SUPPLEMENTAL INFORMATION

Transcriptional super-enhancers connected to cell identity and disease

Denes Hnisz, Brian J. Abraham, Tong Ihn Lee, Ashley Lau, Violaine Saint-André, Alla A. Sigova, Richard A. Young

CONTENTS

References for the validation of candidate master transcription factors
Gene Ontology (GO) Analysis
Trait-associated SNPs
Radar plots
Selection of oncogenes
Cancer hallmark analysis
RNA-Seq
Gene expression analysis

Supplemental References

Supplemental Figure Legends

Supplemental Figure 1. Related to Figure 1.
(A) Metagene representations of mean ChIP-Seq density in regions surrounding constituents of super-enhancers and typical enhancers, active promoters, and the borders of topological domains for the indicated transcription factors in mESCs. For each transcription factor, the leftmost panel shows background-subtracted densities within size-normalized constituents of super-enhancers (red) and typical enhancers (grey) plus adjacent 2.5kb regions. The center panel shows background-subtracted densities within 1kb of RefSeq-defined transcription start sites of active transcripts. Active transcripts were defined as having an RNA Polymerase II mean ChIP-Seq signal >1 rpm/bpin this region. The right panel shows background-subtracted densities within +/-500kb of borders of topological domains in mESC as defined in (Dixon et al., 2012).
(B) Table depicting transcription factor binding motifs at constituent enhancers within typical enhancer regions, and associated p-values.

Supplemental Figure 2. Related to Figure 2.
(A) *(top)* Schematic diagram the shRNAs knockdown of Oct4 in mESCs. *(bottom)* Box plots of fold change expression for all enhancer-associated genes, typical enhancer-associated genes and super-enhancer –associated genes 3, 4 and 5 days after knockdown. Expression values were normalized to the values measured in ESCs transduced with a control shRNA against GFP.
(B) *(top)* Schematic diagram the shRNAs knockdown of the Mediator subunit Med12 in mESCs. *(bottom)* Box plots of fold change expression for all enhancer-associated genes, typical enhancer-associated genes and super-enhancer –associated genes 3, 4 and 5 days after knockdown. Expression values were normalized to the values measured in ESCs transduced with a control shRNA against GFP.
(C) *(top)* Schematic diagram the shRNAs knockdown of the cohesin subunit Smc1 in mESCs. *(bottom)* Box plots of fold change expression for all enhancer-associated genes, typical enhancer-associated genes and super-enhancer –associated genes 3, 4 and 5 days after knockdown. Expression values were normalized to the values measured in ESCs transduced with a control shRNA against GFP.

Supplemental Figure 3. Related to Figure 3.
(A) Comparison of the abilities of enhancer surrogate marks (p300, H3K27ac, H3K4me1, DNase hypersensitivity) to identify super-enhancers and super-enhancer –associated genes in mESCs. *(top)* ChIP-Seq binding profiles for OSN (merged binding profiles of the transcription factors Oct4, Sox2 and Nanog), Mediator (Med1), p300, H3K27ac, H3K4me1 and DNase hypersensitivity at the *POLE* and *miR-290-295* loci in mESCs. Red dots indicate the median enrichment of all bound regions in the respective ChIP-Seq datasets, and are positioned at maximum 20% of the axis height. (rpm/bp: reads per million per base pair) *(middle left)* Venn diagrams showing the overlap between super-enhancer domains identified by Oct4, Sox2, Nanog and Med1 data versus those identified using p300, H3K27ac, H3K4me1 ChIP-Seq data or DNase hypersensitivity data. *(middle right)* Venn diagrams showing the overlap between super-enhancer—associated genes identified by Oct4, Sox2, Nanog and Med1 data versus those identified by p300, H3K27ac or H3K4me1, ChIP-Seq data or DNase hypersensitivity data. We note that even though p300 appears to be an excellent enhancer surrogate to identify super-enhancers, p300 ChIP-Seq data for a large set of human samples are as yet not available. Super-enhancers identified by H3K27ac but not by OSN-Mediator are characterized by low Mediator ChIP-Seq signal, and associate with genes linked to ubiquitous biological processes such as transcription, indicating that OSN-Med1-identified super-enhancers are a subset of the super-enhancers identified by H3K27ac, and suggest that transcription factors

