Super-enhancers are transcriptionally more active and cell-type-specific than stretch enhancers

Aziz Khan[1, 2, *], Anthony Mathelier[1, 3] and Xuegong Zhang[2, 4]

[1] Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, University of Oslo, Oslo, Norway
[2] Key Lab of Bioinformatics/Bioinformatics Division, BNRIST (Beijing National Research Center for Information Science and Technology), Department of Automation, Tsinghua University, Beijing 100084, China
[3] Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, 0372 Oslo, Norway
[4] School of Life Sciences, Tsinghua University, Beijing, 100084, China

* Corresponding author: aziz.khan@ncmm.uio.no

Supplementary data

Additional figures and data

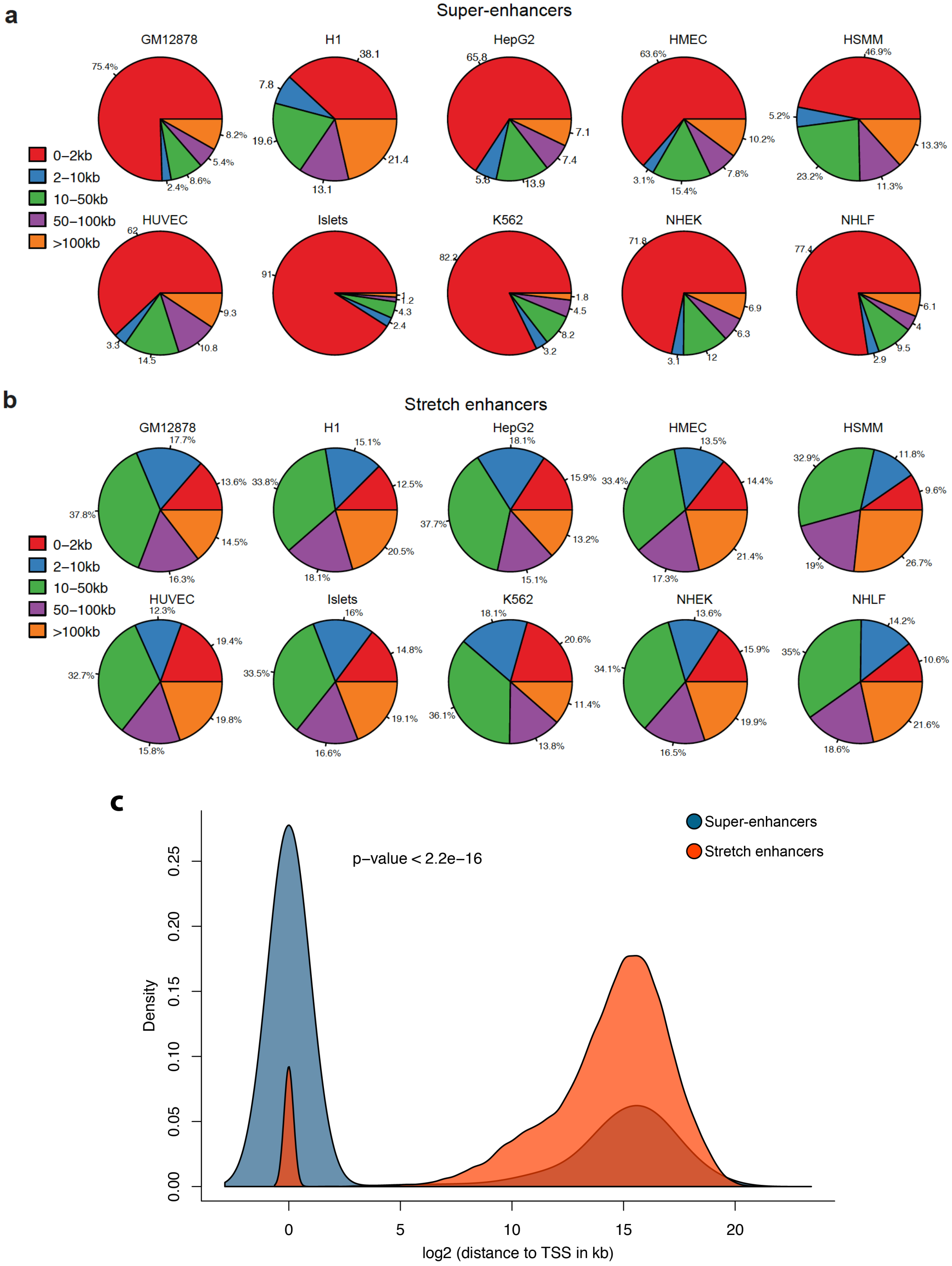

Figure S1: Distribution of distance to TSS for super-enhancers (**a**) and stretch enhancers (**b**) in ten cell types. (**c**) The density plot shows the continuous distribution of log2 of distance to TSS in kb from super-enhancers and stretch enhancers (p-value < 2.2e-16, Wilcoxon signed rank test).

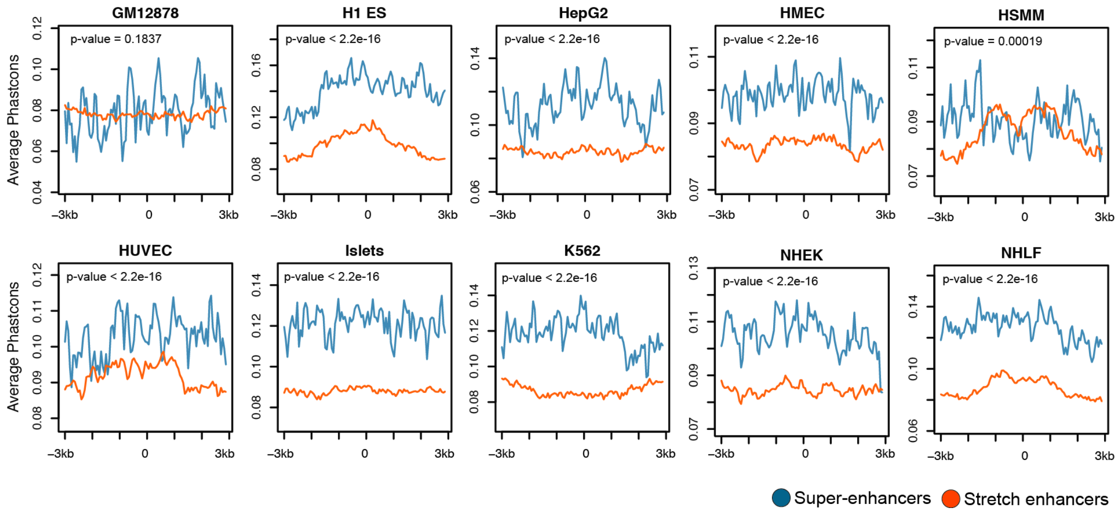

Figure S2: Evolutionary conservation score (phastCons) of super-enhancers and stretch enhancers in 10 cell types. P-value is computed using the Wilcoxon rank sum test.

