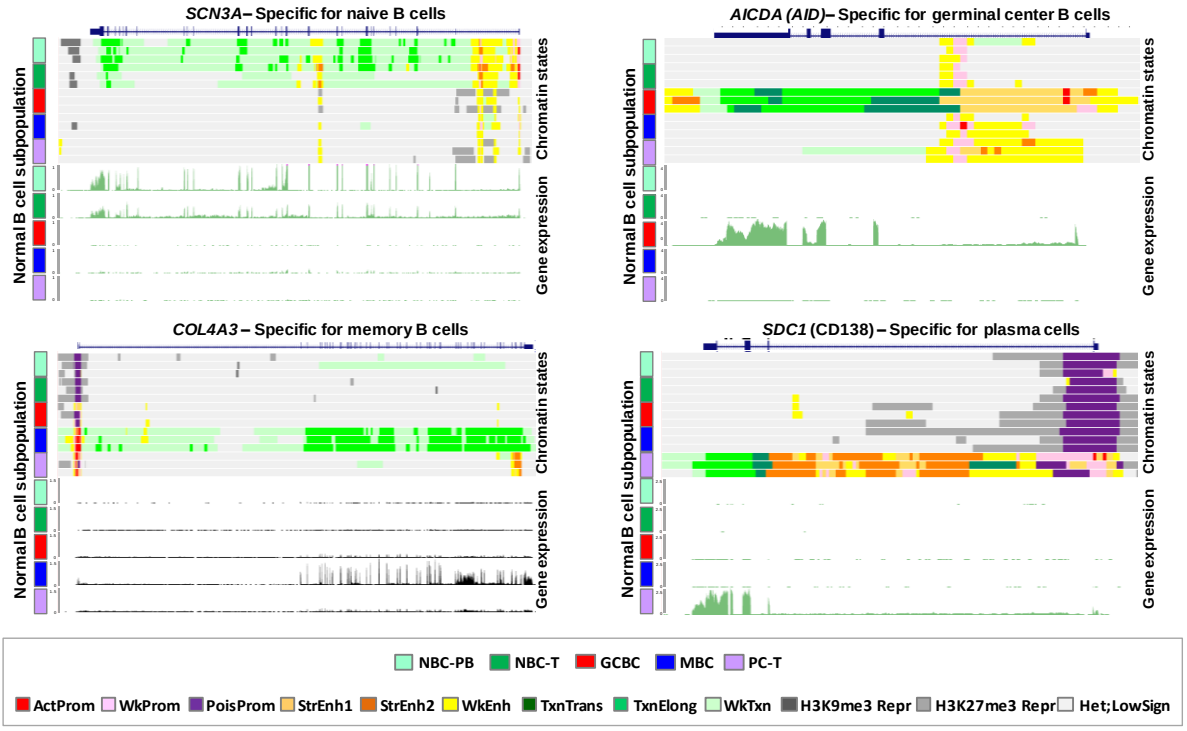
Supplementary Figure 14. Chromatin activation and upregulation in *GFI1*. Example of an identified region highly active in U-CLL, GCBCs and PCs, targeting *GFI1*. Indicated are median H3K27ac levels (orange), median ATAC-seq levels (dark red) and median levels of RNA-seq (green). U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; nai-CLL, naive like CLL; int-CLL, intermediate CLL; mem-CLL, memory like CLL; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal center B cell; MBC, memory B cell; PC-T, plasma cell from tonsil.



