

## **SUPPLEMENTAL MATERIAL**

### Automatic identification of informative regions with epigenomic changes associated to hematopoiesis

Enrique Carrillo-de-Santa-Pau<sup>1,6</sup>, David Juan<sup>1,2,6</sup>, Vera Pancaldi<sup>1,3</sup>, Felipe Were<sup>1</sup>, Ignacio Martin-Subero<sup>4</sup>, Daniel Rico<sup>5,7</sup> & Alfonso Valencia<sup>1,3,7</sup> on behalf of The BLUEPRINT Consortium<sup>8</sup>

<sup>1</sup>Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO)

<sup>2</sup>Institut de Biología Evolutiva, Consejo Superior de Investigaciones Científicas–Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, Barcelona, Catalonia 08003, Spain

<sup>3</sup>Barcelona Supercomputing Centre (BSC), Barcelona, Spain.

<sup>4</sup>Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Department of Anatomic Pathology, Pharmacology and Microbiology, University of Barcelona, Barcelona, Spain.

<sup>5</sup>Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom.

<sup>6</sup>Co-first authors

<sup>7</sup>Co-senior and corresponding authors [daniel.rico@newcastle.ac.uk](mailto:daniel.rico@newcastle.ac.uk), [alfonso.valencia@bsc.es](mailto:alfonso.valencia@bsc.es)

<sup>8</sup>Member of International Human Epigenome Consortium (IHEC). For more information, see <http://www.blueprint-epigenome.eu> and <http://ihec-epigenomes.org>

## Evaluation of different strategies for clustering hematopoietic samples

One of the advantages of our MCA approach is that it can be used for dealing with any type of categorical multivariate data. It means that other sources of epigenomic data associated to cell differentiation could also be used to explore their ability to discriminate different cell types. In the manuscript, we show results using five chromatin states aggregated by manual curation out of the 11 chromatin states obtained from ChromHMM. These analyses use six different ChIP-seq experiments per sample allowing us to explore how different types of epigenomic states contribute to hematopoiesis. However, for different experimental setups, it is also interesting to explore how much every histone combination contributes to cellular differentiation.

In order to address this point, we first performed MCA analyses based on each of the 11 original chromatin states. It is important to realize that these states are defined out of the combination of the six used histone marks and that they couldn't be defined independently. However, each of these chromatin states represents a good discrimination of different histone combinatorics. As a consequence, it might be reasonable to assume that only some of these states are the ones driving hematopoietic cell specificity. For each chromatin state we collapse all the other chromatin states in an "other" state. We run our MCA protocol for each of these two-states genome segmentations. Our results show that genomic distributions of none of the eleven original states are able to provide a good separation of the seven cell types (see **S3 Fig**). As expected from previous works (Kundaje et al., 2015; Lara-Astiaso et al., 2014), elongation (E1 and E2) and enhancer states (E8 and E9) show some discriminatory power. In particular, enhancer states MCA-spaces show a coherent distribution of all the cell types, while elongation states are more noisy. However, our approach fails to fully take advantage of this information because for these states, the second component is not informative enough. This means that while many E1-elongation and E8- and E9-enhancer regions are able to discriminate Monocytes and Macrophages from the rest of the cell types, far fewer contain information about the segregation of the other cell types. Interestingly, the E2-elongation state seems to contain information to discriminate B cells, but not the other cell types. Importantly, even using this second axis, it seems impossible to discriminate all the cell types in a non-supervised way in any of the cases.

Given the differences among these MCA states and the one based on the five aggregated states used in the main manuscript, we decided to repeat the same analysis but using the four aggregated states with more than one original chromatin state (Repressed/Polycomb state remains unaggregated, see **S4 Fig**). Our results show that only the elongation state (E1 and E2) is able to separate all the cell types (see **S4A Fig**). In fact, in the MCA space based on elongation states, different cell types are clearly separated and distributed as a mirror image of the MCA space used in the main text. It is evident, that combining the complementary information from the elongation states is key for this analysis. In contrast, combination of the enhancer states does not improve their ability to separate different cell types, because they contain more redundant information (see **S4C Fig**). This is an interesting observation because the main difference between E1- and E2-elongation states is the higher enrichment of peaks of H3K36me3 in E2 state. These results suggest that difficulties in discriminating cell types based on transcriptional information could be related to variability in the levels of gene expression in different samples of the same cell type that might mask the transcriptional signature of cell type specificity

We have confirmed that elongation and enhancer states are the most informative ones for cell type discrimination and both of them can be used to explore cell specificity. Therefore, could we use just the histone peaks that are known to define these regions? In order to answer this question, we first performed the same MCA analysis segmenting the genome into H3K36me3 peak regions as defined by chromHMM (see **S5A Fig**). Surprisingly, our results show that the MCA space based on H3K36me3 peaks does not discriminate different cell types. This implies that chromHMM states are capturing some important information about transcription elongation missing in the peaks. ChromHMM is based on a multivariate HMM that takes advantage of neighboring regions to define chromatin states and classify different regions. We suggest that this feature might be particularly suitable for detecting regions where transcription elongation is taking place. In such scenario, E1 and E2 state could reflect different relationships with the “moving” peaks of H3K36me3, that when combined allows to provide a more robust definition of elongating regions. In addition, we also analysed those regions presenting peaks of H3K4me1 but not of H3K4me3 (an enhancer definition previously used in Lara-Astiaso et al., 2014). The obtained MCA space is highly consistent with those obtained for any of the enhancer states (see **S5B Fig**). This means that enhancers, although less informative of cell type identity, are more robust than elongation to different definitions. As a whole, we have shown that using chromatin states is a valuable step for taking advantage of gene transcription as an important source of cell type specificity and that our approach is flexible enough to be used in different experimental designs

## Supporting Figures

S1 Fig. Related to Fig 1. Chromatin States definitions emission probabilities and genomic annotation enrichments.

- A) Heatmap with emission probabilities and genomic annotation enrichments
- B) Table with ChromHMM emission probabilities
- C) Chromatin States descriptions based on emission probabilities and genomic annotation enrichment

S2 Fig. Related to Fig 2. MCA by chromosome

S3 Fig. Related to Fig 2. Sample spaces from Multiple Correspondence Analysis for each chromatin state

S4 Fig. Related to Fig 2. Sample spaces from Multiple Correspondence Analysis for the four aggregated chromatin states

S5 Fig. Related to Fig 2. Sample spaces from Multiple Correspondence Analysis for histone peaks defining elongating and enhancer regions

S6 Fig. Related to Fig 3. The most abundant Chromatin Determinant Region patterns

S7 Fig. Related to Fig 3. Boxplot with the sizes of the Chromatin Determinant Region patterns

S8 Fig. Related to Fig 2. Diagram with the cell types disposition in the MCA and the number of cell type and lineage specific CDRs

S9 Fig. Related to Fig 3. UCSC screenshots with CDRs affecting ABHD16B (A) and LINC00494 (B)

S10 Fig. Related to Fig 2. Boxplots showing ATAC-seq signal across cell types in CDRs overlapping with the regions in the CYBERSORT signature matrix from Corces et al (2016).

S11 Fig. Related to Fig 3. Chromatin Determinant Regions genomic annotation

- A) Histogram with CDRs distance to closest Transcription Start sites
- B) Barplot with CDRs proportion to genomic annotation
- C) Table with CDRs genomic annotation and loci enrichment

S12 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active HSCs Chromatin Determinant Regions

S13 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active T cell Chromatin Determinant Regions

S14 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Naïve B cells Chromatin Determinant Regions

S15 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Germinal Centre B cells Chromatin Determinant Regions

S16 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Neutrophil Chromatin Determinant Regions

S17 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-

represented in active Monocyte Chromatin Determinant Regions

S18 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Macrophage Chromatin Determinant Regions

S19 Fig. Related to Fig 3. ERG2 expression in macrophages (red), monocytes (blue) and neutrophils (green) from BLUEPRINT data analysis portal (<http://blueprint-data.bsc.es>)

S20 Fig. Related to Fig 4. Epigenomes of leukemias projected into the “Healthy” blood cells chromatin space

S21 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in commonly altered regions in leukemia

S22 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in AML alterations

S23 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in CLL alterations

S24 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in MCL alterations

## **Supporting Tables**

- S1 Table. Related to Fig 2. Healthy and disease samples included in the analyses and ChIP-seq quality measures. (In S1\_Table.xls)
- S2 Table. Related to Fig 1. Correlation across ChromHMM training models with 5, 7, 9, 11, 13 and 20 chromatin states. (see below)
- S3 Table. Related to Fig 1. Correlations between ChromHMM training models with 20 and 11 chromatin states. (see below)
- S4 Table. Related to Fig 3. Chromatin Determinant Regions (In S4\_Table.xls)
- S5 Table. Related to Fig 3. Frequencies for Chromatin Determinant Region patterns (In S5\_Table.xls)
- S6 Table. Related to Fig 3. Chromatin Determinant Region annotations (In S6\_Table.csv)
- S7 Table. Related to Fig 3. Differential expression analysis on 7764 genes with associated CDRs and 28 Transcription Factors in figure 3C (In S7\_Table.xls)
- S8 Table. Related to Fig 3. Gene Ontology (Biological Process) & REACTOME enrichment results from HOMER by cell type (In S8\_Table.xls)
- S9 Table. Related to Fig 4. Leukemia-altered Chromatin Determinant Regions annotations (In S9\_Table.xlsx)
- S10 Table. Related to Fig 4. COSMIC and MSigDB enrichments for leukemia-altered CDRs (In S10\_Table.xls)
- S11 Table. Related to Fig 4. Gene Ontology-biological process and REACTOME enrichments for leukemia-altered CDRs (In S11\_Table.xls)

## **Supporting Files**

- S1\_File. Related to Fig 3. HOMER enrichment analyses for Chromatin Determinant Regions (In S1\_File.zip)

**Table S2.**- Correlations across ChromHMM training models with 5, 7, 9, 11, 13 and 20 chromatin states

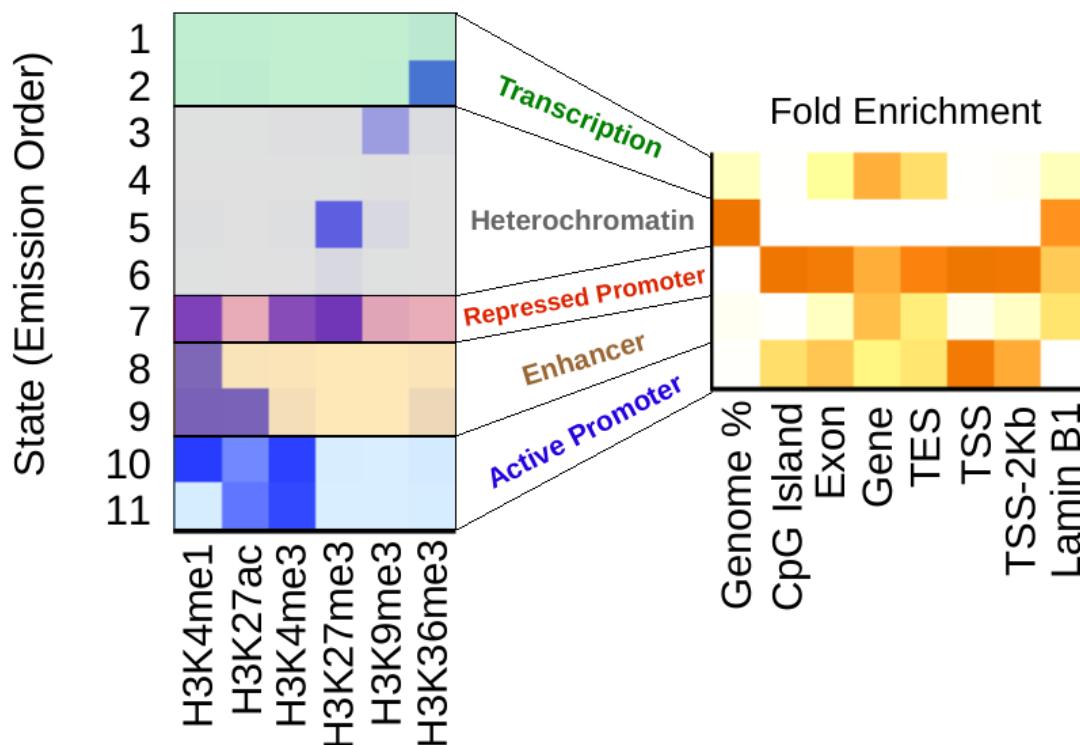
Model 20 states	emissions_5_states	emissions_7_states	emissions_9_states	emissions_11_states	emissions_13_states	emissions_20_states	Model 11 higher state
<b>1</b>	0.41	0.77	1.00	0.99	1.00	1.00	<b>7</b>
<b>2</b>	0.83	0.98	1.00	1.00	1.00	1.00	<b>5 &amp; 6</b>
<b>3</b>	0.84	0.98	1.00	1.00	1.00	1.00	<b>5 &amp; 6</b>
<b>4</b>	0.95	0.95	0.99	0.83	0.82	1.00	<b>4</b>
<b>5</b>	0.84	1.00	1.00	1.00	1.00	1.00	<b>3</b>
<b>6</b>	0.83	1.00	1.00	1.00	1.00	1.00	<b>3</b>
<b>7</b>	0.62	0.79	0.78	0.78	0.78	1.00	<b>3</b>
<b>8</b>	1.00	1.00	1.00	1.00	1.00	1.00	<b>2</b>
<b>9</b>	1.00	1.00	1.00	1.00	1.00	1.00	<b>2</b>
<b>10</b>	0.89	0.89	0.95	0.95	0.95	1.00	<b>1</b>
<b>11</b>	0.66	0.65	0.75	0.75	1.00	1.00	<b>1</b>
<b>12</b>	0.97	0.99	0.99	0.99	0.98	1.00	<b>8</b>
<b>13</b>	0.98	1.00	1.00	1.00	1.00	1.00	<b>8</b>
<b>14</b>	0.98	1.00	1.00	1.00	1.00	1.00	<b>8</b>
<b>15</b>	0.82	0.99	0.99	0.99	0.99	1.00	<b>9</b>
<b>16</b>	0.42	0.85	0.87	0.85	0.88	1.00	<b>9</b>
<b>17</b>	0.92	0.90	0.90	0.95	1.00	1.00	<b>10</b>
<b>18</b>	0.74	0.71	0.69	0.92	0.93	1.00	<b>10</b>
<b>19</b>	0.83	0.85	0.85	0.83	0.89	1.00	<b>11</b>
<b>20</b>	0.85	0.88	0.89	0.98	1.00	1.00	<b>11</b>