other than cell type specific master transcription factors may form large domains in the genome. *(bottom, heatmap)* Heatmap representation of H3K27ac and Med1 ChIP-Seq densities at the 392 super-enhancer regions identified by H3K27ac in mESC. Super-enhancer regions are shown along with 5kb flanking distance up- and downstream. Color scale reflects the density of H3K27ac signal at the super-enhancer regions. *(bottom, GO analysis)* Gene Ontology analysis the indicated gene sets.
B) Heatmap showing the classification of super-enhancer-associated genes across 26 human cell and tissue types. Each row is a gene and red color indicates the gene being associated with a super-enhancer in the respective cell type.
C) Heatmap showing the classification of typyical enhancer-associated genes across 26 human cell and tissue types. Each row is a gene and red color indicates the gene being associated with a super-enhancer in the respective cell type.
D) Super-enhancers can overlap with Locus Control Regions (LCR), Transcription Initiation Platforms (TIP) and DNA methylation valleys (DMV).

Supplemental Figure 4. Related to Figure 4.
(A) Summary of trait-associated SNPs in the union of super-enhancers, typical enhancers and regulatory regions (defined as H3K27ac binding peaks) in 86 human cell and tissue samples. Displayed is the percentage of the total 4,378 trait-associated non-coding SNPs falling into these regions. The percentage of the genome (3.4 billion bases) covered by the union of these regions in the 86 human cell and tissue types is also displayed. SNP enrichment is defined as the percent of SNPs contained in the percent of the genome covered for these regions.
B) The SNP enrichment values of non-coding SNPs linked to the highlighted traits and diseases in the union of super-enhancers, typical enhancers and regulatory regions (defined as H3K27ac binding peaks) in 86 human cell and tissue samples.

Supplemental Figure 5. Related to Figure 5.
Radar plots showing the density of non-coding SNPs linked to selected diseases in the super-enhancer domains and typical enhancers identified in 12 human cell and tissue types. The center of the plot is 0, and a colored dot on the respective axis indicates the SNP density (SNP/10MB sequence) in the super-enhancer domains or typical enhancers of each cell and tissue type. Lines connecting the density values to the origin of the plot are added to improve visualization.

Supplemental Figure 6. Related to Figure 6.
ChIP-Seq binding profiles for H3K27ac at the *c-MYC* locus in the indicated cancer samples. Super-enhancers are highlighted as black bars above the binding profile. Purple track indicates the topological domains identified in hESC in (Dixon et al., 2012). (rpm/bp: reads per million per base pair)

Supplementary Table 1: Catalogue of ChIP-Seq densities in constituents of super-enhancers and typical enhancers and corresponding control samples in mESCs. Reads per million mapped reads per base pair densities with background subtraction are shown in constituents of super-enhancers and typical enhancers (sheet 1) as well as in control regions randomly shifted on the same chromosome (sheet 2).

Supplementary Table 2: Catalogue of 86 human cell and tissue types in which super-enhancers and super-enhancer—associated genes were identified. The tables containing the super-enhancers and super-enhancer—associated genes are found in Supplemental Data 1. The catalogue highlights the cell and tissue types. Representative datasets (Rep. Dataset) were used in Figure 3A, 3B, 3C and Figure

S3B,C. Datasets used for SNP analyses were included for Figures 4B, 5, and Figure S5. Cancer ID identifies the samples included on Figure 6A and Figure S6. Additional column indicates whether the sample was used for analyses on Figure 6B-E. CML stands for chronic myelogenous leukemia.