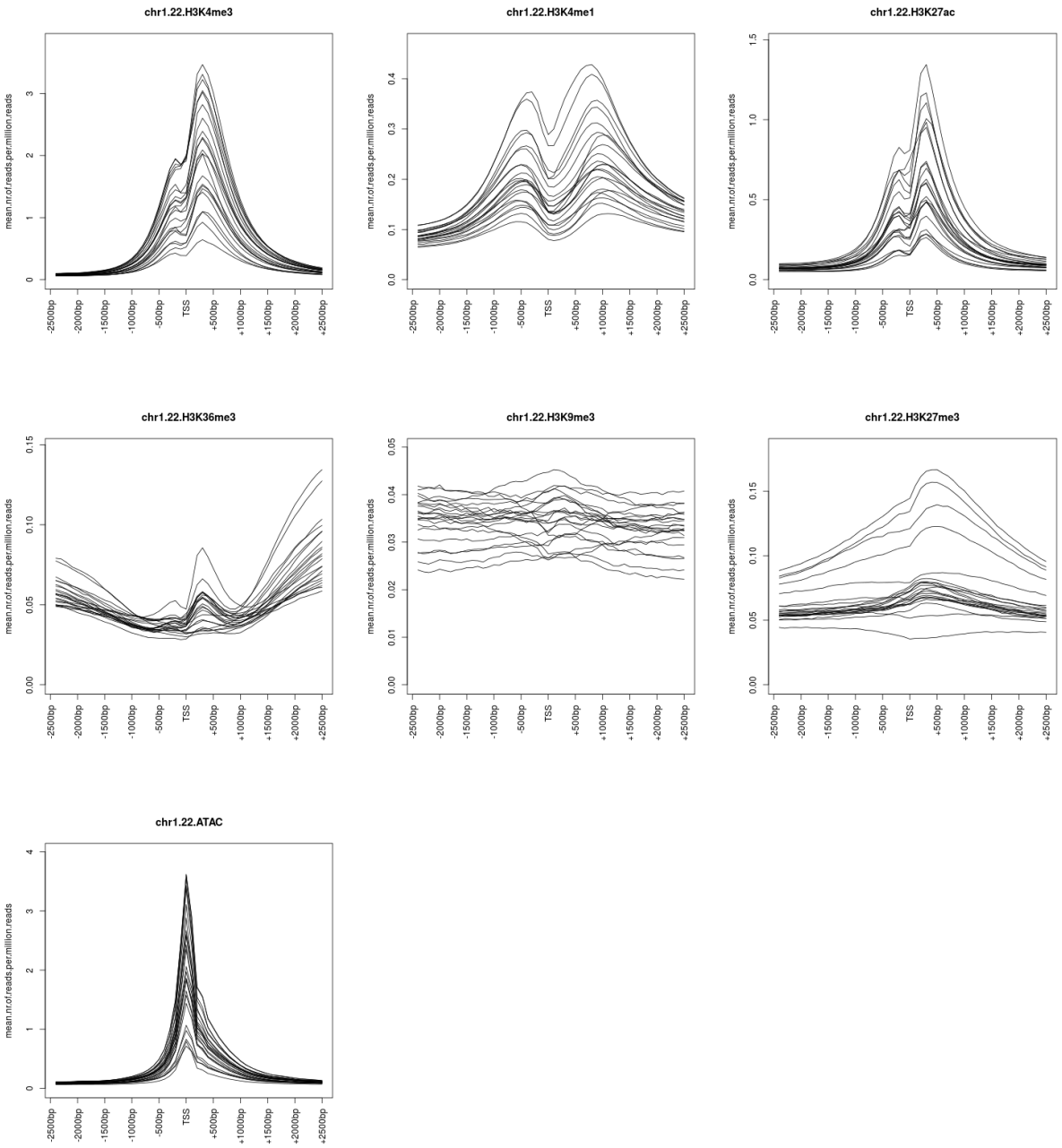
Supplementary Figure 15. Chromatin activation and accessibility in M-CLLs with and without driver mutations. Heatmap of *de novo* regions with increased and decreased H3K27ac and ATAC-seq levels in M-CLL cases. M-CLL cases are separated in those with and without driver mutations. Signal intensities are indicated as row z-scores. Driverless M-CLLs show a similar chromatin profile as M-CLLs with driver mutations. U-CLL, CLL with unmutated IGHV; M-CLL, CLL with mutated IGHV; nai-CLL, naive like CLL; int-CLL, intermediate CLL; mem-CLL, memory like CLL; NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal center B cell; MBC, memory B cell; PC-T, plasma cell from tonsil.