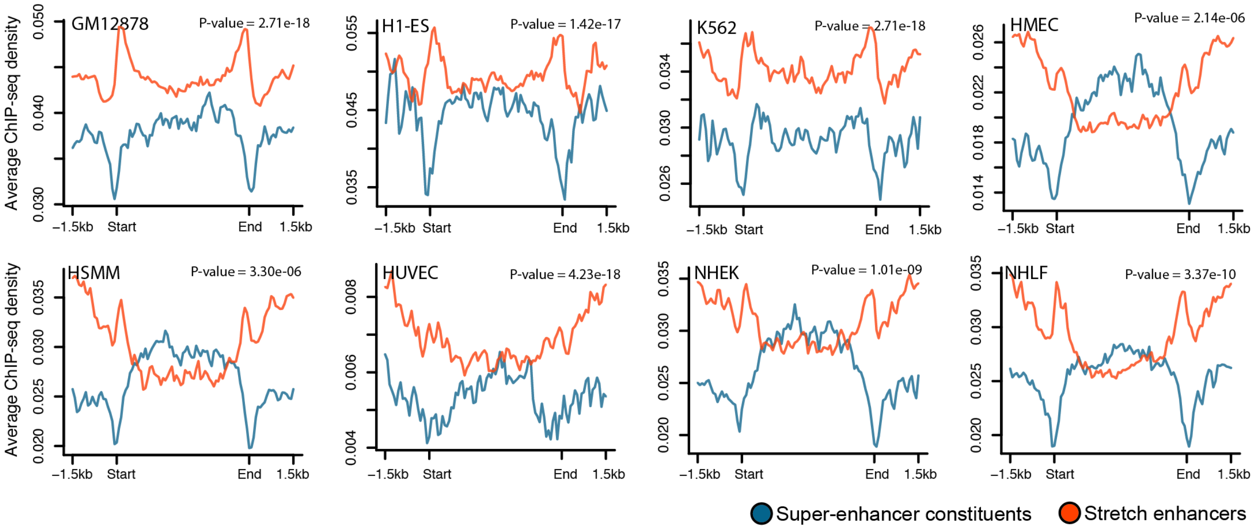

Figure S3. Genome-wide profile of poised mark H3K27me3 at the constituents of super-enhancers and stretch enhancers in H1-ES, K562, GM12878, HMEC, HSMM, HUVEC, NHEK, and NHLF cell-lines. The p-values are computed by using Wilcoxon signed rank test.

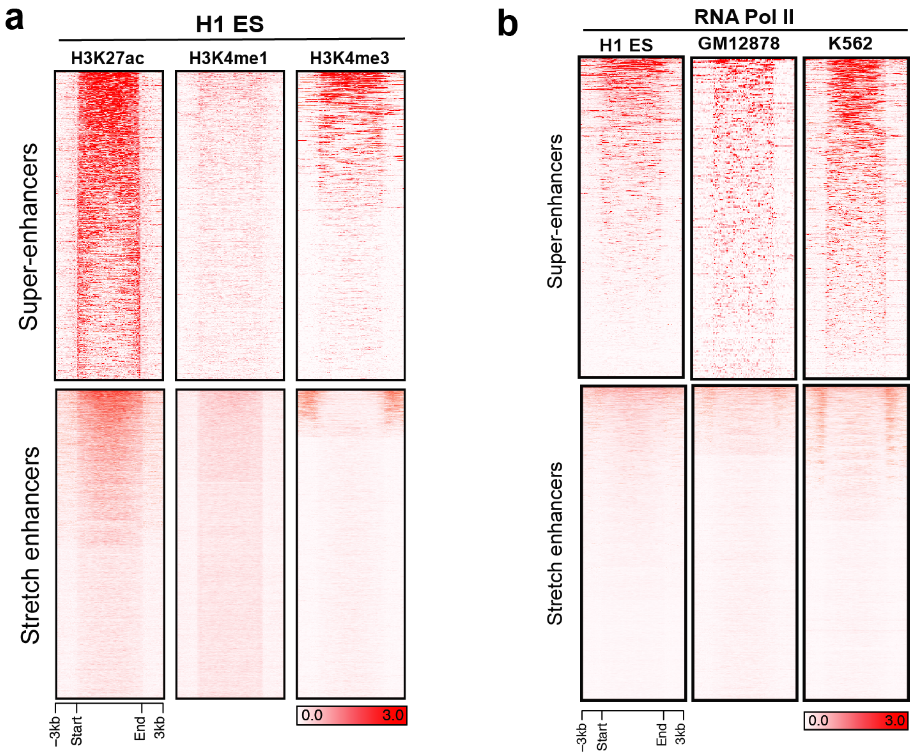

Figure S4. Chromatin architecture and Pol II at super-enhancers and stretch enhancers (a) Heatmap of H3K27ac, H3K4me, H3K4me3 and H3K27me3 at super-enhancers and stretch enhancers in H1-ESC (b) Heatmap of RNA Pol II at super-enhancers and stretch enhancers in H1-ESC, GM12878 and K562.

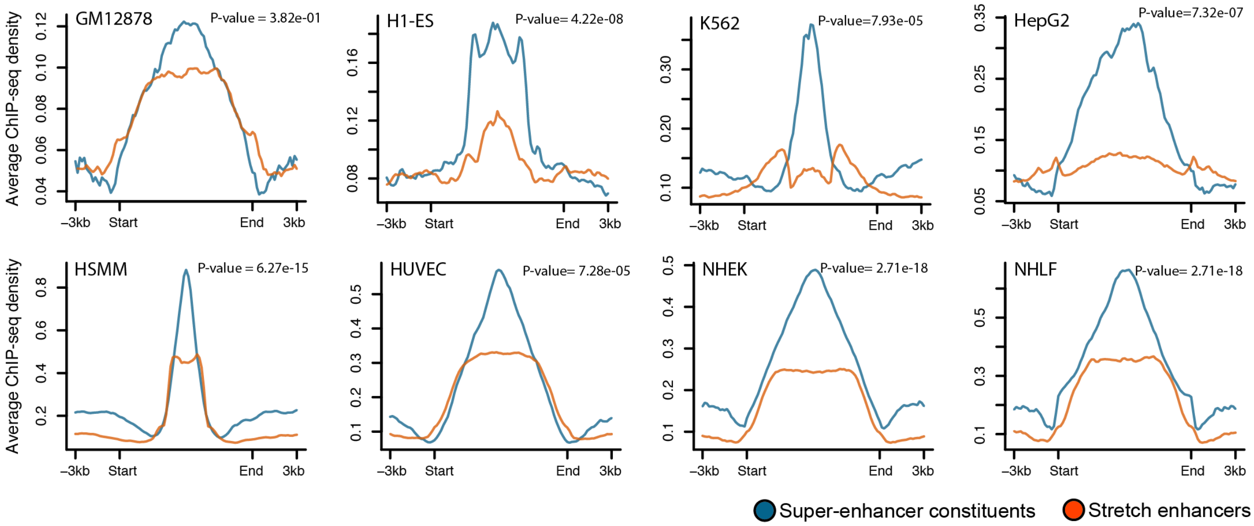

Figure S5: Open chromatin (DNase hypersensitivity) profile at super-enhancers and stretch enhancers in H1-ES, K562, GM12878, HepG2, HUVEC, NHEK, NHLF, and HSSM cell-lines. The p-values are computed by using Wilcoxon paired rank test.