Supplemental Table 3: Catalogue of candidate master transcription factors in 86 human cell and tissue samples. Super-enhancer—associated transcription factor genes are displayed. Classification as a transcription factor was determined by inclusion in both the AnimalTFDB and TcoF transcription factor databases. Candidate master transcription factors are ranked by their signal of H3K27ac at their associated super-enhancers.

Supplemental Table 4: GWAS SNPs used in this study and their distributions in super-enhancers and typical enhancers. Single-nucleotide polymorphisms associated with traits were downloaded from the NHGRI genome-wide association study database (08/09/2013). Only SNPs that had a dbSNP identifier, and whose trait association was replicated in at least two studies were considered in the downstream analyses (sheet 1). Unique SNPs found outside of coding regions are listed in Sheet 2, and were used for analyses on Figures 4A right panel, 4B, 5, Figures S4 and S5. Sheets 3 and 4 show the number of unique, non-coding, replicated SNPs found in the super-enhancers and typical enhancers of the 86 human cell and tissue samples, respectively.

Supplemental Table 5: Cancer hallmark genes associated with super-enhancers acquired in colorectal cancer, T cell leukemia and pancreatic cancer. The fourth sheet lists all genes that acquire super-enhancer compared to their healthy counterparts, respectively.

Supplemental Table 6: GEO accession identifiers for all samples used in the study.

Supplemental Data 1: Zip file containing 86 files, each containing stitched enhancers, associated genes, and ChIP-Seq signal. Columns are: enhancer ID, chromosome, start, end, associated gene, enhancer rank, is enhancer a super-enhancer (1:yes, 0:no), H3K27ac ChIP-Seq density (rpm/bp), read density in corresponding input sample (rpm/bp).

Supplemental Data 2: Zip file containing 86 files, each a pair of UCSC Genome Browser tracks. Tracks displayed in black are all stitched enhancers. Tracks displayed in red are super-enhancers.

Supplemental Experimental Procedures

ChIP-Seq

ChIP-Seq for Ronin was previously described (Dejosez et al., 2010). ChIP-Seq for Mbd3 in the murine ESC line V6.5 was performed with an anti-Mbd3 antibody (Santa Cruz, SC-9402) as described (Whyte et al., 2012). ChIP-Seq for CBP in the murine ESC line V6.5 was performed with an anti-CBP antibody (Santa Cruz, SC-9402) as described (Mullen et al., 2011). ChIP-Seq for H3K27ac in RPMI-8402 cells was performed with an anti-H3K27ac antibody (Abcam, ab4729) as described (Sanda et al., 2012).

Several as yet unreleased ChIP-Seq datasets were generously shared by the NIH Roadmap Epigenome project (Bernstein et al., 2010).

ChIP-Seq density analysis

ChIP-Seq read density was measured as described in (Lin et al., 2012). Briefly, ChIP-Seq reads were extended 200bp and the density of reads per base pair was calculated. This density was normalized to the millions of mapped reads contributing to the density, measured in reads per million per base pair (rpm/bp).

Percent enhancer signal falling in super-enhancers (Figures 1D and 2C) was calculated using the sum of signal (density * length) of super-enhancers and typical enhancers. Comparison of super-enhancer vs. typical enhancer signal was calculated using mean background-subtracted signal (density * length) of super-enhancers divided by the sum of super-enhancers and typical enhancers.

Definition of enhancers and super-enhancers in mESC

Genomic coordinates of murine embryonic stem cell typical enhancers, super-enhancers, typical enhancer constituents and super-enhancer constituents were downloaded from (Whyte et al., 2013).