**Table S3.**- Correlations between ChromHMM training models with 20 and 11 chromatin states. Maximum correlation in the 20 model states are marked in bold.

	1_model11	2_model11	3_model11	4_model11	5_model11	6_model11	7_model11	8_model11	9_model11	10_model11	11_model11	Model 11 higher state
<b>1_model20</b>	-0.463	-0.469	-0.412	-0.221	0.644	0.652	<b>0.991</b>	0.378	-0.073	0.318	-0.055	<b>7</b>
<b>2_model20</b>	-0.317	-0.234	-0.142	0.127	<b>1.000</b>	<b>1.000</b>	0.538	-0.224	-0.389	-0.449	-0.335	<b>5 &amp; 6</b>
<b>3_model20</b>	-0.329	-0.246	-0.128	0.143	<b>1.000</b>	<b>1.000</b>	0.530	-0.227	-0.385	-0.454	-0.337	<b>5 &amp; 6</b>
<b>4_model20</b>	-0.512	-0.391	0.649	<b>0.833</b>	0.649	0.645	0.080	-0.401	-0.555	-0.660	-0.414	<b>4</b>
<b>5_model20</b>	-0.281	-0.206	<b>0.998</b>	0.968	-0.115	-0.120	-0.383	-0.236	-0.387	-0.457	-0.333	<b>3</b>
<b>6_model20</b>	-0.300	-0.225	<b>0.997</b>	0.966	-0.125	-0.130	-0.391	-0.244	-0.374	-0.443	-0.305	<b>3</b>
<b>7_model20</b>	0.396	0.490	<b>0.780</b>	0.630	-0.292	-0.309	-0.643	-0.373	-0.550	-0.632	-0.419	<b>3</b>
<b>8_model20</b>	0.977	<b>1.000</b>	-0.156	-0.316	-0.229	-0.244	-0.483	-0.193	-0.222	-0.403	-0.296	<b>2</b>
<b>9_model20</b>	0.979	<b>1.000</b>	-0.156	-0.317	-0.236	-0.250	-0.486	-0.186	-0.213	-0.399	-0.297	<b>2</b>
<b>10_model20</b>	<b>0.947</b>	0.885	-0.104	-0.264	-0.352	-0.356	-0.506	0.190	0.152	-0.286	-0.502	<b>1</b>
<b>11_model20</b>	<b>0.747</b>	0.602	-0.400	-0.545	-0.433	-0.426	-0.130	0.644	0.518	0.230	-0.335	<b>1</b>
<b>12_model20</b>	0.157	-0.023	-0.290	-0.338	-0.281	-0.262	0.320	<b>0.986</b>	0.671	0.540	-0.318	<b>8</b>
<b>13_model20</b>	-0.013	-0.188	-0.246	-0.266	-0.219	-0.198	0.420	<b>1.000</b>	0.660	0.588	-0.280	<b>8</b>
<b>14_model20</b>	-0.032	-0.204	-0.241	-0.254	-0.193	-0.173	0.447	<b>0.999</b>	0.640	0.583	-0.288	<b>8</b>
<b>15_model20</b>	-0.113	-0.307	-0.385	-0.394	-0.360	-0.336	0.036	0.693	<b>0.994</b>	0.637	0.166	<b>9</b>
<b>16_model20</b>	-0.117	-0.247	-0.324	-0.320	-0.314	-0.297	-0.311	0.158	<b>0.847</b>	0.372	0.411	<b>9</b>
<b>17_model20</b>	-0.282	-0.413	-0.502	-0.558	-0.479	-0.463	0.227	0.462	0.726	<b>0.946</b>	0.708	<b>10</b>
<b>18_model20</b>	-0.238	-0.333	-0.372	-0.438	-0.339	-0.330	0.607	0.666	0.342	<b>0.920</b>	0.388	<b>10</b>
<b>19_model20</b>	-0.306	-0.262	-0.265	-0.334	-0.251	-0.258	0.348	-0.131	-0.160	0.643	<b>0.831</b>	<b>11</b>
<b>20_model20</b>	-0.297	-0.302	-0.347	-0.392	-0.345	-0.344	-0.085	-0.282	0.272	0.561	<b>0.982</b>	<b>11</b>

S1 Fig. Related to Fig 1. Chromatin States definitions, emission probabilities and genomic annotation enrichments.

A



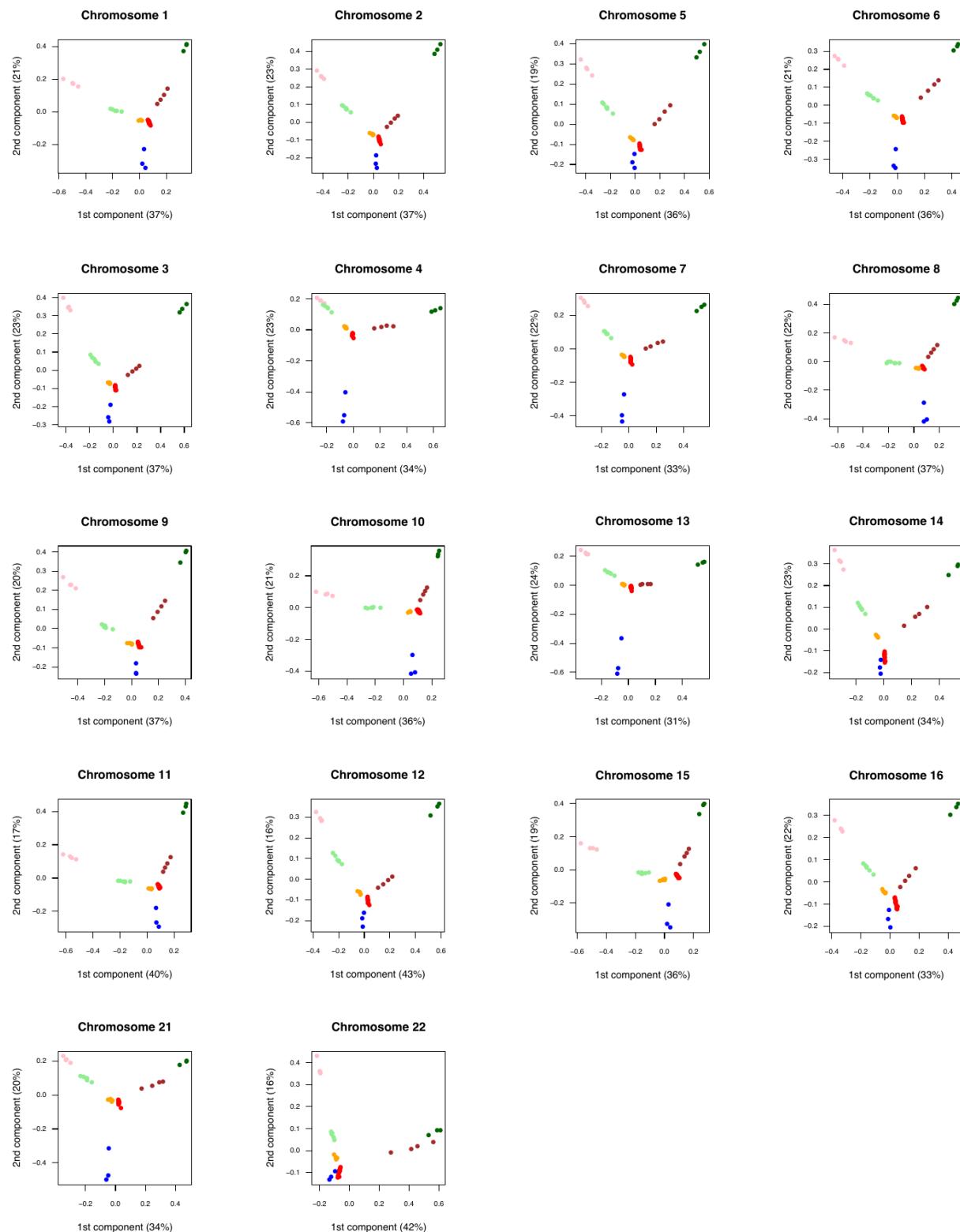
B

STATES	H3K4me1	H3K27ac	H3K4me3	H3K27me3	H3K9me3	H3K36me3
1	0.0083789397	0.0055690833	6.66E-004	2.86E-004	8.92E-004	0.0377074823
2	0.0134763455	0.0161736305	8.62E-004	6.75E-004	0.0072781926	0.6605666861
3	0.0013315356	0.0026166689	0.0096648726	0.0135653097	0.363112689	0.0246882243
4	5.62E-004	6.76E-004	4.93E-004	0.0015522292	0.0042164807	3.74E-004
5	0.0130833231	0.0011377025	0.0127860985	0.6944957102	0.0417521816	0.0010188047
6	0.0021542927	0.0011527494	0.0011525308	0.0408574485	0.0029809303	3.06E-004
7	0.6037590334	0.0332430091	0.5485884971	0.6684052213	0.0704342588	0.0242053224
8	0.6912545251	0.0280633723	0.0173963982	0.0015834742	0.0011606909	0.0171964967
9	0.720097335	0.7178205379	0.056774834	8.38E-004	0.0028630358	0.0958220634
10	0.9557643195	0.5535936307	0.9338481998	0.0023170696	0.0010476375	0.0291980914
11	0.0175188988	0.6510026835	0.8955382763	0.0011644844	5.73E-004	0.0151994013