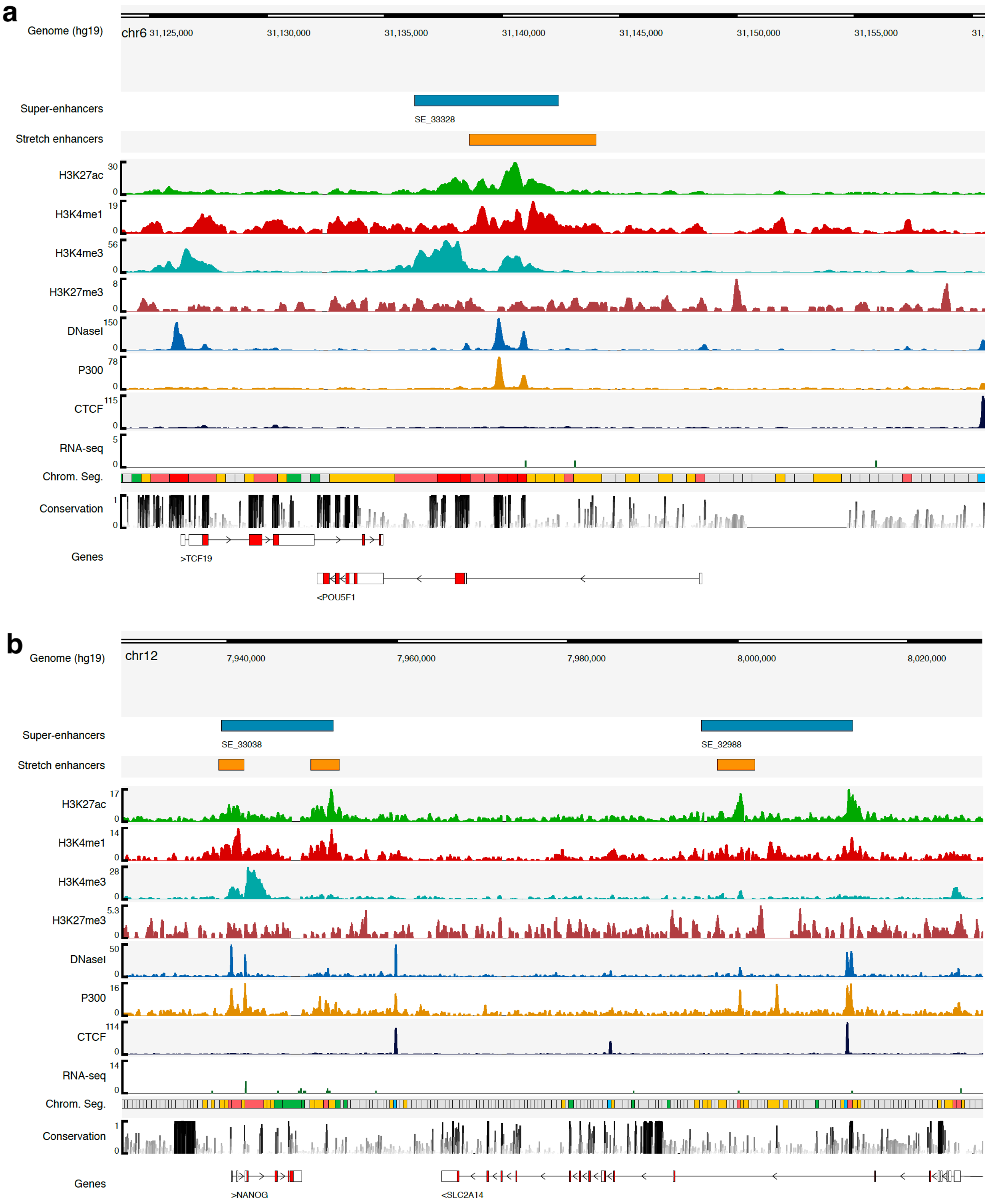

Figure S6: Genomic browser screenshot shows the overlap of super and stretch enhancers with ChIP-seq signal of H3K27ac, H3K4me1, H3K4me3, P300, CTCF, open chromatin (DNaseI), RNA-seq, chromatin segmentation, and conservation at the locus of POU5F/OCT4 (**a**) and NANOG (**b**) genes.

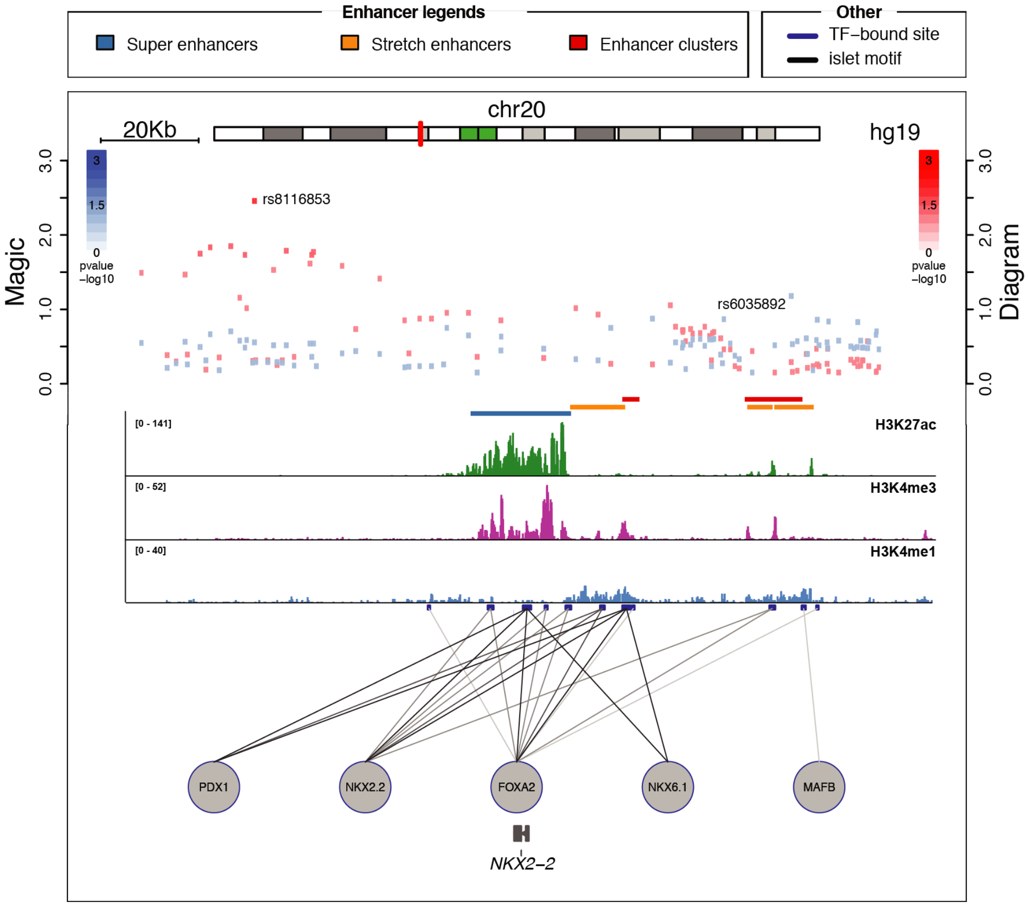

Figure S7: Islet Browser (http://gattaca.imppc.org/isletregulome/) screenshot shows super-enhancers, stretch enhancers and enhancer clusters and binding of Islets specific TFs with ChIP-seq binding for histone modifications H3K27ac, H3K4me3 and H3K4me1.

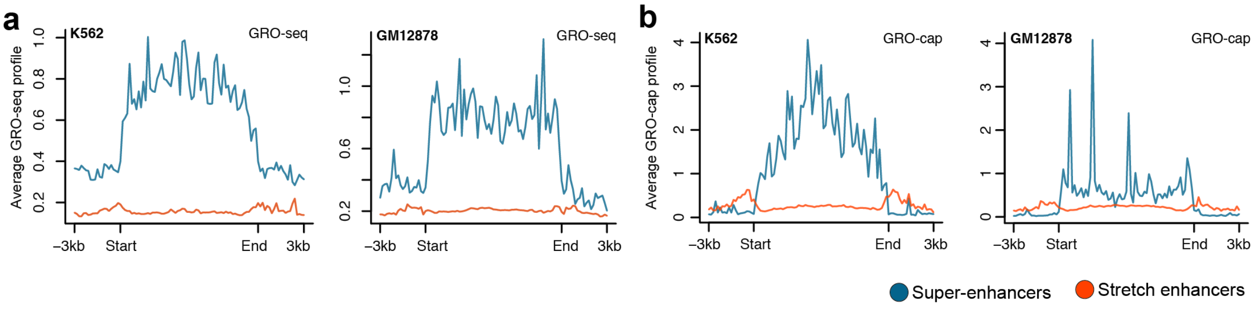

Figure S8: (a) GRO-seq profile at super-enhancers and stretch enhancers in K562 and GM12878 cell-lines. (b) GRO-cap profile at super-enhancers and stretch enhancers in K562 and GM12878 cell-lines.