Threshold for occupancy at enhancers

For Figures 1 and 2, the occupancy of transcription factors, co-factors and chromatin regulators at enhancers in ESCs was determined as follows. The mean ChIP-Seq density at every enhancer constituent was calculated for each transcription factor, co-factor and chromatin regulator listed in Supplemental Table 1. Mean ChIP-Seq density values measured in the corresponding input samples were subtracted. To correct for ChIP-Seq background signal, a minimum value of 0.2 rpm/bp after background subtraction was required for further consideration. To correct for different ChIP-Seq qualities across multiple samples, the mean ChIP-Seq densities of random genomic regions of equivalent sizes of the enhancer constituents were calculated (Supplemental Table 1), and the ratio of the mean ChIP-seq signal at enhancer constituents and the mean ChIP-seq signal at the random genomic positions was calculated. Presence of a factor at enhancers was defined as this ratio >9. We found that these two thresholds largely captured the definition of presence or absence at enhancers for several factors where genomic localization of the factor has previously been analyzed in a similar context.

USCS Browser tracks

To assess the relative enrichment of binding peaks compared to other peaks in ChIP-Seq samples displayed as UCSC Browser tracks (Figure 1B, Figure 2A, Figure S3A) the median enrichment value for all bound regions was calculated. MACS was used to identify enriched regions (as described below), and the background subtracted read density at the enriched regions was determined. The median read density value of all enriched regions is denoted by a red dot on the y-axis of the UCSC Browser tracks.

Figure 6C *(top)*, depicts a described translocation event that between chr8(q24) and chr3(q21) in the multiple myeloma cell line MM1S (Shou et al., 2000). The segment of chromosomal region chr3(q21) depicted on the figure is chr3:122,500,000-124,250,000.

The sites of focal amplification Figure 6C (bottom) in small cell lung cancer were described in (Iwakawa et al., 2013).

Overlaps with previously described large genomic regions

The genomic co-ordinates for the DNA-methylation valleys were obtained from (Xie and Ren, 2013). The cell types pooled on Figure S3D include, hESC, mesenchymal stem cell, mesendoderm, neural progenitor and trophoblast (DMV), and hESC, fetal intestine, fetal large intestine, fetal thymus, fetal hematopoietic progenitor and fetal muscle (SE).

The genomic co-ordinates for the globin LCR and TIP at the IFNAR1 loci were adapted from (Bonifer, 2000; Koch et al., 2011)

Metagenes and Heatmaps

Genome-wide average “meta” representations of ChIP-Seq density at typical enhancers and super-enhancers (Figure 1C, 2B) were created by mapping reads to the enhancer regions and flanking regions. Each enhancer or flanking region was split into 100 equally sized bins. This split all enhancer regions, regardless of their size, into 300 bins. All typical enhancer or super-enhancer regions were then aligned and the average ChIP-seq density in each bin was calculated to create a meta genome-wide average in units of reads per million per base pair. In order to visualize the length disparity between typical and super-enhancer regions, the enhancer region (between its actual start and end) was scaled relative to its median length.

Constituent metagenes (Figure 1E, 2D, Figure S1) were created in a similar fashion. Constituents of super- and typical enhancers, as well as 2.5kb upstream and downstream were each broken into 50 bins. The ChIP-Seq density in these regions was calculated in and combined together to get 150 bins spanning 2.5kb upstream, the constituent enhancer, and 2.5kb downstream. The average combined profiles for the super- or typical enhancers constituents is shown.

Metagenes around expressed promoters were similarly created. Promoters were defined as +/- 1kb around the TSS. 9,667 expressed promoters were defined as those having RNAPII ChIP-Seq density > 1 rpm/bp. The regions were broken into 50 bins and the average read density for expressed promoters is shown (Figure S1A).

Metagenes around boundaries of topological domains (TD) were similarly created. Topological domains defined in (Dixon et al., 2012) were downloaded for the mm9 genome build. Regions interrogated were +/- 500kb from the TD border and were split into 100 bins. The average read density per bin is shown (Figure S1A).