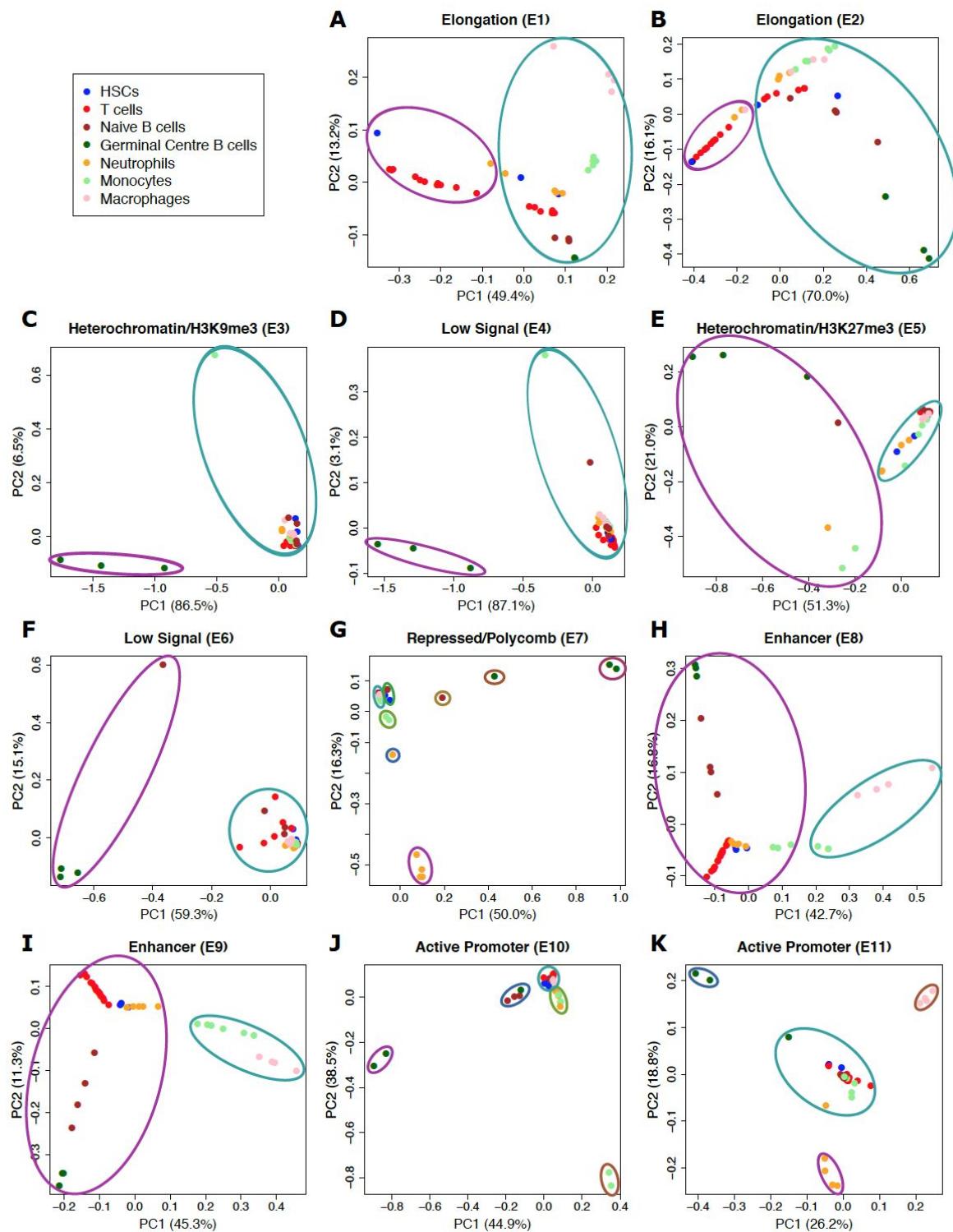
C

<b>State1</b>	Transcription Low signal H3K36me3
<b>State2</b>	Transcription High signal H3K36me3
<b>state3</b>	Heterochromatin High Signal H3K9me3
<b>State4</b>	Low signal
<b>State5</b>	Heterochromatin High signal H3K27me3
<b>State6</b>	Heterochromatin Low signal H3K27me3
<b>State7</b>	Repressed Polycomb Promoter High Signal H3K4me3, H3K4me1 and H3K27me3
<b>State8</b>	Enhancer High Signal H3K4me1
<b>State9</b>	Active Enhancer High Signal H3K4me1 & H3K27Ac
<b>State10</b>	Distal Active Promoter (2Kb) High Signal H3K4me3 & H3K27Ac & H3K4me1
<b>State11</b>	Active TSS High Signal H3K4me3 & H3K27Ac

S2 Fig. Related to Fig 2. MCA by chromosome



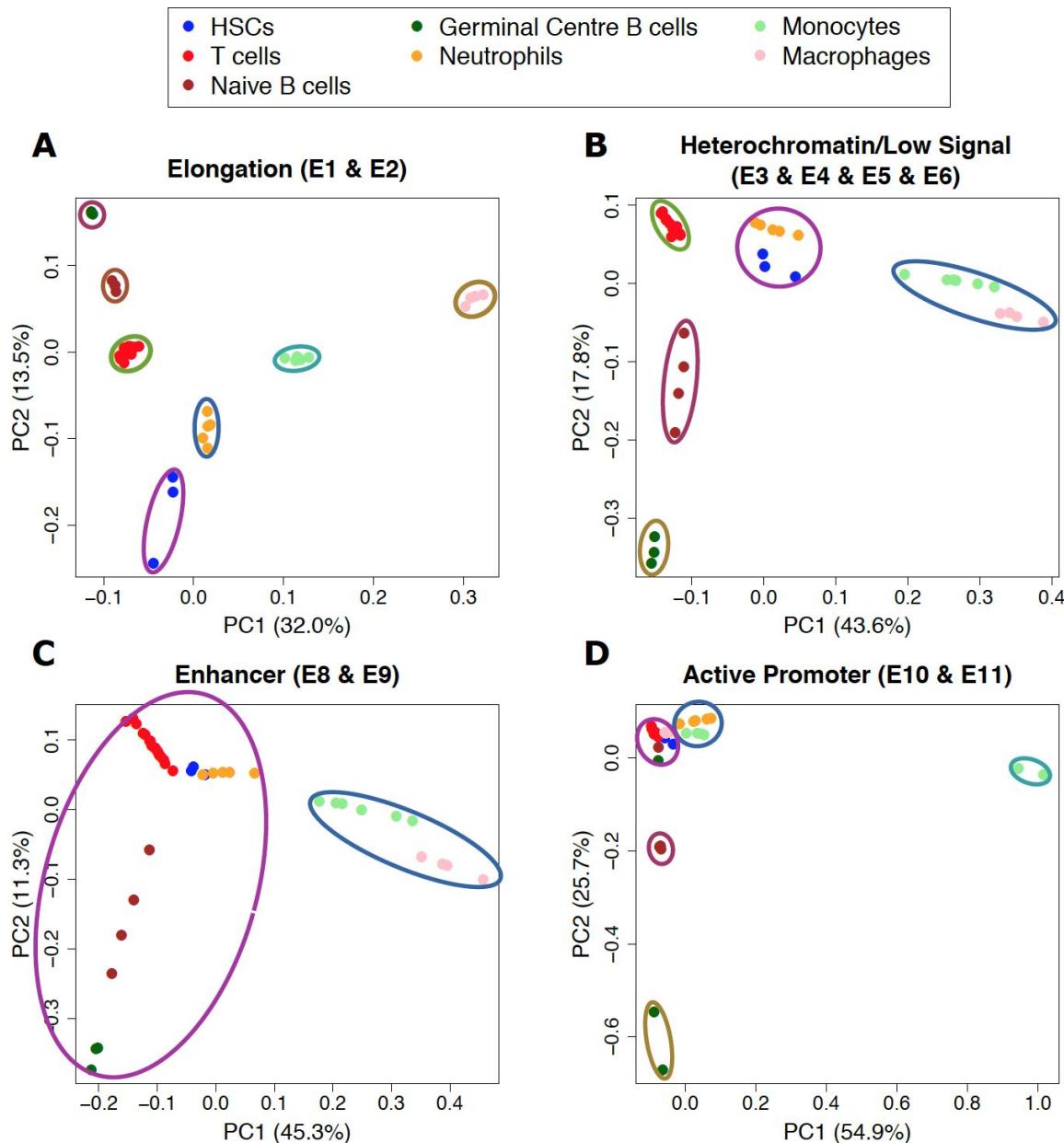
S3 Fig. Related to Fig 2. Sample spaces from Multiple Correspondence Analysis for each chromatin state



Clustering of the samples in MCA spaces for each of eleven chromatin states using the same protocol explained in the main text. Clusters are represented by colored ellipses

surrounding the corresponding samples. For the sake of simplicity, we show the first two components for each MCA space, although the actual number of dimensions is determined independently according to the protocol and might differ for different spaces. **A)** Clustering of the samples in the MCA space for state E1 (elongation). Our protocol selected the first component as informative and detected two clusters unrelated to cell type specificity. Visual inspection including the second component shows certain segregation of Monocyte, Macrophage and Germinal Centre B cells samples. **B)** Clustering of the samples in the MCA space for state E2 (elongation). Our protocol selected the first component as informative and detected two clusters unrelated to cell type specificity. Visual inspection including the second component shows certain segregation of Germinal Centre B cells samples. **C)** Clustering of the samples in the MCA space for state E3 (heterochromatin/H3K9me3). Our protocol selected the first component as informative and detected two clusters separating Germinal Centre B cells samples from the other cell types. **D)** Clustering of the samples in the MCA space for state E4 (low signal). Our protocol selected the first component as informative and detected two clusters separating Germinal Centre B cells samples from the other cell types. **E)** Clustering of the samples in the MCA space for state E5 (heterochromatin/H3K27me3). Our protocol selected the two first components as informative and detected two clusters unrelated to cell type specificity. Visual inspection shows certain segregation of Germinal Centre B cells samples. **F)** Clustering of the samples in the MCA space for state E6 (low signal). Our protocol selected the first component as informative and detected two clusters unrelated to cell type specificity. Visual inspection including the second component shows certain segregation of Germinal Centre B cells samples. **G)** Clustering of the samples in the MCA space for state E7 (repressed/Polycomb). Our protocol selected the two first components as informative and detected eight clusters unrelated to cell type specificity. Visual inspection shows certain segregation of Germinal Centre B cells and Neutrophils samples. **H)** Clustering of the samples in the MCA space for state E8 (enhancer). Our protocol selected the first component as informative and detected two clusters unrelated to cell type specificity. Visual inspection including the second component shows certain segregation of Macrophages, Neutrophils, Naive B cells and Germinal Centre B cells samples. **I)** Clustering of the samples in the MCA space for state E9 (enhancer). Our protocol selected the first component as informative and detected two clusters separating Monocytes and Macrophages from the other cell types. Visual inspection including the second component shows certain segregation of Macrophages, Neutrophils, Naive B cells and Germinal Centre B cells samples. **J)** Clustering of the samples in the MCA space for state E10 (active promoter). Our protocol selected the two first components as informative and detected five clusters poorly related to cell specificity. **K)** Clustering of the samples in the MCA space for state E11 (active promoter). Our protocol selected the three first components as informative and detected four clusters separating Macrophage samples and some Neutrophils and Germinal Centre B cells samples from the other cell types.

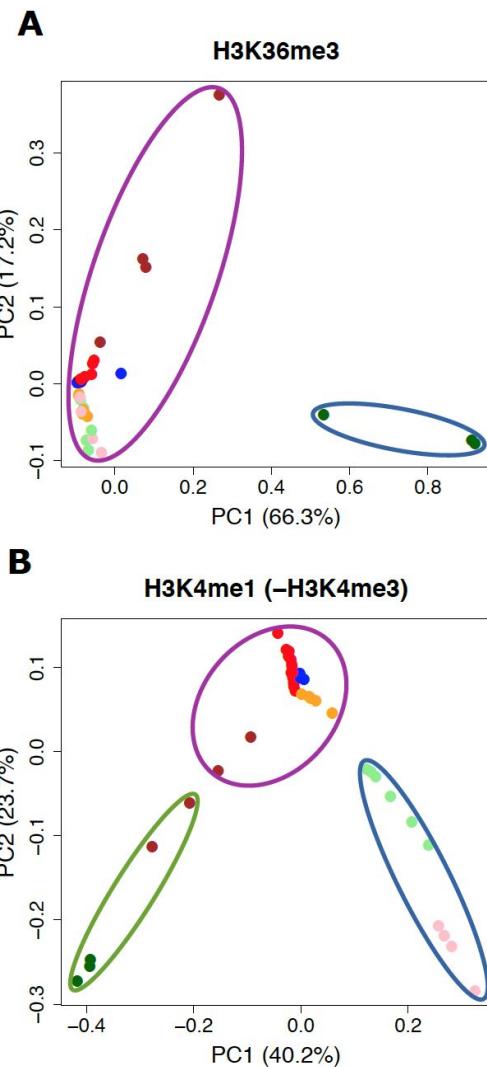
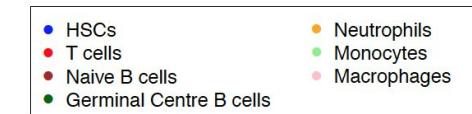
S4 Fig. Related to Fig 2. Sample spaces from Multiple Correspondence Analysis for the four aggregated chromatin states



Clustering of the samples in MCA spaces for each of four aggregated chromatin states, using the same protocol explained in the main text. Clusters are represented by colored ellipses surrounding the corresponding samples. For the sake of simplicity, we show the first two components for each MCA space, although the actual number of dimensions is determined independently according to the protocol and might differ for different spaces. **A)** Clustering of the samples in the MCA space for states E1 and E2 (elongation). Our protocol selected the two first components as informative and detected seven clusters perfectly

segregating all cell types. **B)** Clustering of the samples in the MCA space for state E3, E4, E5 and E6 (heterochromatin/low signal). Our protocol selected the first two components as informative and detected five clusters separating Germinal Centre B cells, Naive B cells and T cells samples. Visual inspection shows certain segregation also for Monocyte, Macrophages, Neutrophils and HSCs samples. **C)** Clustering of the samples in the MCA space for states E8 and E9 (enhancer). Our protocol selected the first component as informative and detected two clusters separating Monocytes and Macrophages from the other cell types. Visual inspection including the second component shows certain segregation also for T cells, Naive B cells and Germinal Centre B cells samples. **D)** Clustering of the samples in the MCA space for states E10 and E11 (active promoter). Our protocol selected the two first components as informative and detected five clusters poorly related to cell specificity.