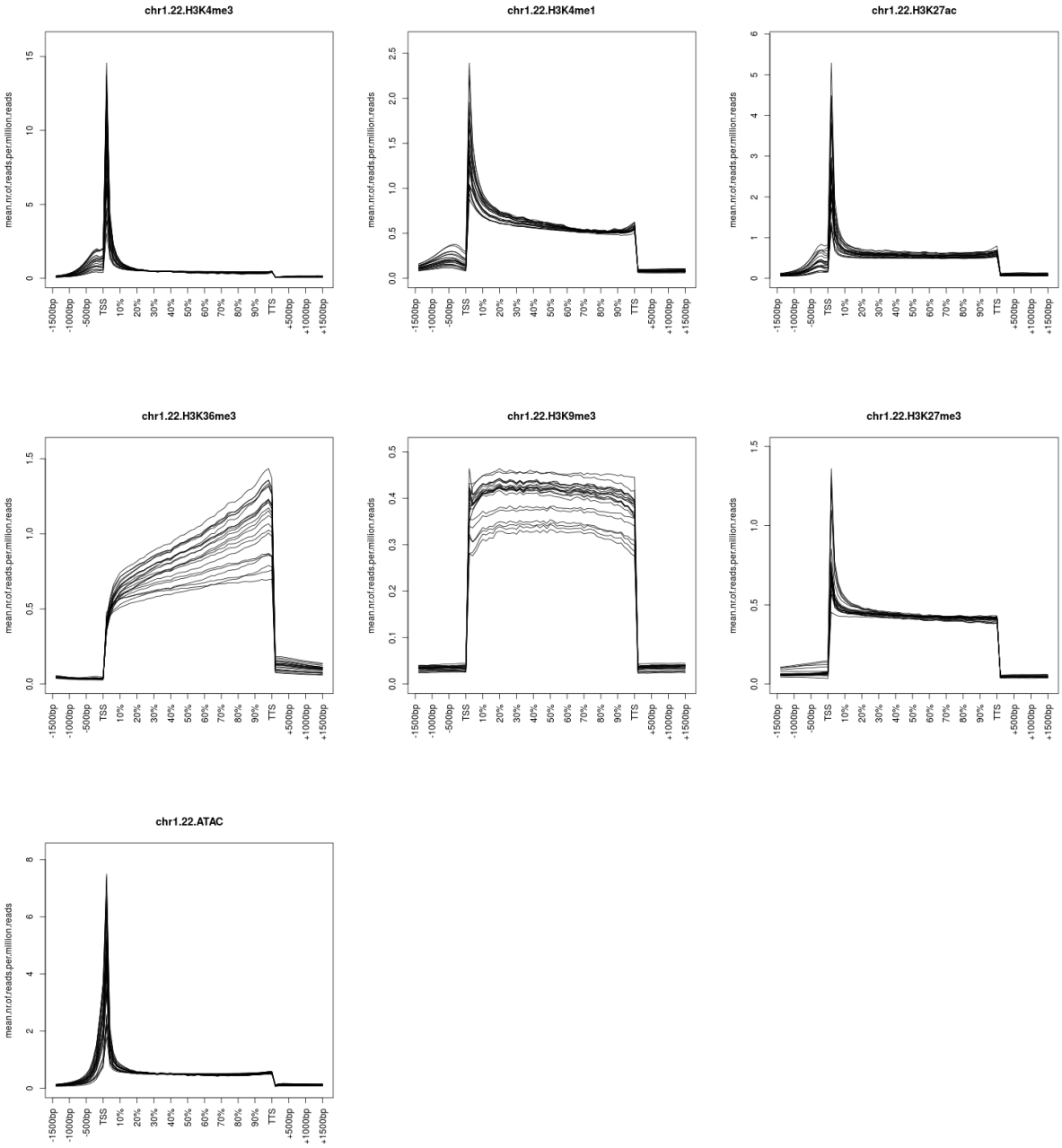
Supplementary Figure 16. Determination of local chromatin activity changes in CLL related to somatic mutations. Strategy to detect regions with somatic mutations that are associated with a local change in chromatin activity or accessibility. The immunoglobulin loci (chr14:105550000-107043718, chr2:88816899-90235370 and chr22:22020000-22962010, all grch38) were excluded in this analysis.