Heatmaps in Figure 3A and Figure S3A were calculated in a manner similar to metagenes. The union of 26 sets of representative super-enhancers resulted in 5,988 regions used for H3K27ac density analysis in Figure 3A. Each element in the union was broken into 50 equally sized bins. Reads were extended 200bp and reads-per-million densities were calculated in each element in the union of super-enhancers. Figure S3A contains a heatmap showing densities in H3K27ac-defined super-enhancers. Regions 5kb upstream and downstream of the super-enhancers are shown. Each region (upstream, super-enhancer, downstream) is broken into 50 equally sized bins. Reads were extended 200bp and reads-per-million normalized densities were calculated in these regions.

Motif analysis

To find sequence motifs enriched in super-enhancers in murine ESCs, we analyzed the genomic sequence under the constituents within super-enhancers. We extracted their sequence from the mm9 genome and used this as input for TRAP using TRANSFAC vertebrates as the comparison library, mouse promoters as the control, and Benjamini-Hochberg as the correction (Thomas-Chollier et al., 2011). To include Tcfcp2l1, we used the Jaspar vertebrates as the comparison library. P-values displayed in Figure 1E correspond to the corrected P in the output. Motif enrichment analysis at typical enhancers (Figure S1B) was done the same way, but only the enhancer constituents with a size smaller than 500 bp were used, because of limitations of TRAP. P-values displayed in Figure S1B correspond to the corrected P in the output.

Matrices used: Oct4: M01124; Sox2: M01272; Nanog: M01123; Klf4: M01588; Esrrb: M01589; Stat3: M01595; Tcf3: M01594; Smad3: M00701; Tcfcp2l1: MA0145.

The motif logos displayed on Figure 1E and Supplemental Figure 1B were downloaded from the Cistrome database (Liu et al., 2011).

Identifying ChIP-Seq enriched regions in human cells

Human sequencing reads were aligned to the human genome build hg19 (GRCh37) using bowtie 0.12.9 (Langmead et al., 2009) using parameters -k 2, -m 2, -n 2, --best. Mouse sequencing reads were aligned to the mouse genome build mm9 (NCBI37) using bowtie 0.12.9 using parameters -k 1, -m 1, -n 2, --best. Regions of enrichment of H3K27ac in human samples were calculated using MACS 1.4.2 (Zhang et al., 2008) using parameters -p 1e-9, --keep-dup=auto, -w –S –space=50, and -g hs on H3K27ac ChIP-Seq with control libraries. MACS peaks were called on mouse ChIP-Seq using –p 1e-9, --keep-dup=auto, -w –S –space=50, -g mm. UCSC Genome Browser (Kent et al., 2002) tracks were generated using MACS wiggle outputs. MACS peaks of human H3K27ac were used as constituent enhancers for super-enhancer identification.

Definition of enhancers and super-enhancers in human cells

Enhancers were stitched and super-enhancers were identified using ROSE (https://bitbucket.org/young_computation/rose), which is an implementation of

the algorithm described in (Loven et al., 2013). Briefly, this algorithm stitches constituent enhancers together if they are within a certain distance and ranks the enhancers by their input-subtracted signal of H3K27ac. It then separates super-enhancers from typical enhancers by identifying an inflection point of H3K27ac signal vs. enhancer rank (Whyte et al., 2013). ROSE was run with a stitching distance of 12,500 bp, i.e. we allowed enhancers within 12,500 bp to be stitched together. In addition, we used a promoter exclusion zone of 2,000 bp, i.e. if a constituent enhancer was wholly contained within a window +/- 1,000 bp around an annotated transcription start site, the constituent enhancer was excluded from stitching.

Stitched enhancers were assigned to the expressed transcript whose TSS was the nearest to the center of the stitched enhancer. Expressed transcripts were defined as having an at least 0.5 mean rpm/bp H3K27ac ChIP-Seq density in a window 500 bases up- and downstream of the TSS. In cases where enhancer-gene assignments were previously verified by experimental techniques, we assigned genes based on those studies. These examples include: *MYC* in multiple myeloma, *IRF4* in multiple myeloma (Loven et al., 2013), *MYC* in colorectal cancer, breast cancer and prostate cancer (Ahmadiyeh et al., 2010; Pomerantz et al., 2009) (all on Figure 6).