S5 Fig. Related to Fig 2. Sample spaces from Multiple Correspondence Analysis for histone peaks defining elongating and enhancer regions



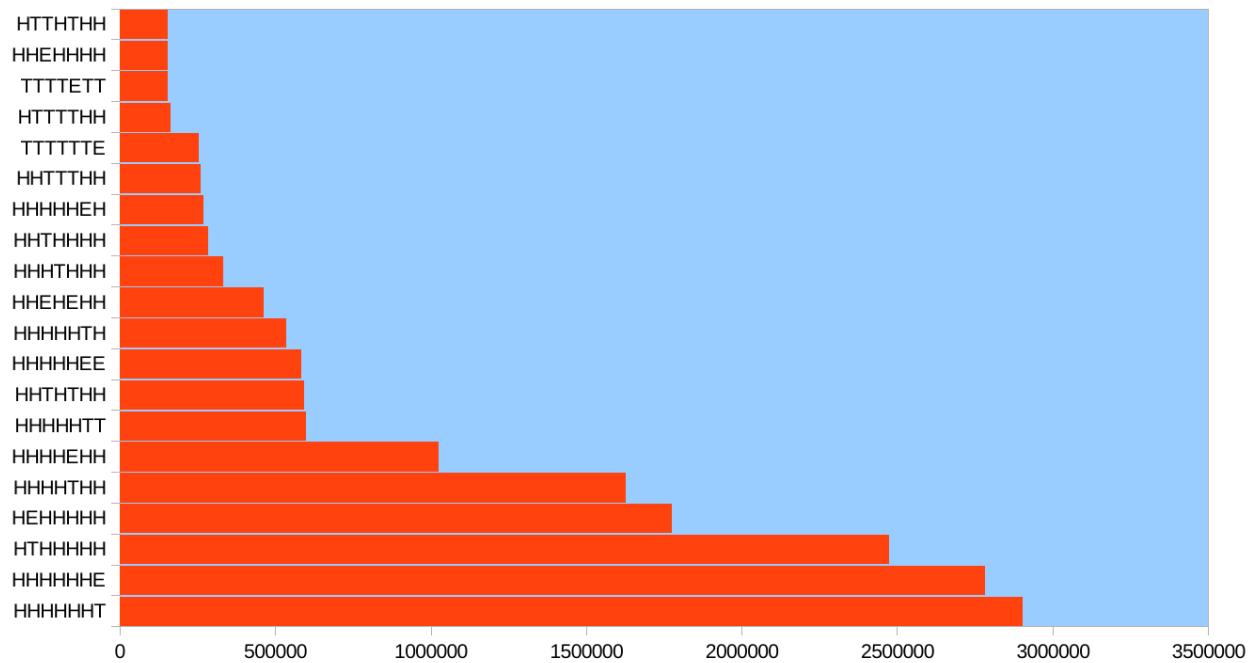
Clustering of the samples in MCA spaces for two selections of histone peaks using the same protocol explained in the main text. Clusters are represented by colored ellipses surrounding the corresponding samples. For the sake of simplicity, we show the first two components for each MCA space, although the actual number of dimensions is determined independently according to the protocol and might differ for different spaces. **A)** Clustering of the samples in the MCA space for H3K36me3 peaks. Our protocol selected the first three components as informative and detected two clusters separating Germinal Centre B cells samples from the other cell types. **B)** Clustering of the samples in the MCA space for H3K4me1 peaks non-overlapping with H3K4me3 peaks. Our protocol selected the first two components as

informative and detected three clusters poorly related to cell specificity. Visual inspection shows certain segregation for Naive B cells, Germinal Centre B cells, Monocytes and Macrophages samples.

S6 Fig. Related to Fig 3. The most abundant Chromatin Determinant Region patterns

Each position in the pattern represents: T cell/HSC/Monocyte/Neutrophil/Macrophage/Naïve B cells/Germinal Centre B cells

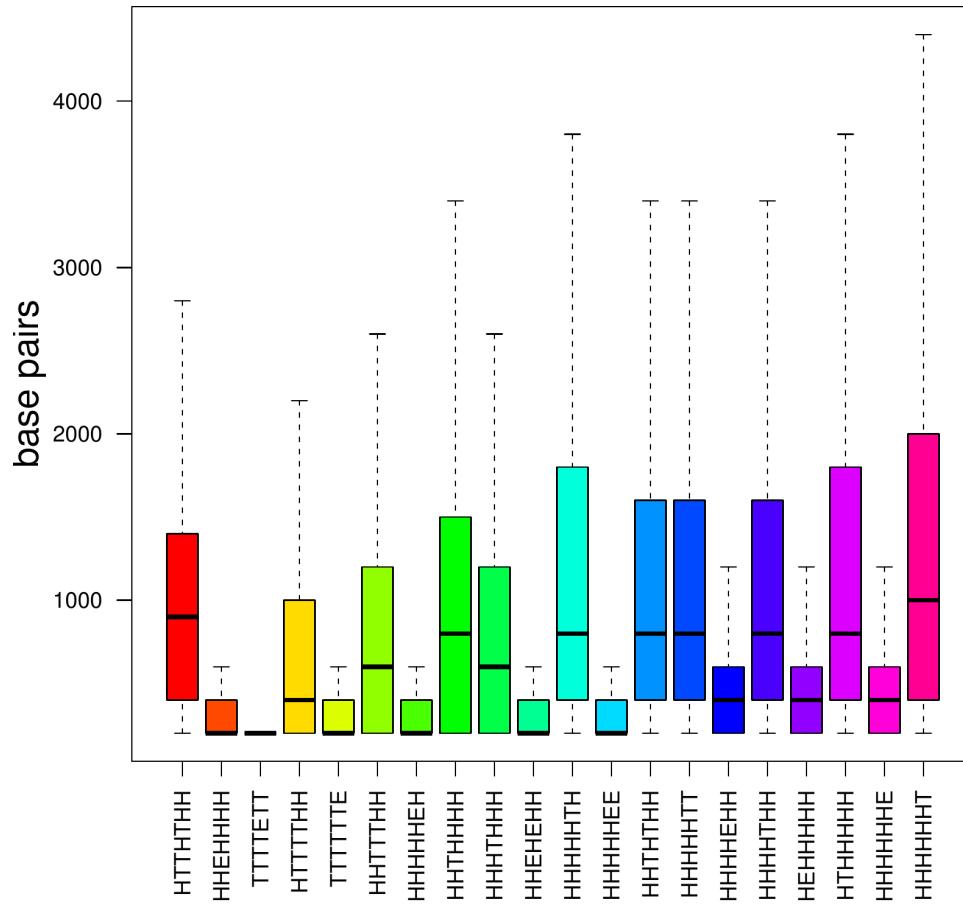
\*A=Active promoter; E=enhancer; H=Heterocromatin; R=Repressed promoter; T=transcription



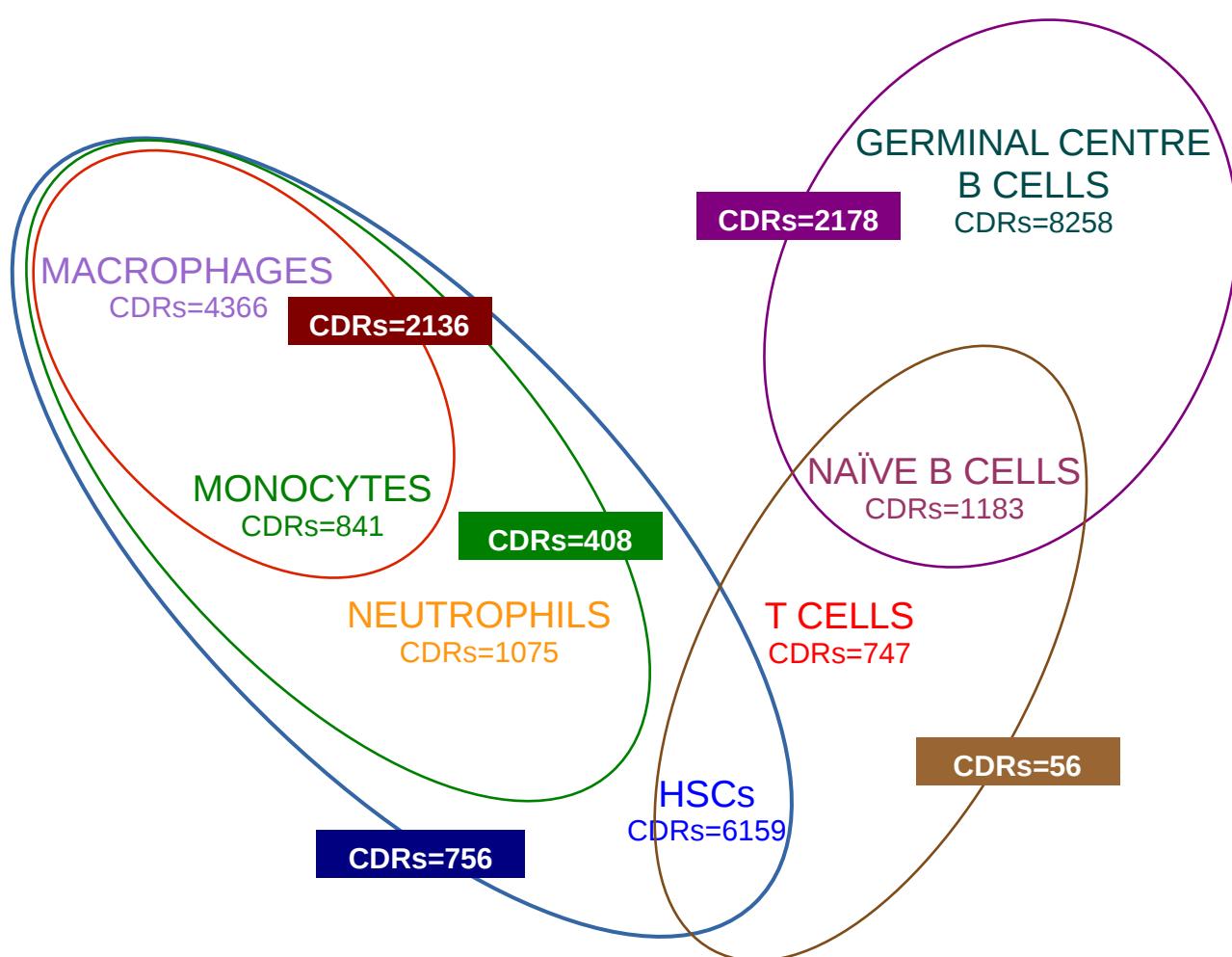
## S7 Fig. Related to Fig 3. Boxplot with the sizes of the Chromatin Determinant Region patterns

Each position in the pattern represents: T cell/HSC/Monocyte/Neutrophil/Macrophage/Naïve B cells/Germinal Centre B cells