Supplementary Figure 17. Chromatin states and gene expression of genes specific for each B cell subpopulation analyzed in the present study. NBC-PB, naive B cell from peripheral blood; NBC-T, naive B cell from tonsil; GCBC, germinal center B cell; MBC, memory B cell; PC-T, plasma cell from tonsil. ActProm, Active Promoter; WkProm, Weak Promoter; PoisProm, poised Promoter; StrEnh1, Strong Enhancer 1; StrEnh2, Strong Enhancer 2; WkEnh, Weak Enhancer; TxnTrans, Transcription Transition; TxnElong, Transcription Elongation; WkTxn, Weak Transcription; H3K9me3 Repr, H3K9me3 Repressed; H3K27me3 Repr, H3K27me3 Repressed; Het;LowSign, Heterochromatin;Low Signal.



Supplementary Figure 18. Read profiles around TSS. Mean number of reads per million reads around the transcriptional start sites (-2500 to +2500bp) of all protein coding genes on chromosome 1-22 for all 6 histone marks and ATAC-seq. Each line represents a sample of the reference epigenome dataset (22 biologically independent samples for the histone marks, 25 biologically independent samples for ATAC-seq). As the goal of this analysis was to determine read profiles around the TSS of each sample and not to compare them to each other, we did not perform any inter-sample normalization.



Supplementary Figure 19. Read profiles across gene bodies. Mean number of reads per million reads within gene bodies (50 windows of 2% of the gene bodies of all protein coding genes on chromosome 1-22, together with -1500 from the transcriptional start site to +1500bp from the transcriptional termination site) for all 6 histone marks and ATAC-seq. Each line represents a sample of the reference epigenome dataset (22 biologically independent samples for the histone marks, 25 biologically independent samples for ATAC-seq). As the goal of this analysis was to determine read profiles across the gene body of each sample and not to compare them to each other, we did not perform any inter-sample normalization.

Supplementary Tables

Supplementary Table 1. Patient characteristics.

Supplementary Table 2. K-means clusters. Genomic coordinates/gene names and levels of (a) the different histone marks, (b) DNA accessibility and (c) gene expression for the regions/genes within each K-means cluster.

Supplementary Table 3. Differential DNA methylation in CLL. Genomic coordinates and DNA methylation levels of *de novo* (a) differentially methylated CpGs (DMCs) and (b) differentially methylated regions (DMRs) in CLL as compared to normal B cells.

Supplementary Table 4. Genes related to K-means clusters histone marks. (a) Genes associated with each histone mark cluster and (b) gene expression levels within regions with *de novo* increase (C1) or decrease (C2) of the different histone marks.

Supplementary Table 5. Gene ontology terms. (a) Enriched gene ontology terms for the genes associated with each histone mark cluster (top 20 terms per cluster). Number of independent target genes for which gene ontology analysis was performed: H3K4me3 (cluster 1: 624, cluster 2: 911), H3K4me1 (cluster 1: 971, cluster 2: 952), H3K27ac (cluster 1: 1,081, cluster 2: 713), H3K36me3 (cluster 1: 233, cluster 2: 235), H3K9me3 (cluster 1: 160, cluster 2: 206) and H3K27me3 (cluster 1: 92, cluster 2: 262). (b) Clustering of gene ontology terms related to regions with *de novo* increase (C1) or decrease (C2) of the different histone marks as in the rows figure 1g. (c) Enriched gene ontology terms for the genes (n=246 independent target genes) associated with *de novo* active regulatory elements in CLL. (d) Enriched gene ontology terms for the genes associated with *de novo* regions differentially active/accessible between U-CLL and M-CLL. Number of independent target genes for which gene ontology analysis was performed: H3K27ac (*de novo* increase in U-CLL: 233, *de novo* increase in M-CLL: 58, *de novo* decrease in U-CLL: 44, *de novo* decrease in M-CLL: 98), ATAC-seq (*de novo* increase in U-CLL: 427, *de novo* increase in M-CLL: 38, *de novo* decrease in U-CLL: 82, *de novo* decrease in M-CLL: 226). (e) Enriched gene ontology terms for the genes associated with regions specifically active/accessible in U-CLL, GCBC and PCT. Number of independent target genes for which the gene ontology analysis was performed: H3K27ac (n=54) and ATAC-seq (n=154).