For the Figure 3A heatmap, the union of super-enhancers for 26 cell types was taken. The rpm/bp density of ChIP-Seq reads (extended by 200bp) was calculated for the corresponding 26 H3K27ac datasets and is represented by color intensity. Each region in the union of super-enhancers was compared to every super-enhancer in the 26 cell types, and the cell types containing a super-enhancer that contacts a region in the union were recorded. The heatmap is sorted first by the number of cell types containing a contacting super-enhancer and then by the order of cell types containing a contacting super-enhancer.

Gene sets and annotations

All analyses were performed using RefSeq (GRCh37/hg19) human gene annotations or RefSeq (NCBI37/mm9) mouse gene annotations (Pruitt et al., 2007).

The high confidence set of transcription factors used for analysis was the intersection of genes identified as transcription factors in two different transcription factor databases (AnimalTFDB and TcoF) (Schaefer et al., 2011; Zhang et al., 2012).

References for the validation of candidate master transcription factors

The references validating the transcription factors listed on Figure 3C, as candidate master transcription factors are as follows:

Brain: NKX2-2 (Briscoe et al., 1999; Panman et al., 2011), OLIG1 (Arnett et al., 2004), BRN2 (Ambasudhan et al., 2011; McEvilly et al., 2002; Pfisterer et al., 2011; Son et al., 2011; Sugitani et al., 2002), SOX10 (Bondurand and Sham, 2013; Britsch et al., 2001; Lee et al., 2008), SOX2 (Bergsland et al., 2011; Cavallaro et al., 2008; Ferri et al., 2004; Han et al., 2012; Ring et al., 2012).

Heart: TBX20 (Cai et al., 2005; Cai et al., 2013; Takeuchi et al., 2005), TBX5 (Ieda et al., 2010; Nadeau et al., 2010; Qian et al., 2012; Song et al., 2012), MEF2A (Naya et al., 2002; Schlesinger et al., 2011), NKX2-5 (Lyons et al., 1995; Schlesinger et al., 2011), GATA4 (Ieda et al., 2010; Nadeau et al., 2010; Qian et al., 2012; Song et al., 2012; Turbendian et al., 2013; Watt et al., 2004).

Skeletal muscle: MYOD1 (Bergstrom et al., 2002; Davis et al., 1987; Rudnicki et al., 1993; Tajbakhsh et al., 1997), PITX2 (Gherzi et al., 2010; Lin et al., 1999), SIX1 (Grifone et al., 2004; Yajima et al., 2010), TEAD4 (Benhaddou et al., 2012).

Lung: NFIB (Hsu et al., 2011), TBX5 (Arora et al., 2012), CEBPA (Martis et al., 2006), TBX2 (Ludtke et al., 2013), TBX3 (Ludtke et al., 2013).

Adipose tissue: PPARG (Lehrke and Lazar, 2005; Rosen et al., 1999; Rosen et al., 2000; Schupp et al., 2009), CEBPB (Cao et al., 1991; Kajimura et al., 2009), CEBPD (Rosen et al., 2000; Tanaka et al., 1997), CREB1 (Reusch et al., 2000).

B cell: IKZF3 (Ferreiros-Vidal et al., 2013; Kioussis, 2007; Ma et al., 2010), PAX5 (Busslinger, 2004; Medvedovic et al., 2011), BACH2 (Kallies and Nutt, 2010), OCT2 (Gstaiger et al., 1996; Wirth et al., 1995), IKZF1 (Ferreiros-Vidal et al., 2013; Kioussis, 2007; Ma et al., 2010), IRF8 (Busslinger, 2004).

Gene Ontology (GO) Analysis

For gene ontology analysis, a subset of 26 datasets, representing the diversity of tissues in the collection used for this study, were first selected. For each tissue, the genes that were associated with super-enhancers in that tissue and no more than two other tissues in the subset were analyzed using DAVID (http://david.abcc.ncifcrf.gov/home.jsp). For each tissue, the three top scoring categories (ie the categories with the lowest p-values) were selected for display. A threshold p-value score of 2E-05 was incorporated as a minimum requirement filter for scoring as a top category.