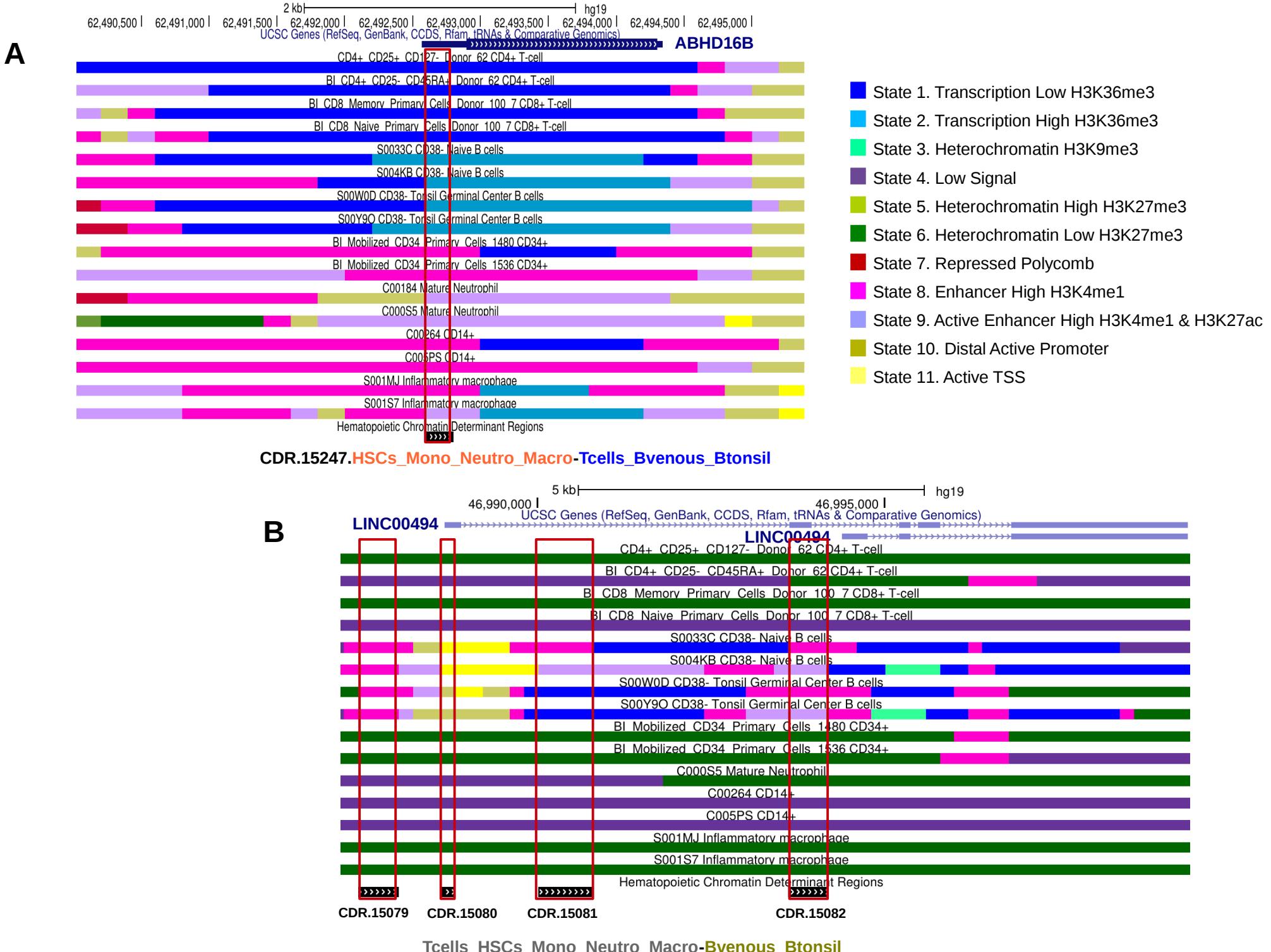
\*A=Active promoter; E=enhancer; H=Heterochromatin; R=Repressed promoter; T=transcription



S8 Fig. Related to Fig 3. Diagram with the cell types disposition in the MCA and the number of cell type and lineage specific CDRs

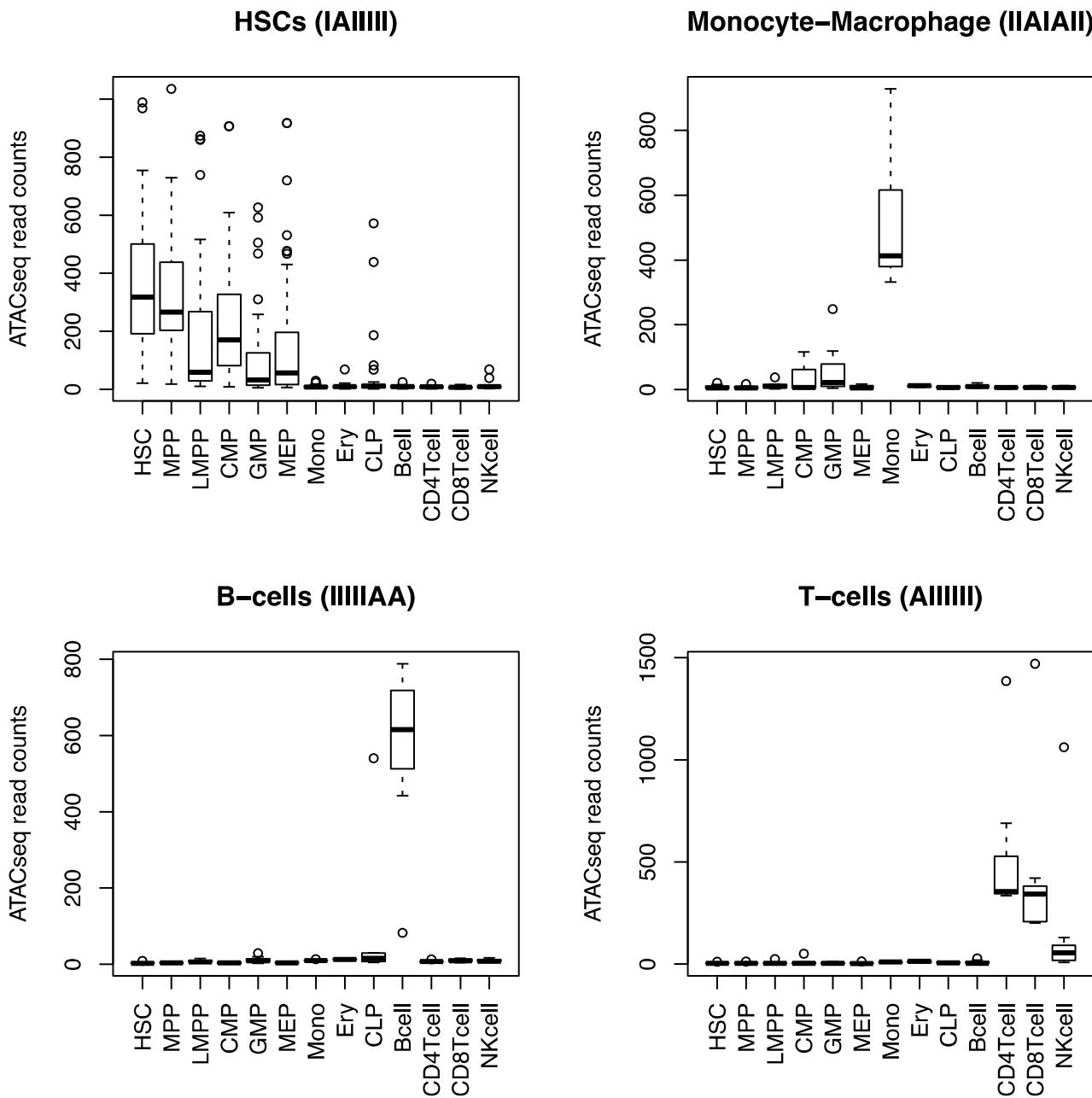


S9 Fig. Related to Fig 3. UCSC screenshots with CDRs affecting ABHD16B (A) and LINC00494 (B)

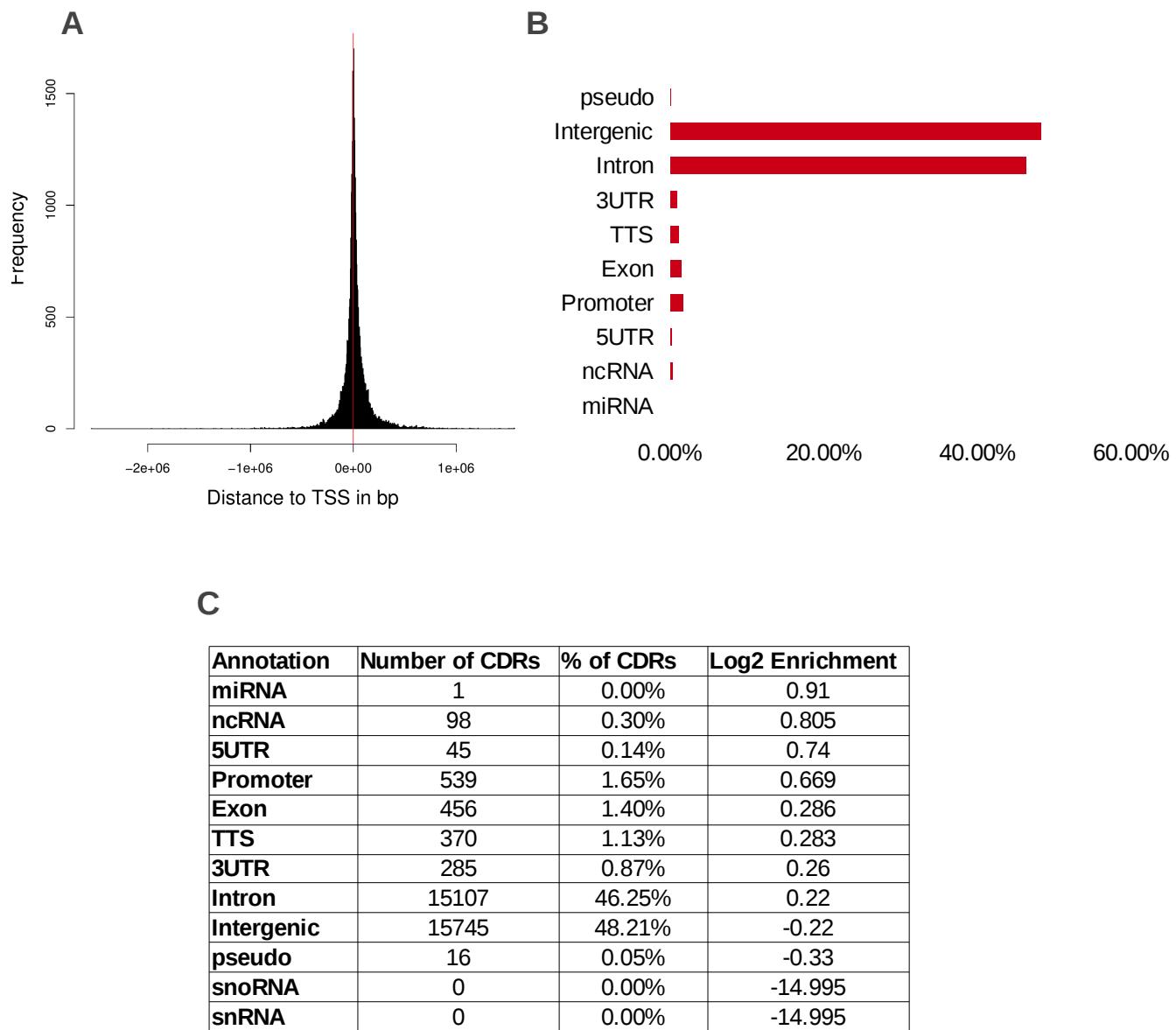


S10 Fig. Related to Fig 2. Boxplots showing ATAC-seq signal across cell types in CDRs overlapping with the regions in the CYBERSORT signature matrix from Corces et al (2016).