Supplementary Table 6. Chromatin states related to K-means clusters histone marks. (a) Individual and (b) mean percentages of chromatin states in regions within the different histone mark clusters. The background represents the total genomic occupation (in percentage) of each chromatin state.

Supplementary Table 7. *De novo* active and inactive regulatory elements in CLL. Coordinates and characteristics of *de novo* active and inactive regulatory elements in CLL including target genes and their gene expression levels. Differential regions were determined as indicated in supplementary figure 10 comparing CLL (n=7 biologically independent samples) with normal B cells (n=15 biologically independent samples). Target genes were defined as indicated in supplementary figure 11a comparing CLL (n=7 biologically independent samples) with normal B cells (n=15 biologically independent samples), using an FDR cutoff of 0.05.

Supplementary Table 8. Enriched transcription factor motifs. (a) Transcription factor binding motif enrichment results, including their expression levels, for *de novo* active and inactive regions in CLL. Number of regions analyzed vs. background were: *de novo* active regions (n=934 independent genomic loci) versus the background (n=1,868 independent genomic loci), *de novo* inactive regions (n=61 independent genomic loci) versus the

background (n=122 independent genomic loci). (b) Transcription factor binding motif enrichment results, including their expression levels, for regions with a *de novo* change in chromatin activity/accessibility in U-CLL or M-CLL. At maximum, the top 50 significant terms are shown. Number of regions analyzed vs. background were: *de novo* increased accessibility in U-CLL (n= 2,125 vs. 4,250 independent genomic regions) or M-CLL (n=175 vs. 350 independent genomic regions) and *de novo* decreased accessibility in U-CLL (n=238 vs. 476 independent genomic regions) or M-CLL (n=1,065 vs. 2,130 independent genomic regions). Statistical significance was determined using the one-tailed Wilcoxon rank-sum test and the p-values were adjusted using the Bonferroni correction.

Supplementary Table 9. Differentially active/accessible regions U-CLL and M-CLL. Coordinates, characteristics and target genes of regions with differential H3K27ac and ATAC-seq levels between U-CLL and M-CLL. Differential regions were determined using the two-sided nbinomWaldTest in the DEseq2 package comparing U-CLL and M-CLL, corrected for multiple testing (Benjamini-Hochberg). Sample sizes: H3K27ac, U-CLL (n=39 biologically independent samples) vs. M-CLL (n=63 biologically independent samples) and ATAC-seq, U-CLL (n=38 biologically independent samples) vs. M-CLL (n=66 biologically independent samples). Target genes were defined as indicated in supplementary figure 11a comparing U-CLL (n=34 biologically independent samples) with M-CLL (n=42 biologically independent samples), using an FDR cutoff of 0.01.

Supplementary Table 10. Overlap cell of origin signature DNA methylation and ATAC-seq. Overlap of regions with differentially higher ATAC-seq levels in M-CLL as compared to U-CLL and the DNA methylation cell of origin signature.

Supplementary Table 11. Patient groups somatic mutations and structural variants. List of patients included to make comparisons between CLL cases (a) with and without somatic mutations and between driverless and non-driverless CLLs and (b) between CLLs with and without structural variants.