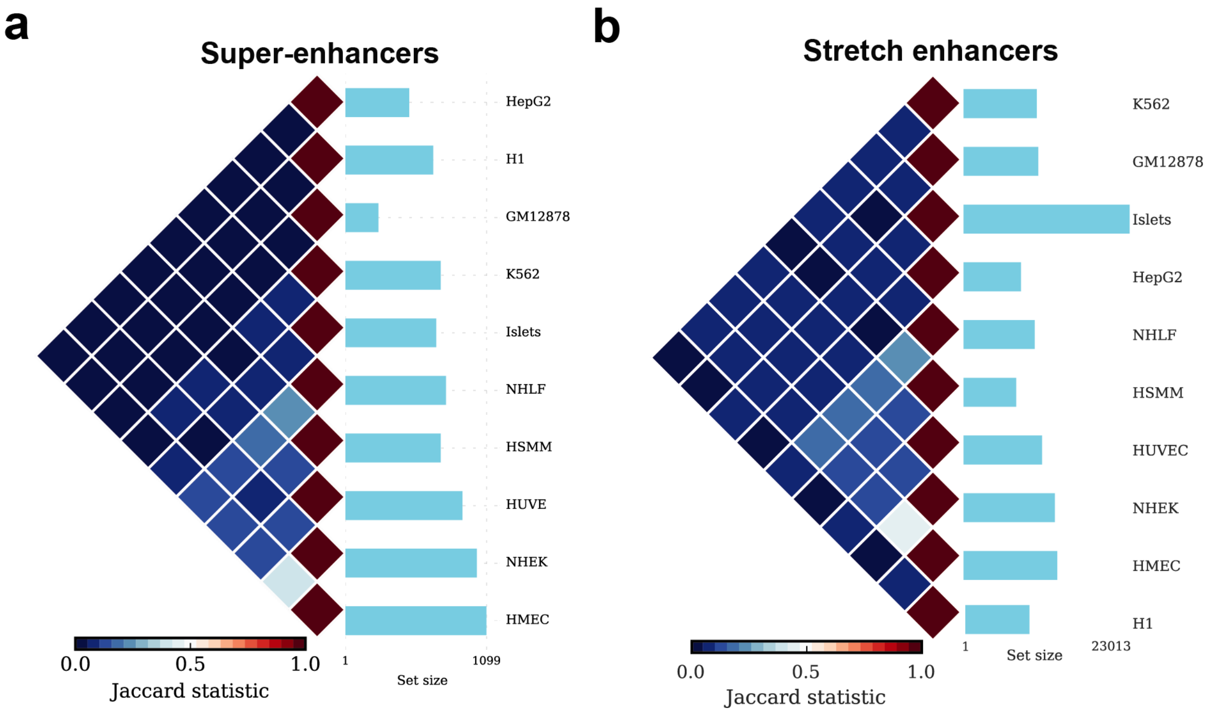

Figure S9: Jaccard's statistics for pairwise overlap of super-enhancers (**a**) and stretch enhancers (**b**) in 10 human cell types. We used pairwise Jaccard score and computed the p-value for pairwise combinations between super and stretch enhancers by using Wilcoxon signed rank test with continuity correction, which shows the difference is statistically significant (p-value = 4.311005e-10).

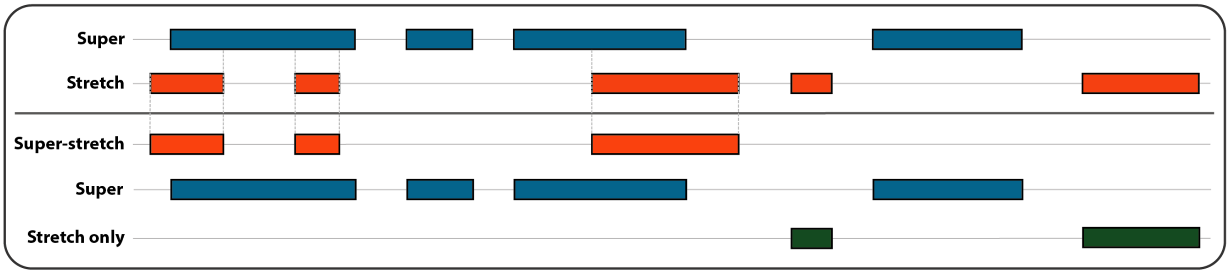

Figure S10: **Three categories of enhancers based on the overlap analysis**. Based on the overlap analysis of super and stretch enhancers, regions are divided in to three categories: (i) super-stretch – all the stretch enhancers which overlap with super-enhancers at least one base pair. (ii) super – all the super-enhancers (since majority overlap with stretch) and (iii) stretch only – those stretch enhancers which does not overlap with super-enhancers (these are referred as stretch in the following text).

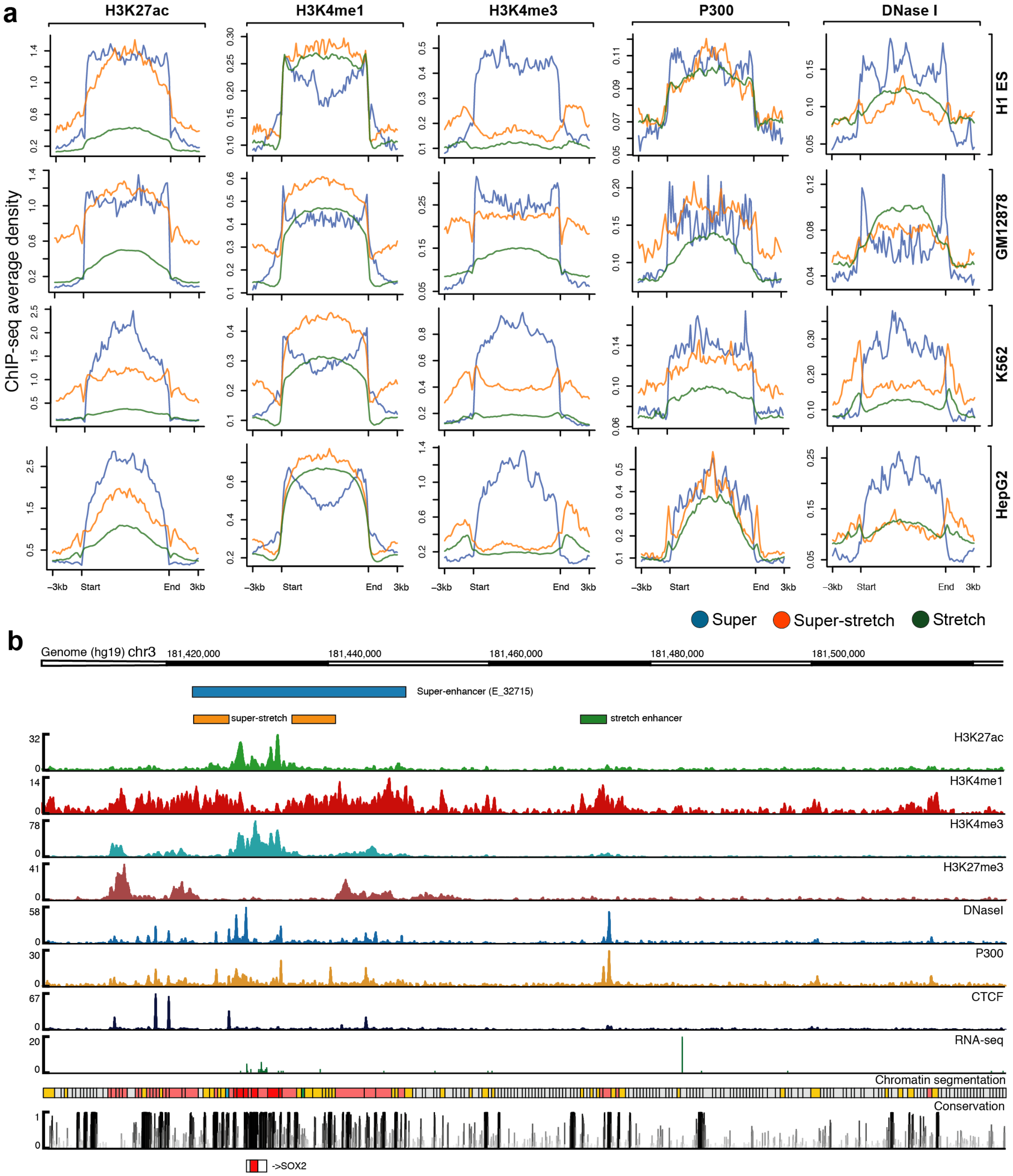

Figure S11. **Chromatin architecture at three groups of enhancers** (**a**) Genome-wide profile of H3K27ac, H3K4me1, H3K4me3, P300 and Dnase I at super, super-stretch and stretch enhancers in H1-ES, GM12878, K562 and HepG2. (**b**) Genomic browser screenshot shows the overlap of super, super-stretch and stretch enhancers with ChIP-seq signal of H3K27ac, H3K4me1, H3K4me3, P300, CTCF, open chromatin (DNaseI), RNA-seq, chromatin segmentation, and conservation at the locus of Sox2 gene.