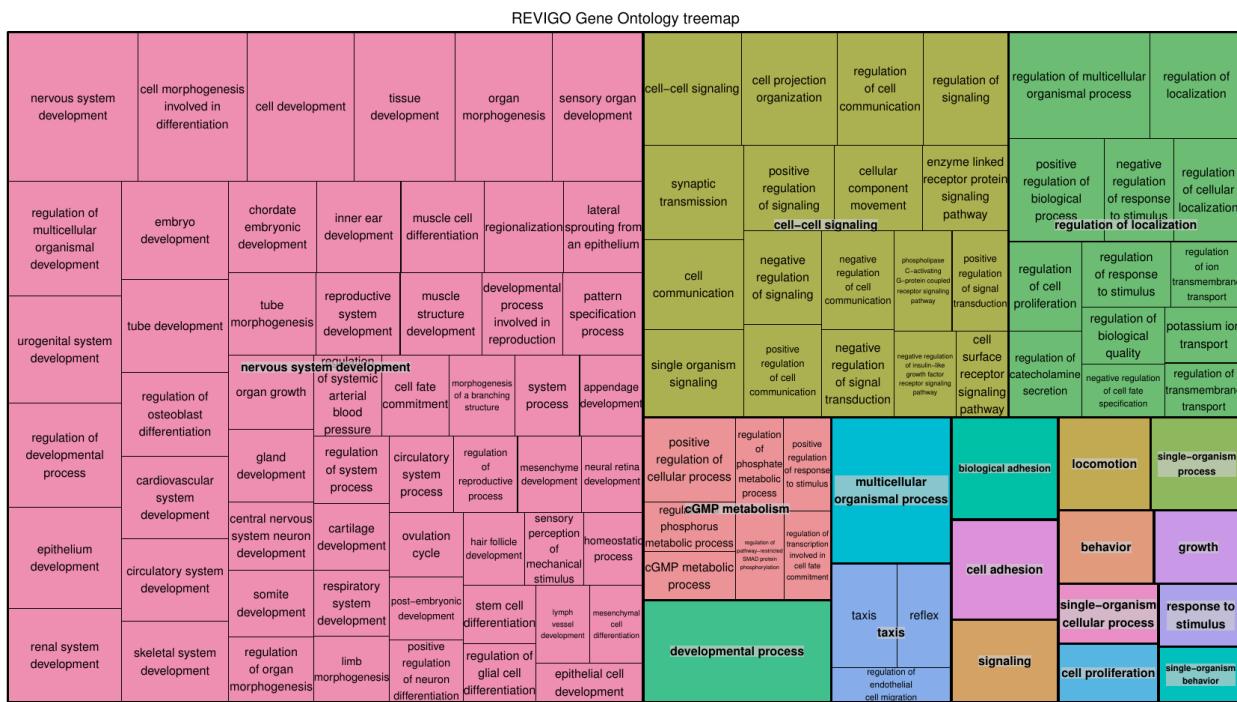
Top left with CDRs active in HSCs; Top-right with CDRs active in monocytes and macrophages;  
 Bottom left with CDRs active in B cells; Bottom right with CDRs active in T cells



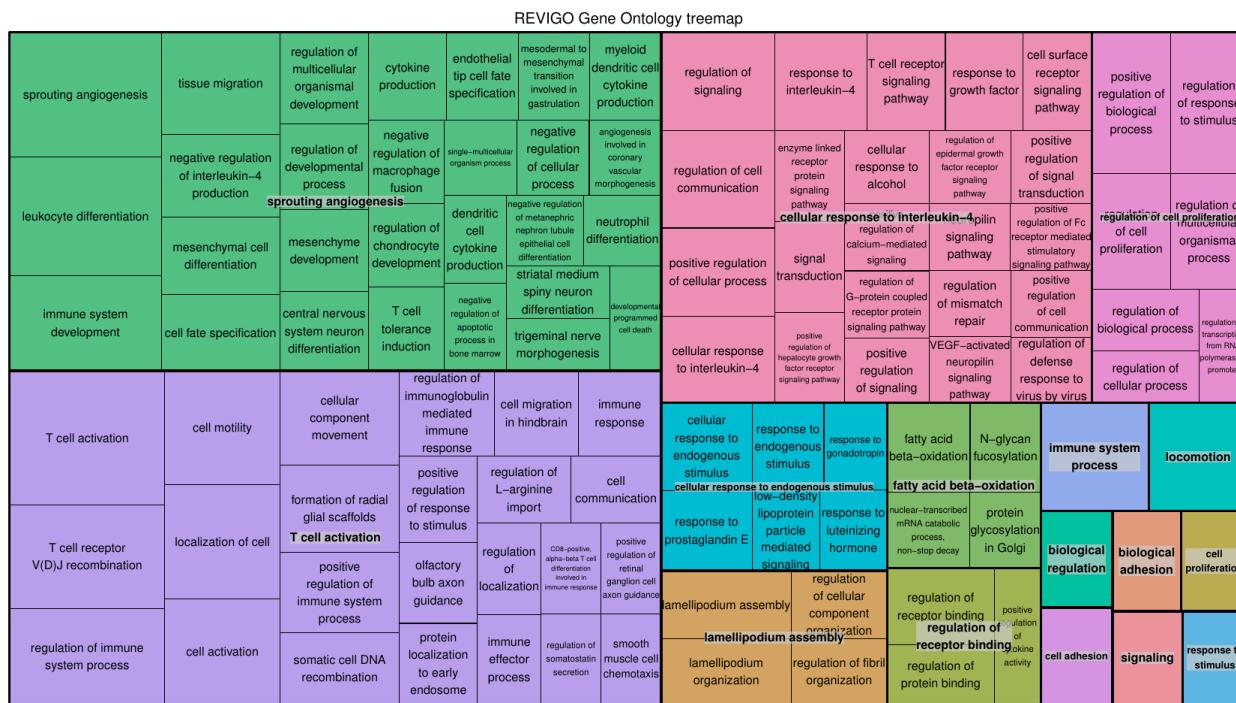
S11 Fig. Related to Fig 3. Chromatin Determinant Region genomic annotation



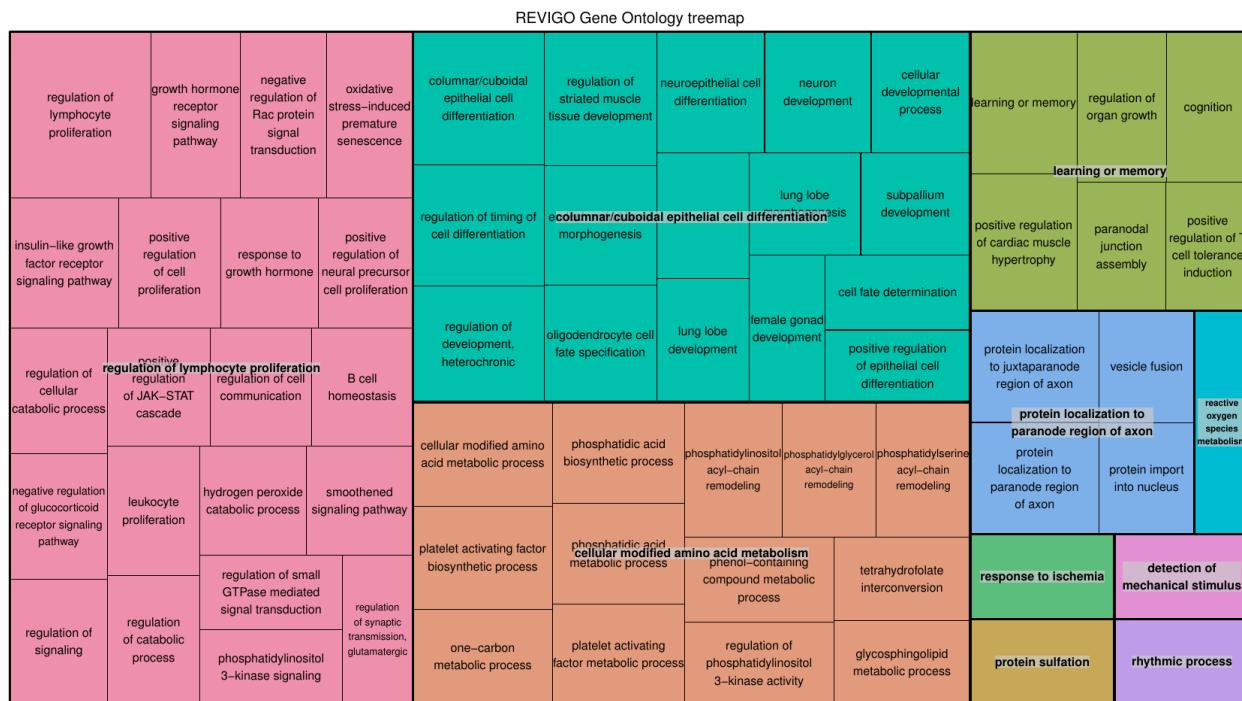
S12 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active HSCs Chromatin Determinant Regions



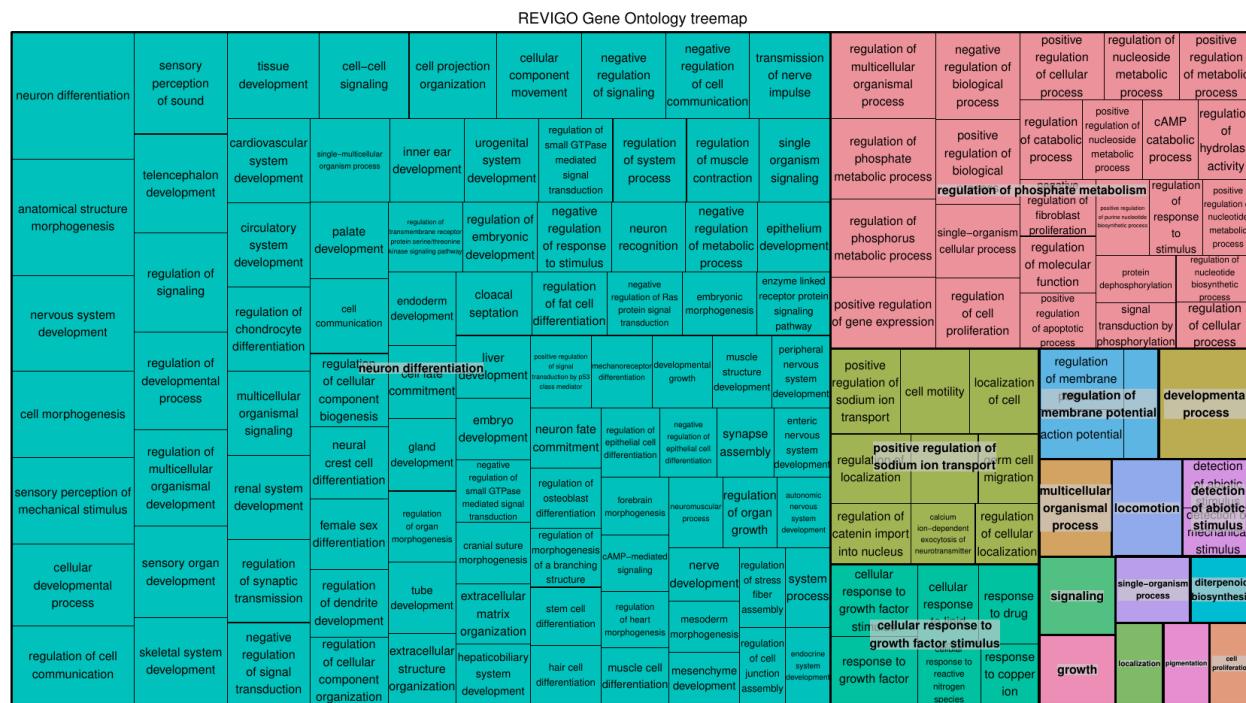
S13 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active T cell Chromatin Determinant Regions



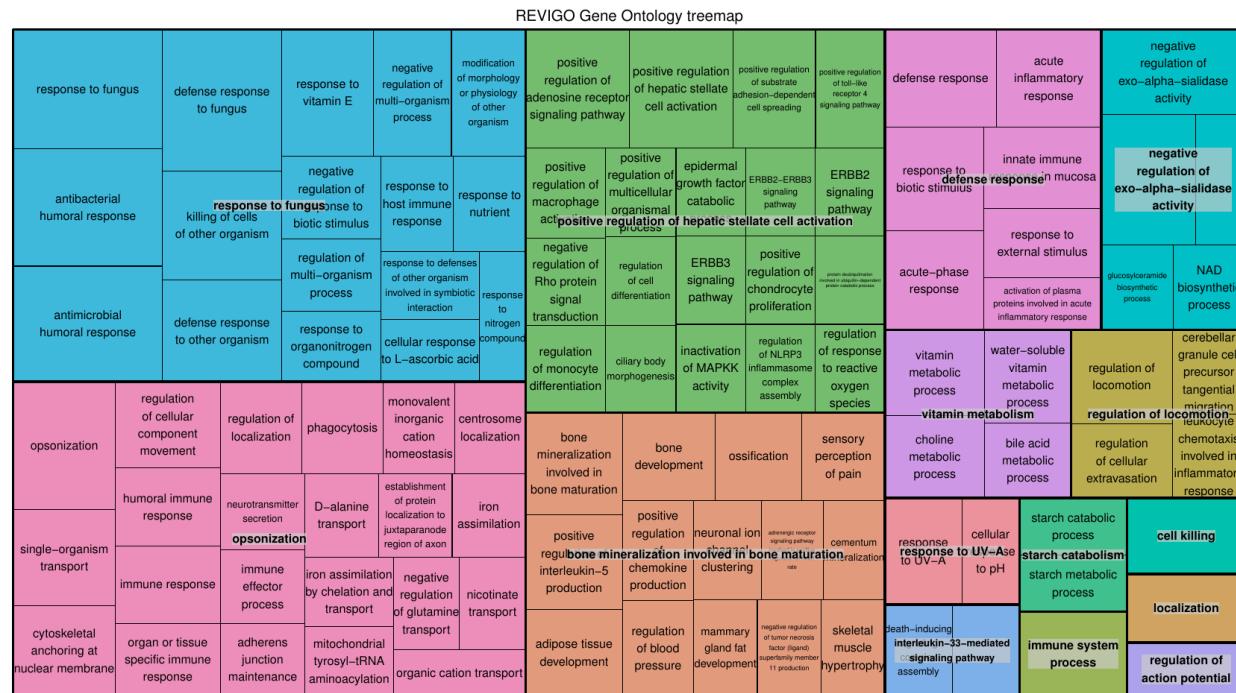
S14 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Naïve B cells Chromatin Determinant Regions