Supplementary Table 12. Differential regions and target genes related to somatic genetic alterations. List of (a) regions with differential H3K27ac and ATAC-seq levels in CLLs with and without mutations/structural variants and (b) their target genes. Statistical significance (supplementary table 12a) was determined using the two-sided nbinomWaldTest in the DEseq2 package, corrected for multiple testing (Benjamini-Hochberg). Target genes were defined as indicated in supplementary figure 11a, using an FDR cutoff of 0.05. Sample sizes: *MYD88*-MT vs. *MYD88*-WT (H3K27ac: n=5 vs. 57, ATAC-seq: n=6 vs. 59, RNA-seq: n=4 vs. 38 biologically independent samples), *SF3B1*-MT vs. *SF3B1*-WT (H3K27ac: n=7 vs. 95, ATAC-seq: n=7 vs. 97, RNA-seq: n=6 vs. 71 biologically independent samples), *ATM*-MT vs. *ATM*-WT (H3K27ac: n=10 vs. 28, ATAC-seq: n=10 vs. 27, RNA-seq: n=9 vs. 24 biologically independent samples), *TP53*-MT vs. *TP53*-WT (H3K27ac: n=5 vs. 97, ATAC-seq: n=5 vs. 99, RNA-seq: n=5 vs. 72 biologically independent samples), I*GLL5*-MT vs. I*GLL5*-WT (H3K27ac: n=6 vs. 56, ATAC-seq: n=7 vs. 58, RNA-seq: n=3 vs. 39 biologically independent samples), *NOTCH1*-MT vs. *NOTCH1*-WT (H3K27ac: n=9 vs. 29, ATAC-seq: n=9 vs. 28, RNA-seq: n=7 vs. 26 biologically independent samples), *SYNE1*-MT vs. *SYNE1*-WT (H3K27ac: n=6 vs. 96, ATAC-seq: n=6 vs. 98, RNA-seq: n=6 vs. 71 biologically independent samples), *MGA*-MT vs. *MGA*-WT (H3K27ac: n=5 vs. 33, ATAC-seq: n=5 vs. 32, RNA-seq: n=5 vs. 28 biologically independent samples), driverless vs. with mutations in driver genes (H3K27ac: n=15 vs. 47, ATAC-seq: n=15 vs. 50, RNA-seq: n=10 vs. 32 biologically independent samples), tri12 vs. non-tri12 (H3K27ac: n=14 vs. 88, ATAC-seq: n=13 vs. 91, RNA-seq: n=11 vs. 66 biologically independent samples), del10q vs. non-del10q (H3K27ac: n=5 vs. 97, ATAC-seq: n=5 vs. 99, RNA-seq: n=5 vs. 72 biologically independent samples), del17p vs. non-del17p (H3K27ac: n=6 vs. 96, ATAC-seq: n=6 vs. 98, RNA-seq: n=6 vs. 71 biologically independent samples), del13q vs. non-del13q (H3K27ac: n=45 vs. 57, ATAC-seq: n=46 vs. 58, RNA-seq: n=36 vs. 41 biologically independent

samples), del11q vs. non-del11q (H3K27ac: n=8 vs. 30, ATAC-seq: n=8 vs. 29, RNA-seq: n=7 vs. 26 biologically independent samples), amp2p vs. non-amp2p (H3K27ac: n=5 vs. 33, ATAC-seq: n=5 vs. 32, RNA-seq: n=5 vs. 28 biologically independent samples).

Supplementary Table 13. Regions considered IG or SHM target loci in this study. For each gene associated with the SHM, the gencode22 location (extended 1MB up- and downstream) was used**,** SHM = somatic hypermutation machinery, IG = immunoglobulin**.** Other regions (Fig. 6h) are defined as those that do not appear in this table.

Supplementary Table 14. Data quality measures. Data quality measures of (a) Chip-seq and ATAC-seq, (b) whole genome bisulfite sequencing and (c) RNA-seq data.