Trait-associated SNPs

Trait-associated SNPs were downloaded from the NHGRI database of genome-wide association studies on August 9, 2013, which contained 13,957 entries/rows. Since SNPs more strongly associated with a trait are suggested to have a higher likelihood of being causative (Maurano et al., 2012), we only considered SNPs that have a dbSNP identifier and were found to be associated with a trait in at least two independent studies. 5,303 such SNP-trait or disease associations were used for the left and center panels of Figure 4A. 4,912 non-coding SNP-trait associations were used for 4B, 5, and Supplemental Figure 5. 4,378 unique SNPs located outside coding regions were used for Figure 4A right panel and Supplemental Table S4.

Figure 4A right panel shows the distance from trait-associated, non-coding SNPs to the nearest border of a region in the union of 86 super-enhancer sets. SNPs within these regions were assigned to the 0 bin.

Significance of the number of SNPs in super-enhancers was calculated using a permutation test. Super-enhancer—sized regions were randomly shifted on

the chromosome of origin 10,000 times. The number of SNPs falling in these shifted regions was counted. No repetition resulted in the same or greater number of trait-associated SNPs in super-enhancer—sized regions.

Radar plots

The density of trait-associated non-coding SNPs in super-enhancer domains and typical enhancers of individual cell and tissue samples were calculated by first counting the number of SNPs that are found in these regions. The numbers were then divided by the number of base pairs super-enhancer domains and typical enhancers cover of genome in these cells, and multiplied by 10 million, to give a SNP/10MB dimension (Figure 4B, 5, Figure S5).

Selection of oncogenes

Proto-oncogenes for display in Figure 6A were selected based on their presence in the COSMIC (Catalogue Of Somatic Mutations In Cancer) or AOGIC (Amplified and Overexpressed Genes in Cancer) databases (Forbes et al., 2010; Santarius et al., 2010).

Forbes et al., Nucleic Acids Research, 2010; Santarius et al., Nat. Rev. Cancer, 2010

Cancer hallmark analysis

The following gene ontology categories were used as proxies for the characteristic hallmark capabilities that are thought to be acquired in cancers.

- Angiogenesis:
  - GO:0001525 – Angiogenesis
- Enabling Replicative Immortality:
  - GO:0032200 – Telomere organization
  - GO:0090398 – Cellular senescence
  - GO:0090399 – Replicative senescence
- Activating Invasion:
  - GO:0034330 – Cell junction organization
  - GO:0016477 – Cell migration
  - GO:0010718 – Positive regulation of epithelial to mesenchymal transition
  - GO:0007155 – Cell adhesion
- Genome Instability:
  - GO:0006281 – DNA repair
  - GO:0051383 – Kinetochore organization
  - GO:0007065 – Sister chromatid cohesion
  - GO:0000819 – Sister chromatid segregation
  - GO:0051988 – Regulation of attachment of spindle microtubules to kinetochore
  - GO:0030997 – Regulation of centriole-centriole cohesion
  - GO:0046605 – Regulation of centrosome cycle
  - GO:0090224 – Regulation of spindle organization
  - GO:0010695 – Regulation ofspindle pole body separation
  - GO:0031577 – Spindle checkpoint
- Resisting Cell Death:

GO:0060548 – Negative regulation of cell death
GO:0012501 – Programmed cell death
GO:0010941 – Regulation of cell death

Disrupting Cellular Energetics:
GO:0006091 – Generation of precursor metabolites and energy

Sustaining proliferative signaling:
GO:0007166 – Cell surface receptor signaling pathway
GO:0070848 – Response to growth factor stimulus

Tumor-Promoting Inflammation:
GO:0006954 – Inflammatory response
GO:0045321 – Leukocyte activation