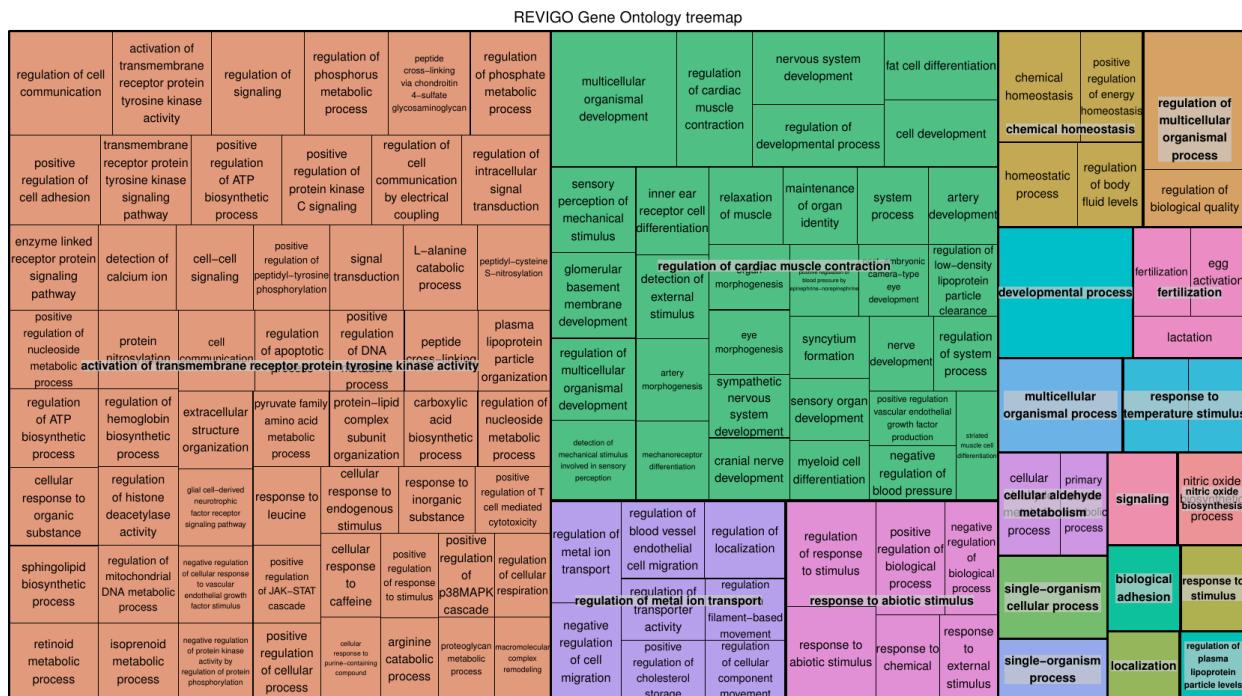
S15 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Germinal Centre B cells Chromatin Determinant Regions



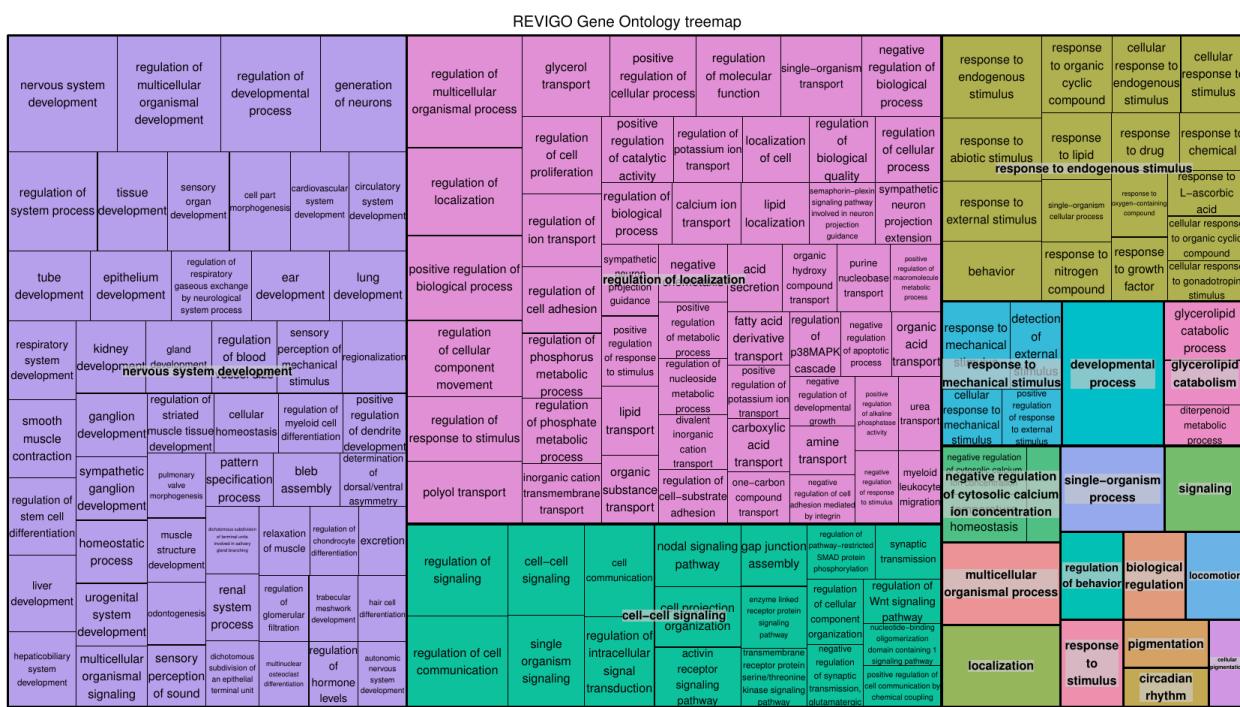
S16 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Neutrophil Chromatin Determinant Regions



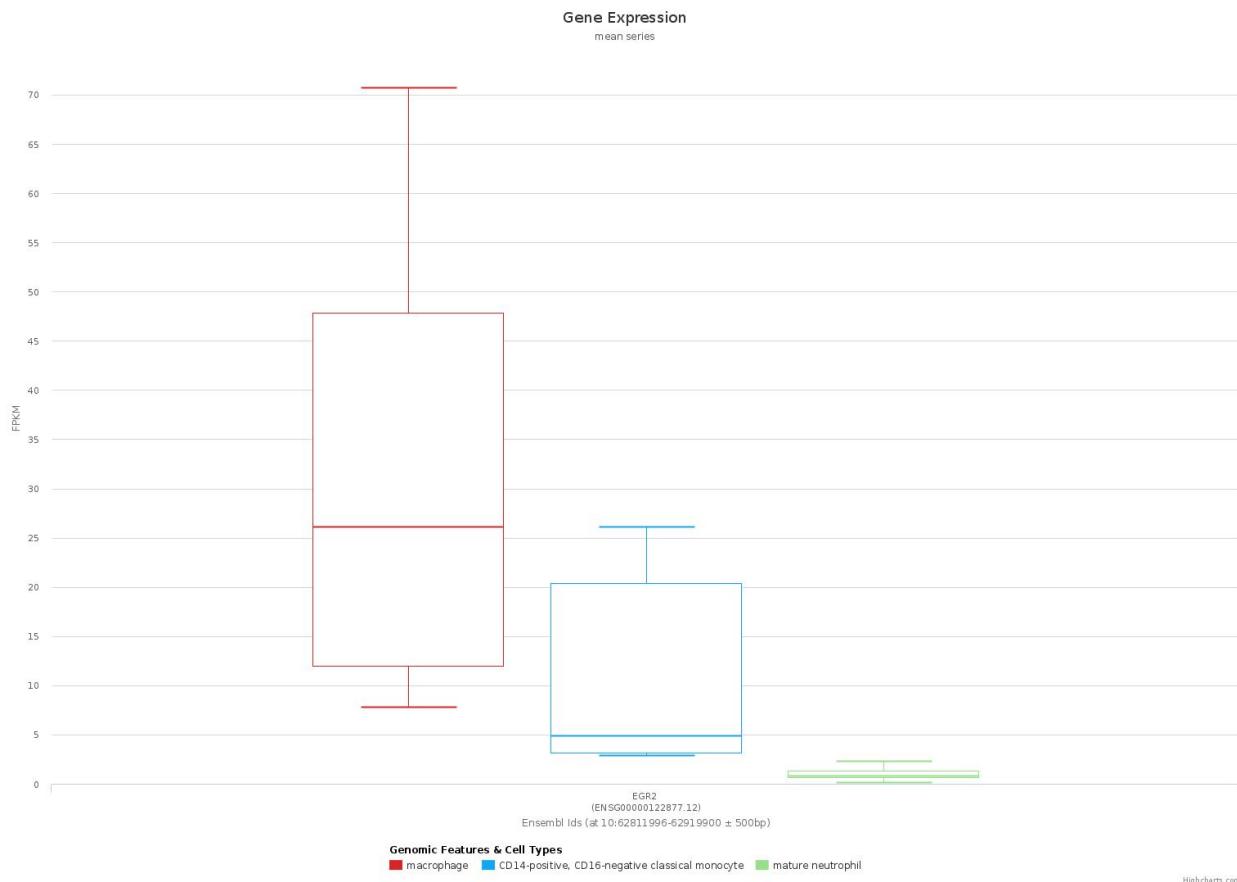
**S17 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Monocyte Chromatin Determinant Regions**



S18 Fig. Related to Fig 3. REVIGO treemap summarizing gene ontology biological process categories over-represented in active Macrophage Chromatin Determinant Regions

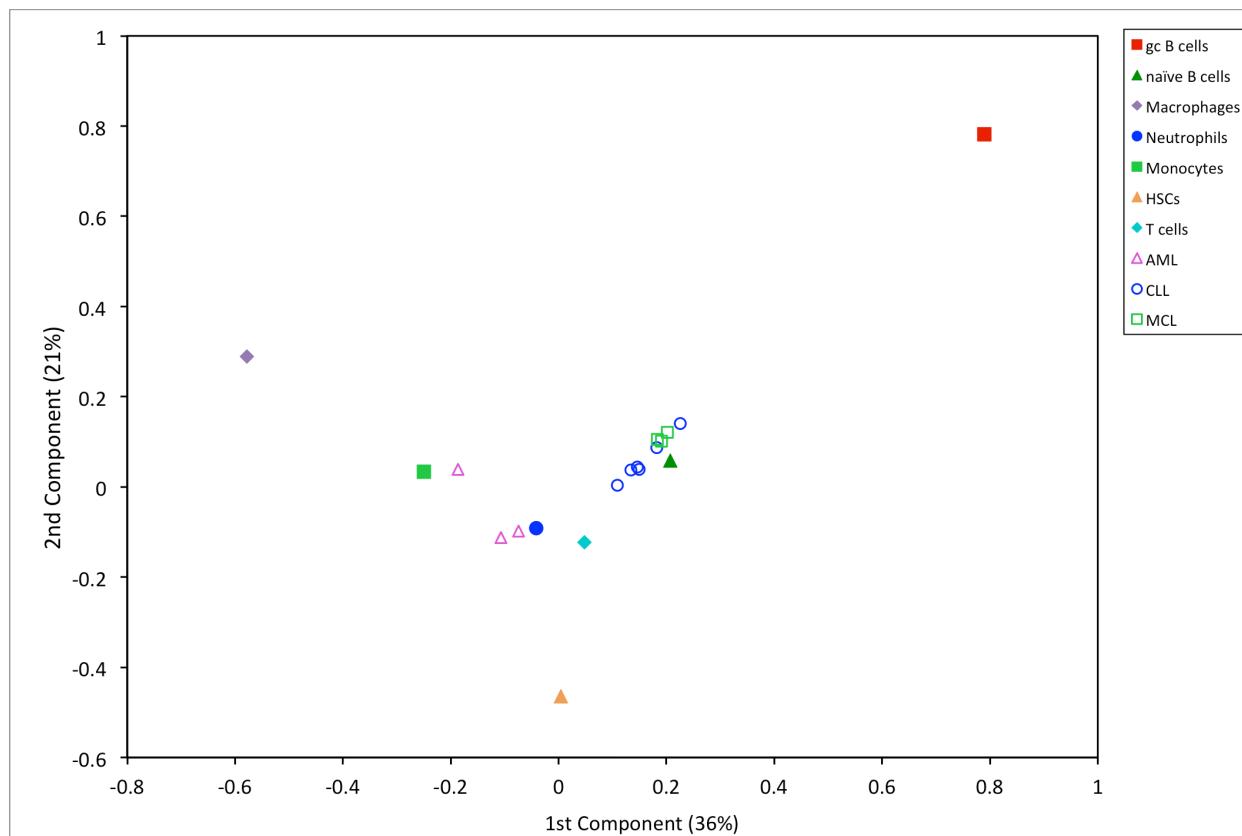


S19 Fig. Related to Fig 3. ERG2 expression in macrophages (red), monocytes (blue) and neutrophils (green) from BLUEPRINT data analysis portal (<http://blueprint-data.bsc.es>)