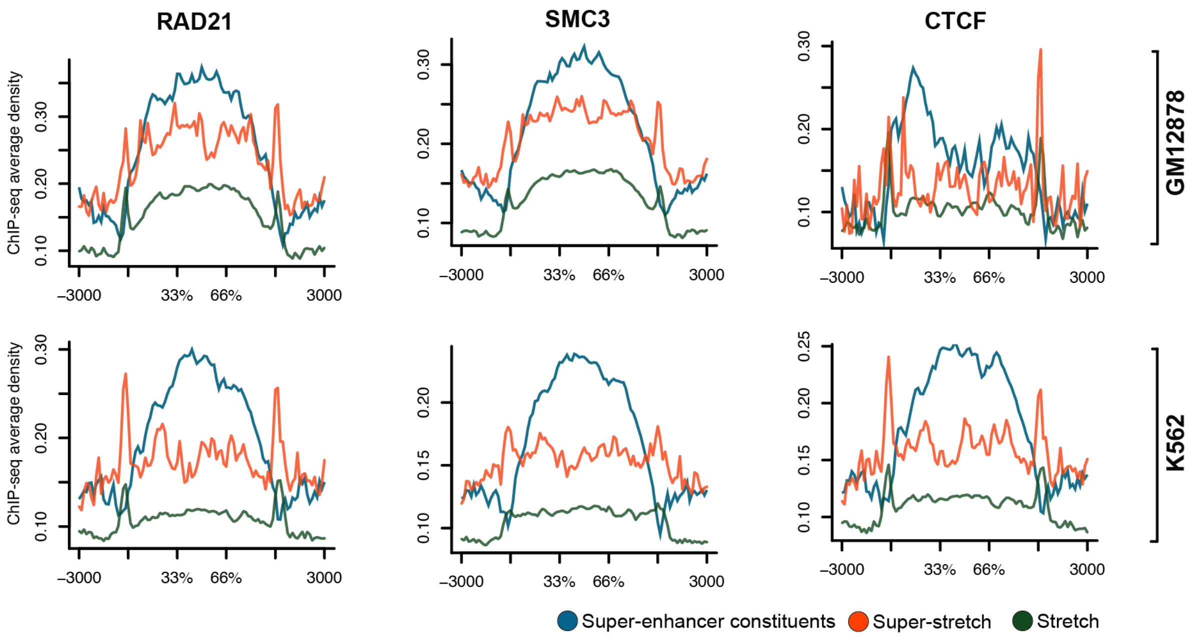

Figure S12: Average profile of Cohesin components (RAD21 and SMC3) and CTCF at the constituents of super-enhancers, super-stretch and stretch enhancers in K562 and GM12878 cell-lines.

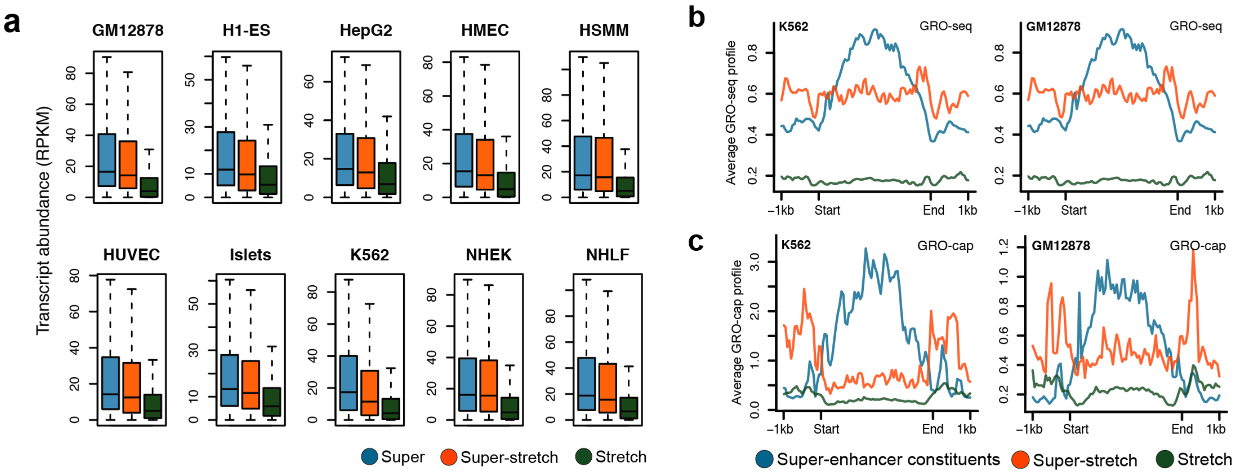

Figure S13: (**a**) Transcriptional abundance (RPKM) of genes associated with super, super-stretch and stretch enhancers across 10 cell types. (**b**) GRO-seq profile at the constituents of super-enhancers, super-stretch and stretch enhancers in K562 and GM12878 cell-lines. (**c**) GRO-cap profile at the constituents of super-enhancers, super-stretch and stretch enhancers in K562 and GM12878 cell-lines.

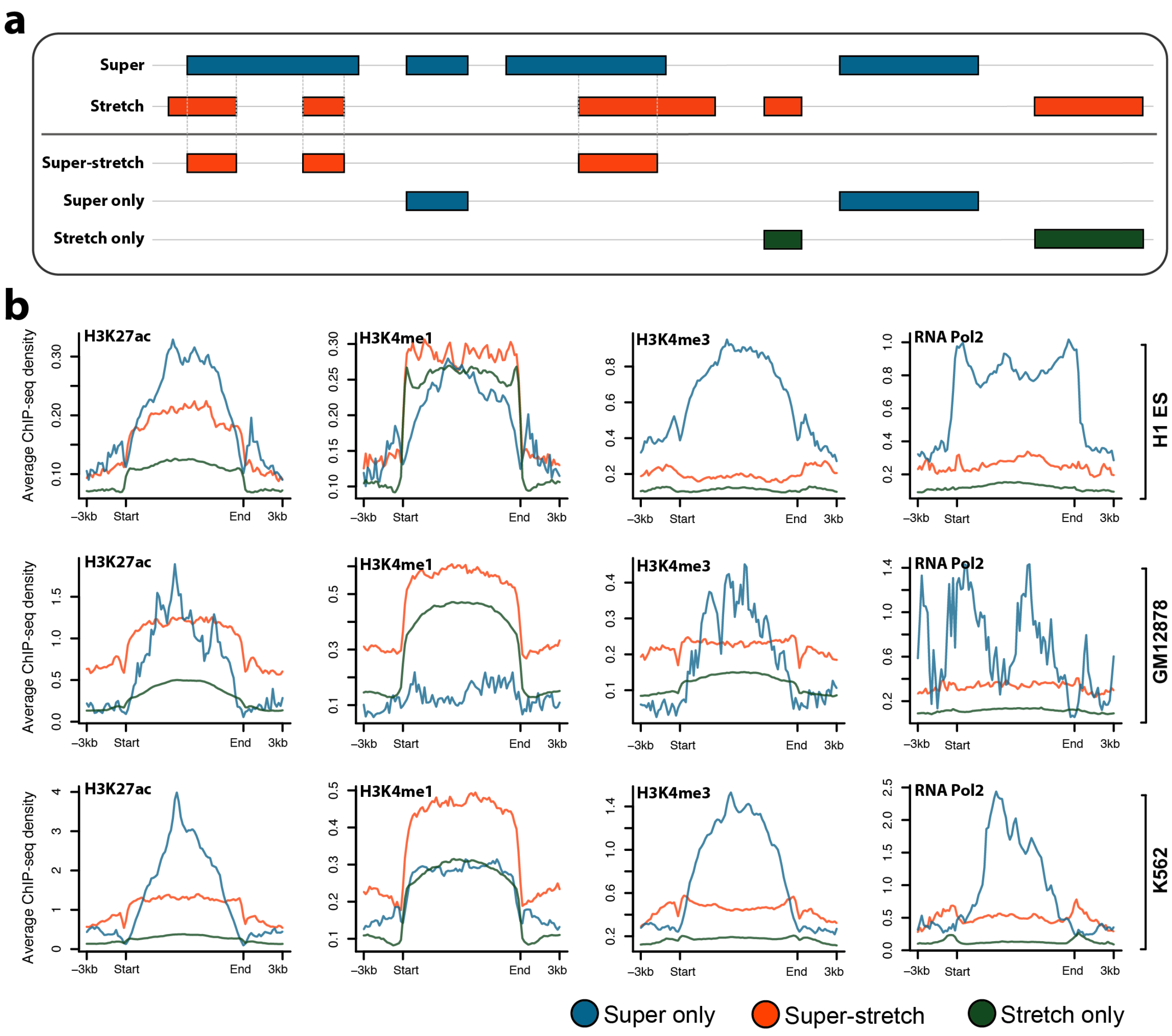

Figure S14: **Chromatin architecture at three categories of enhancers** (**a**) Based on the overlap analysis of super and stretch enhancer are divided into three categories: (i) super only – those super-enhancers which does not overlap with stretch enhancers (we used constituents of super-enhancers for this plot, although we get similar results with the whole super-enhancers) (ii) stretch only – those stretch enhancers which does not overlap with super-enhancers and (iii) super-stretch – there are the exact overlapping regions of super and stretch enhancers. (**b**) Genome-wide profile of H3K27ac, H3K4me1, H3K4me3, and RNA Pol2 at the three categories defined in (**a**) super only (constituents), super-stretch and stretch only enhancers in H1-ES, GM12878, and K562.

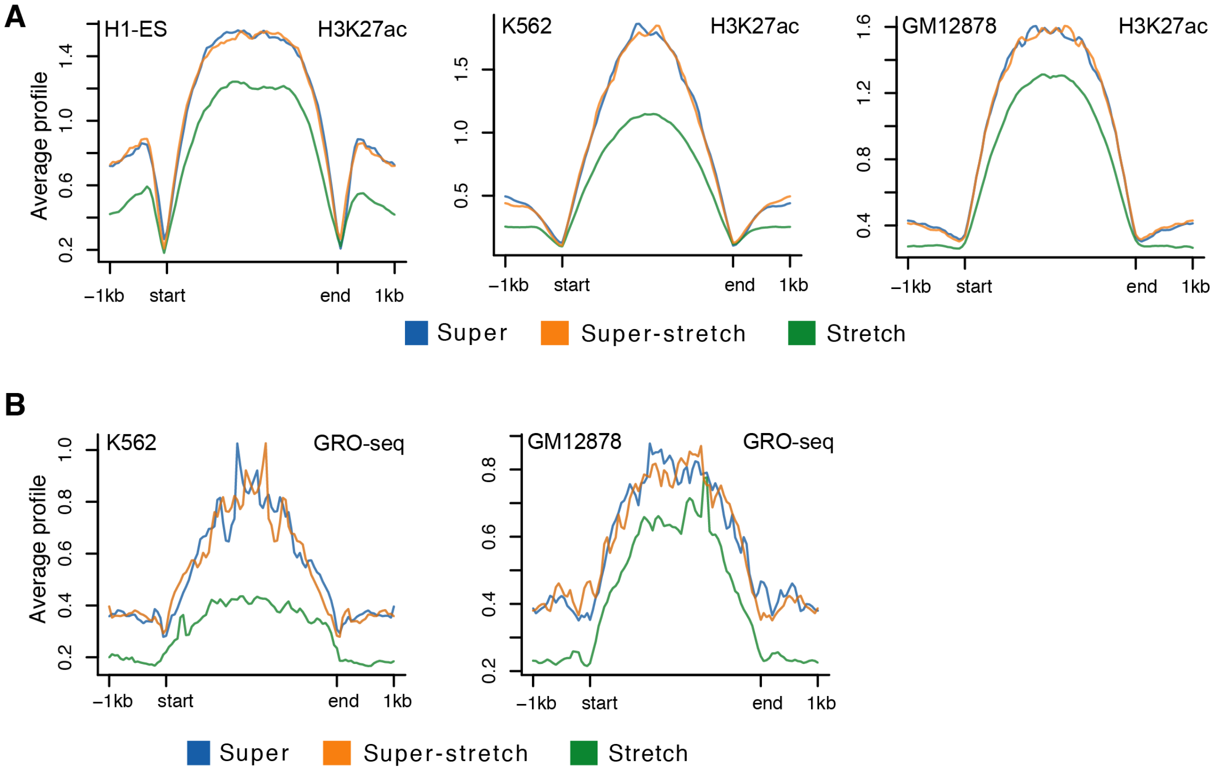

Figure S15: **Redefinition of stretch enhancers using only H3K27ac** (**A**) Genome-wide H3K27ac profile at the constituents of super-enhancers, super-stretch and stretch enhancers defined using only H3K27ac in H1-ES (p-value = 1.13e-06, Wilcoxon rank sum test), K562(p-value = 0.00014, Wilcoxon rank sum test) and GM12878 (p-value = 1.536e-05, Wilcoxon rank sum test) cell-lines. (**B**) Average GRO-seq profile at the constituents of super-enhancers, super-stretch and stretch enhancers defined using only H3K27ac in K562 (p-value = 1.743e-1, Wilcoxon rank sum test) and GM12878 (p-value = 1.009e-08, Wilcoxon rank sum test) cell-lines.