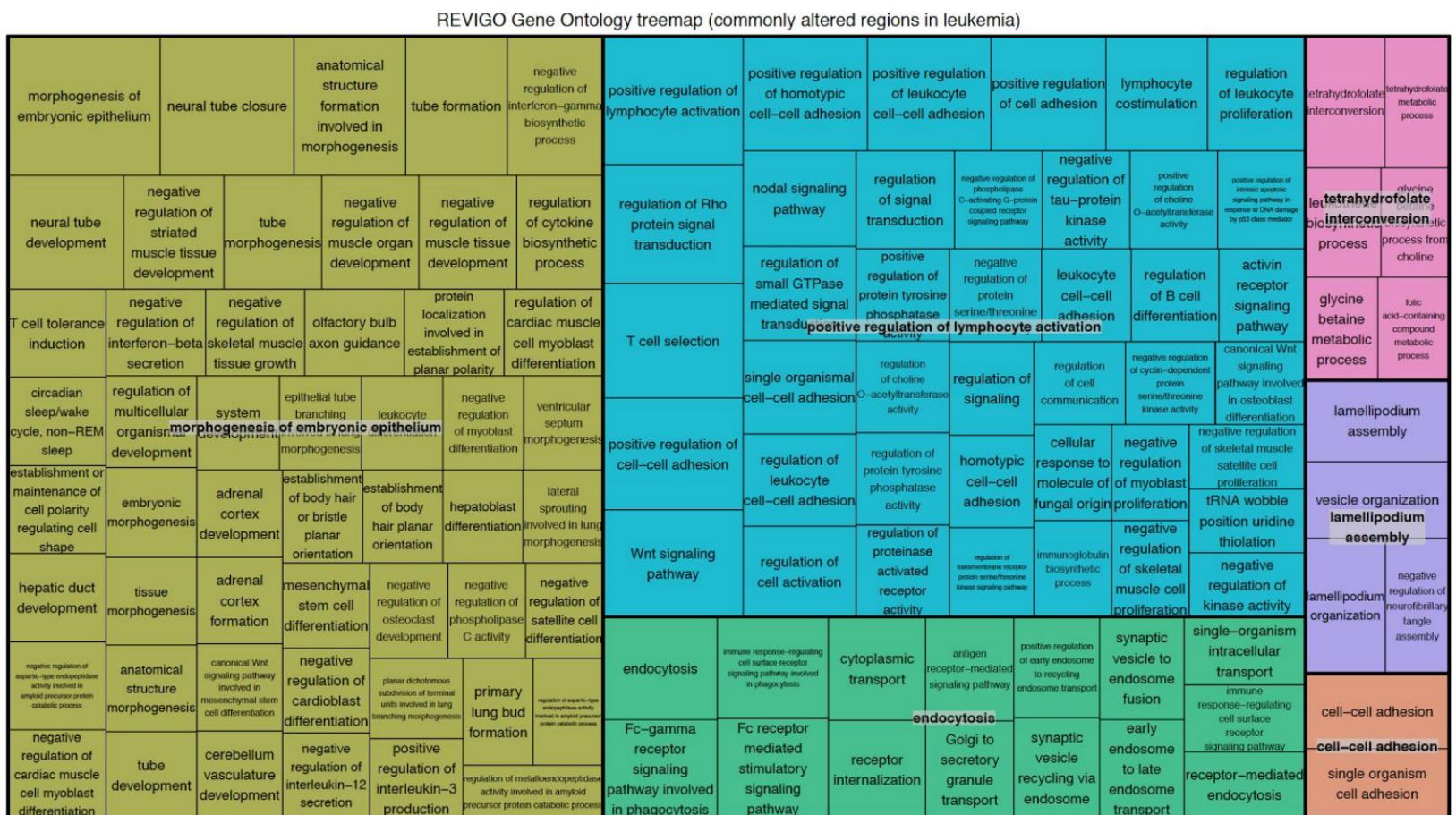


S20 Fig. Related to Fig 4. Epigenomes of leukemias projected into the "Healthy" blood cells chromatin space.

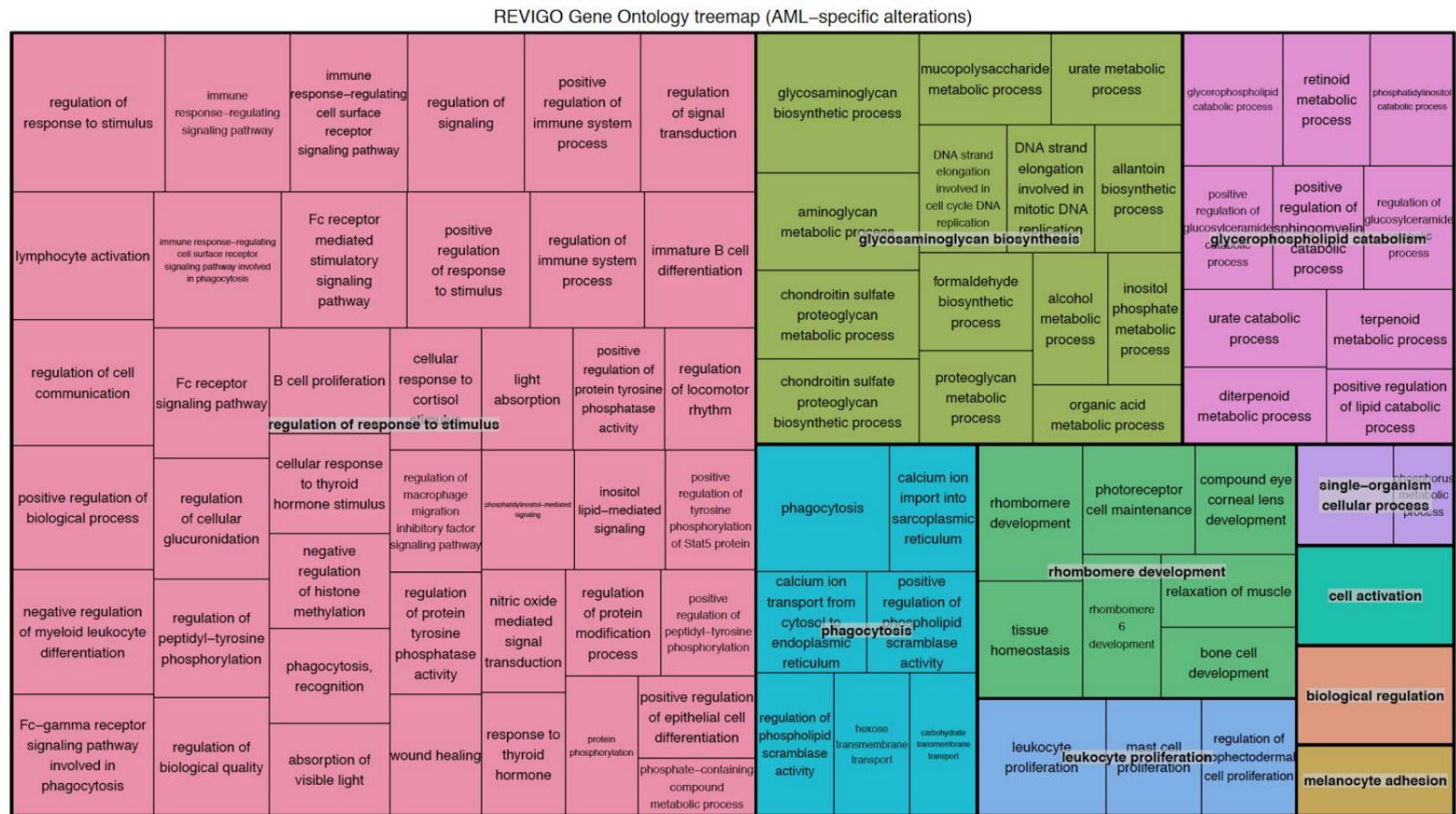
gc B cell: Germinal Centre B cells; AML: Acute Myeloid Leukemia; Leukemia; CLL: Chronic Lymphocytic Leukemia; MCL: Mantle Cell Lymphoma



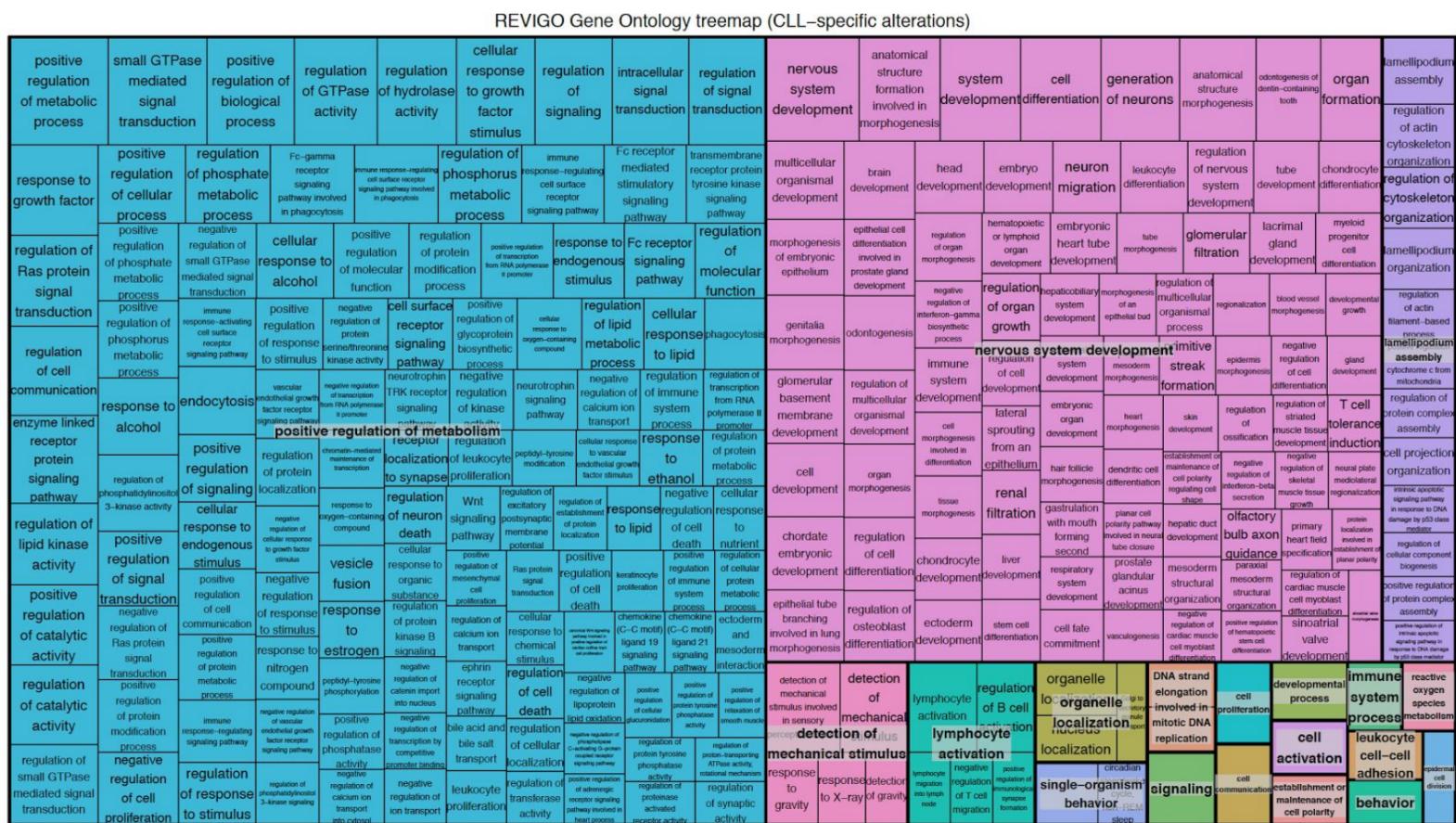
S21 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in commonly altered regions in leukemia



S22 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in AML alterations



## S23 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in CLL alterations



## S24 Fig. Related to Fig 4. REVIGO treemap summarizing gene ontology in MCL alterations