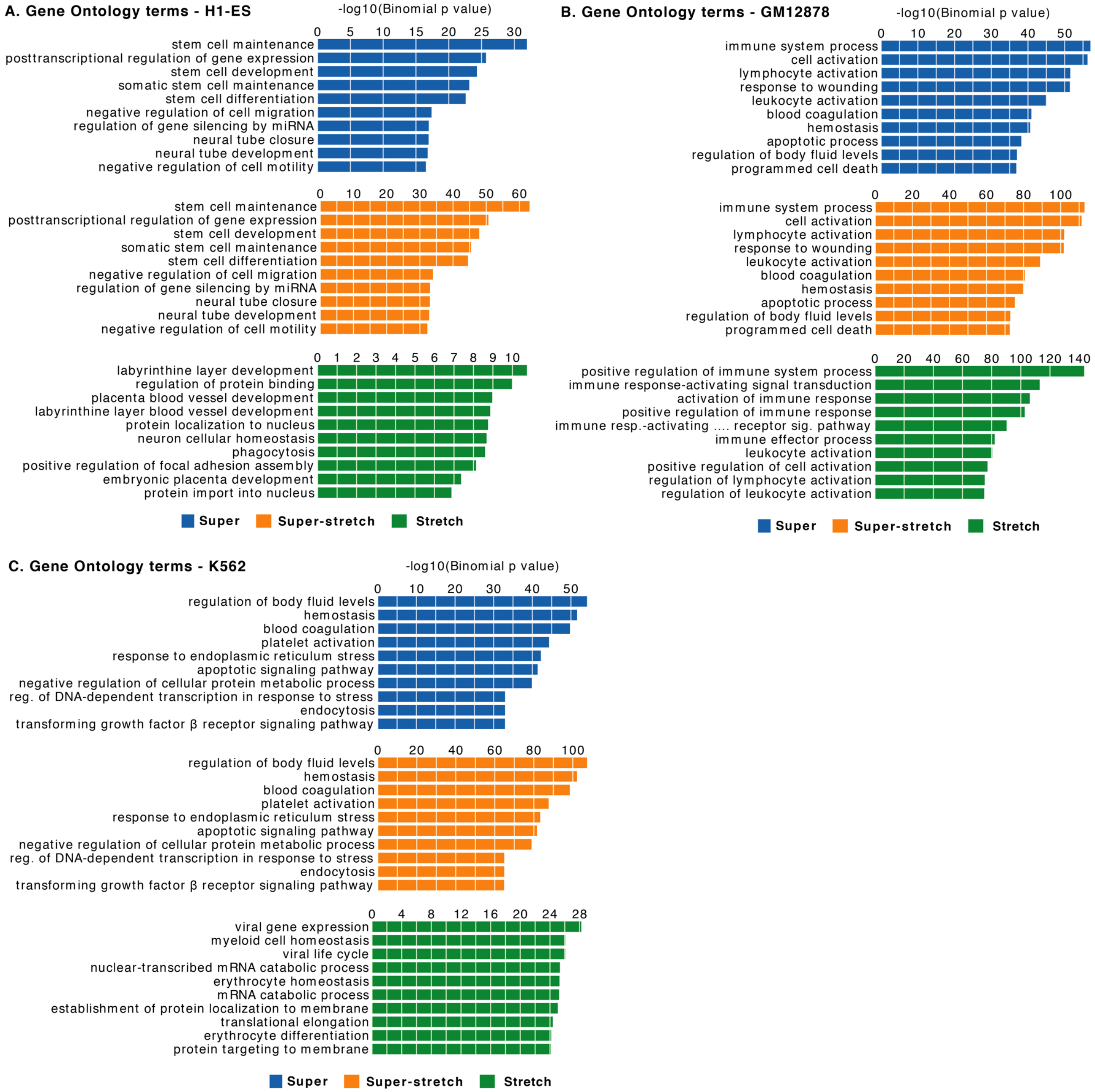

Figure S16: **GO analysis for redefined stretch enhancers using H3K27ac**. Comparison of top 10 GO terms at the constituents of super-enhancers (blue), super-stretch (orange) and stretch enhancers (green) defined using only H3K27ac in three cell types, H1-ES (**A**), GM12878 (**B**) and K562 (**C**).

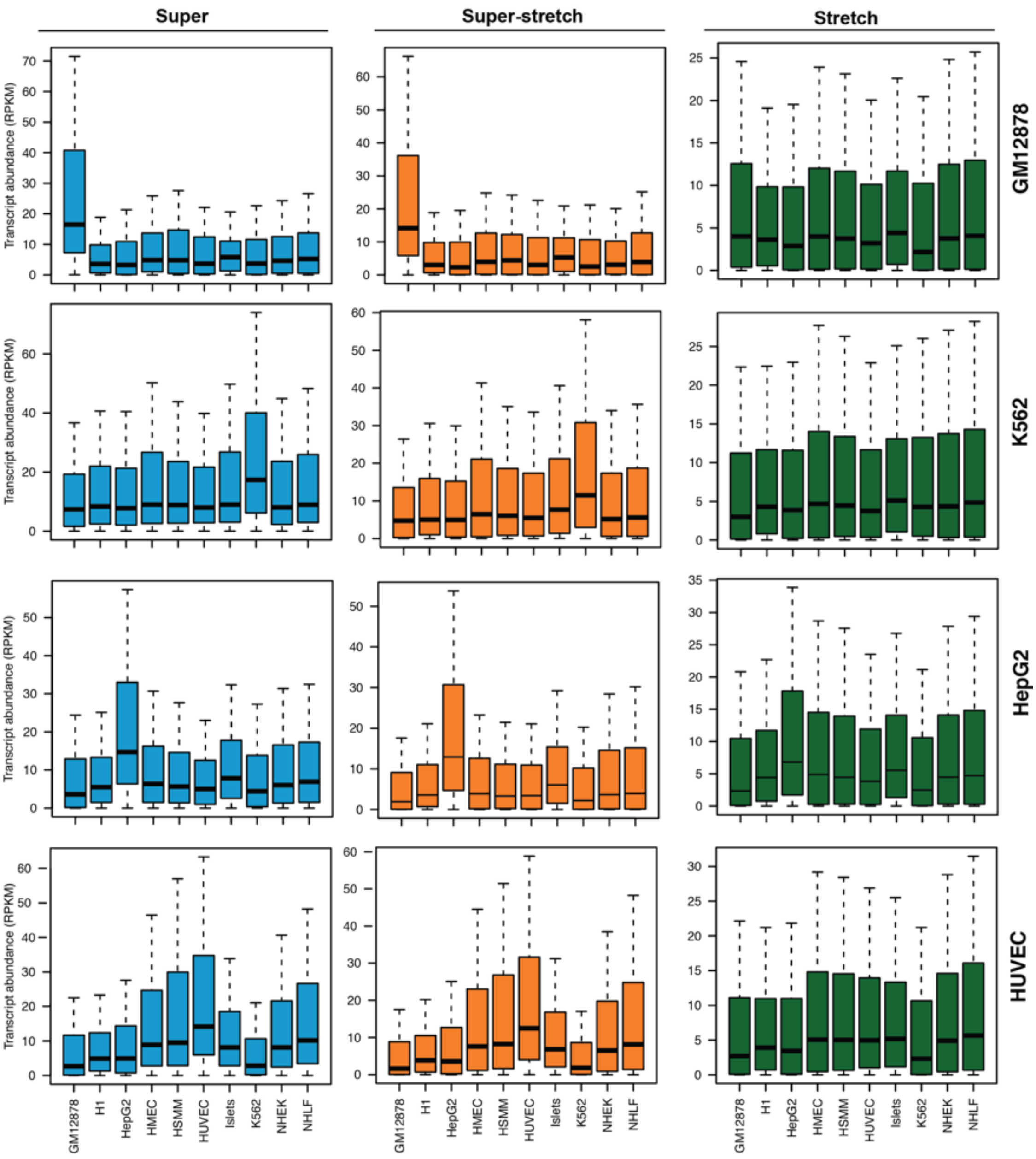

Figure S17: Transcriptional abundance (RPKM) of cell-type-specific genes in GM12878, K562, HepG2 and HUVEC.

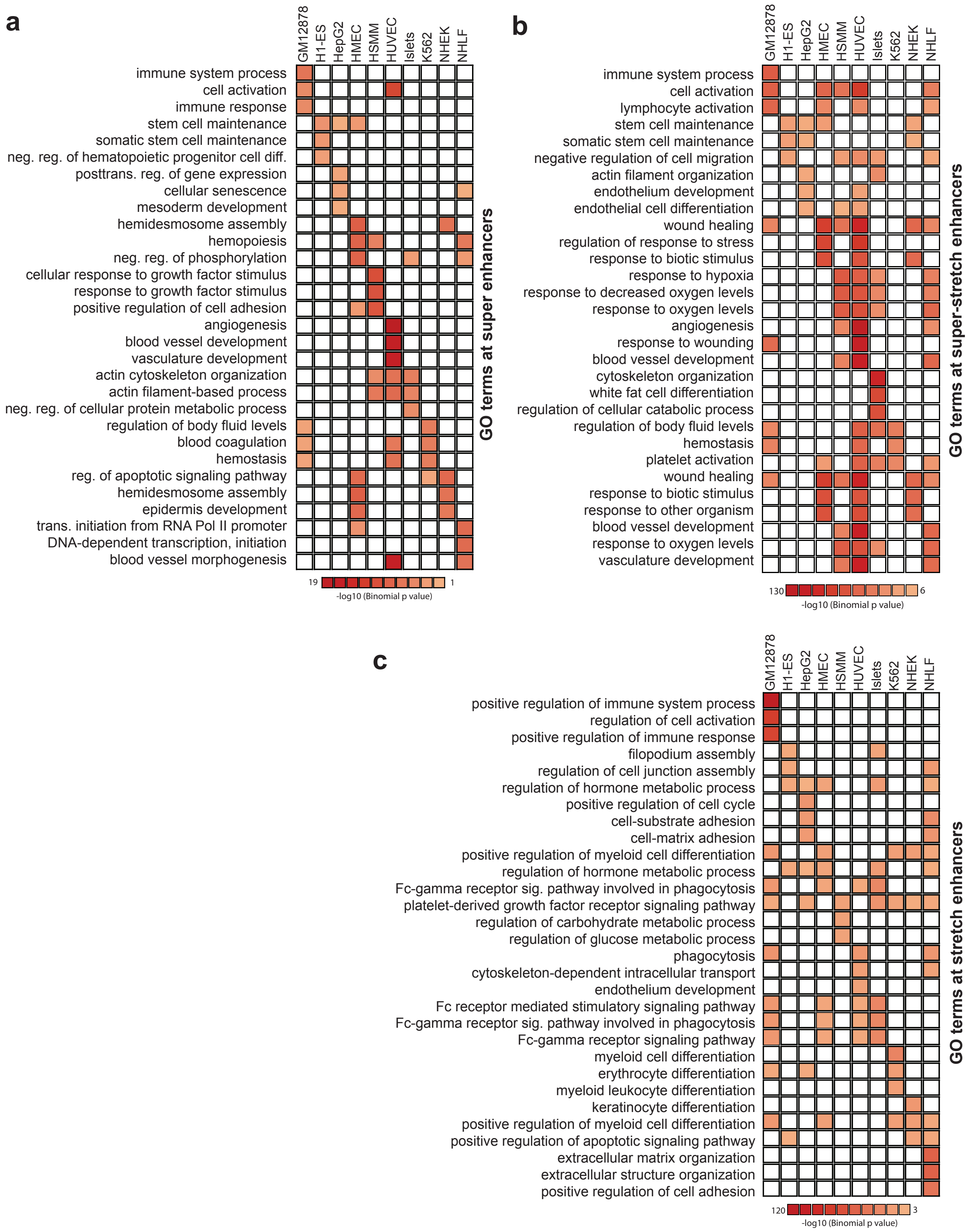

Figure S18: Comparison of top three GO terms from ten cell types in (**a**) super-enhancers, (**b**) super-stretch enhancers and (**c**) stretch enhancers.

Table S1: ChIP-seq and DNAse-seq datasets used in this study.

Cell type	**Factor**	**Data identifiers**
GM12878	Input	GSM733742
GM12878	CTCF	ENCFF000ROM, ENCFF000RON
GM12878	H3K27ac	GSM733771
GM12878	RAD21	ENCFF311CJK, ENCFF800DLO
GM12878	SMC3	ENCFF415OZI, ENCFF797XWM
GM12878	CTCF	wgEncodeBroadHistoneGm12878CtcfStdAln
GM12878	H3K27ac	wgEncodeBroadHistoneGm12878H3k27acStdAln
GM12878	H3K27me3	wgEncodeBroadHistoneGm12878H3k27me3StdAln
GM12878	H3K4me1	wgEncodeBroadHistoneGm12878H3k4me1StdAln
GM12878	H3K4me3	wgEncodeBroadHistoneGm12878H3k4me3StdAln
GM12878	P300	wgEncodeHaibTfbsGm12878P300Pcr1xAln
GM12878	Dnase	wgEncodeOpenChromDnaseGm12878Aln
GM12878	P300	wgEncodeSydhTfbsGm12878P300bStdAln
GM12878	Pol2	wgEncodeSydhTfbsGm12878Pol2StdAln
H1-hESC	H3K27me3	wgEncodeBroadHistoneH1hescH3k27me3StdAln
H1-hESC	H3K4me1	wgEncodeBroadHistoneH1hescH3k4me1StdAln
H1-hESC	H3K4me3	wgEncodeBroadHistoneH1hescH3k4me3StdAln
H1-hESC	P300	wgEncodeBroadHistoneH1hescP300kat3bAln
H1-hESC	Pol2	wgEncodeHaibTfbsH1hescPol2V0416102Aln
H1-hESC	Dnase	wgEncodeOpenChromDnaseH1hescAln
H1-hESC	Dnase	wgEncodeOpenChromDnaseH1hescAln
H1-hESC	Input	SRR067970
HepG2	H3K4me1	wgEncodeBroadHistoneHepg2H3k04me1StdAln
HepG2	H3K27ac	wgEncodeBroadHistoneHepg2H3k27acStdAln
HepG2	H3K4me3	wgEncodeBroadHistoneHepg2H3k4me3StdAln
HepG2	P300	wgEncodeHaibTfbsHepg2P300V0416101Aln
HepG2	Dnase	wgEncodeOpenChromDnaseHepg2Aln
HUVEC	H3K27ac	wgEncodeBroadHistoneHuvecH3k27acStdAln
HUVEC	H3K4me1	wgEncodeBroadHistoneHuvecH3k4me1StdAln
HUVEC	H3K4me3	wgEncodeBroadHistoneHuvecH3k4me3StdAln

HUVEC	Dnase	wgEncodeOpenChromDnaseHuvecAln
HUVEC	Pol2	wgEncodeSydhTfbsHuvecPol2StdAln
HUVEC	Pol2	wgEncodeSydhTfbsHuvecPol2StdAln
K562	Input	GSM733780
K562	CTCF	ENCFF487UYG, ENCFF496SZR
K562	RAD21	ENCFF084HTD, ENCFF330BAK
K562	SMC3	ENCFF359UWD, ENCFF845WVY
K562	H3K27ac	wgEncodeBroadHistoneK562H3k27acStdAln
K562	H3K27me3	wgEncodeBroadHistoneK562H3k27me3StdAln
K562	H3K4me1	wgEncodeBroadHistoneK562H3k4me1StdAln
K562	H3K4me3	wgEncodeBroadHistoneK562H3k4me3StdAln
K562	P300	wgEncodeBroadHistoneK562P300StdAln
K562	Dnase	wgEncodeOpenChromDnaseK562Aln
K562	Pol2	wgEncodeSydhTfbsK562Pol2StdAln
NHEK	Dnase	wgEncodeOpenChromDnaseNhekAln
NHLF	Dnase	wgEncodeUwDnaseNhlfAln
NHEK	Dnase	ENCFF564HCA
HUVEC	Dnase	ENCFF757PTA
HSMM	Dnase	ENCFF192DFJ
NHLF	Dnase	ENCFF937ITN
HSMM	Dnase	wgEncodeOpenChromDnaseHsmmAln
HepG2	H3K27me3	wgEncodeBroadHistoneHepg2H3k27me3StdAln
HepG2	H3K27me3	ENCFF730WFM
HUVEC	H3K27me3	wgEncodeBroadHistoneHuvecH3k27me3StdAln
NHEK	H3K27me3	wgEncodeBroadHistoneNhekH3k27me3StdAln
HSMM	H3K27me3	wgEncodeBroadHistoneHsmmH3k27me3StdAln
NHLF	H3K27me3	wgEncodeBroadHistoneNhlfH3k27me3StdAln
NHEK	H3K27me3	ENCFF285QUM
HMEC	H3K27me3	wgEncodeBroadHistoneHmecH3k27me3StdAln

Table S2: GRO-seq and GRO-cap datasets used in this study.

Cell type	Data type	GEO ID
K562	GRO-seq	GSM1480325
	GRO-cap	GSM1480321
GM12878	GRO-seq	GSM1480326
	GRO-cap	GSM1480323