Avoiding Immune Destruction:
GO:0002507 – Tolerance induction
GO:0001910 – Regulation of leukocyte mediated cytotoxicity
GO:0019882 – Antigen processing and presentation
GO:0002767 – Immune response-inhibiting cell surface receptor signaling pathway

Evading Growth Suppressors:
GO:0007049 – Cell cycle
GO:0008283 – Cell proliferation

RNA-Seq

RNA extraction, purification, quality control and sequencing was performed as described (Sigova et al., 2013). Reads were aligned using TopHat (Trapnell et al., 2009) as a paired-end library with parameters –library-type fr-firststrand –microexon-search –coverage-search. RNA-Seq reads were not extended for density analyses, and reads that mapped to exonic sequences were removed for the analysis on Figure 2C.

Gene expression analysis

Microarray gene expression data used on Figure S2 were previously generated and described (Kagey et al., 2010; Whyte et al., 2013).

Supplemental References

Ahmadiyeh, N., Pomerantz, M.M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., He, H.H., Brown, M., Liu, X.S., Davis, M.*, et al.* (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proceedings of the National Academy of Sciences of the United States of America *107*, 9742-9746.

Ambasudhan, R., Talantova, M., Coleman, R., Yuan, X., Zhu, S., Lipton, S.A., and Ding, S. (2011). Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. Cell stem cell *9*, 113-118.

Arnett, H.A., Fancy, S.P., Alberta, J.A., Zhao, C., Plant, S.R., Kaing, S., Raine, C.S., Rowitch, D.H., Franklin, R.J., and Stiles, C.D. (2004). bHLH transcription factor Olig1 is required to repair demyelinated lesions in the CNS. Science *306*, 2111-2115.
Arora, R., Metzger, R.J., and Papaioannou, V.E. (2012). Multiple roles and interactions of Tbx4 and Tbx5 in development of the respiratory system. PLoS genetics *8*, e1002866.
Benhaddou, A., Keime, C., Ye, T., Morlon, A., Michel, I., Jost, B., Mengus, G., and Davidson, I. (2012). Transcription factor TEAD4 regulates expression of myogenin and the unfolded protein response genes during C2C12 cell differentiation. Cell death and differentiation *19*, 220-231.
Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. Genes & development *25*, 2453-2464.
Bergstrom, D.A., Penn, B.H., Strand, A., Perry, R.L., Rudnicki, M.A., and Tapscott, S.J. (2002). Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. Molecular cell *9*, 587-600.
Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R.*, et al.* (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nature biotechnology *28*, 1045-1048.
Bondurand, N., and Sham, M.H. (2013). The role of SOX10 during enteric nervous system development. Developmental biology.
Bonifer, C. (2000). Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? Trends in genetics : TIG *16*, 310-315.
Briscoe, J., Sussel, L., Serup, P., Hartigan-O'Connor, D., Jessell, T.M., Rubenstein, J.L., and Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. Nature *398*, 622-627.
Britsch, S., Goerich, D.E., Riethmacher, D., Peirano, R.I., Rossner, M., Nave, K.A., Birchmeier, C., and Wegner, M. (2001). The transcription factor Sox10 is a key regulator of peripheral glial development. Genes & development *15*, 66-78.
Busslinger, M. (2004). Transcriptional control of early B cell development. Annual review of immunology *22*, 55-79.
Cai, C.L., Zhou, W., Yang, L., Bu, L., Qyang, Y., Zhang, X., Li, X., Rosenfeld, M.G., Chen, J., and Evans, S. (2005). T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. Development *132*, 2475-2487.
Cai, X., Zhang, W., Hu, J., Zhang, L., Sultana, N., Wu, B., Cai, W., Zhou, B., and Cai, C.L. (2013). Tbx20 acts upstream of Wnt signaling to regulate endocardial cushion formation and valve remodeling during mouse cardiogenesis. Development *140*, 3176-3187.
Cao, Z., Umek, R.M., and McKnight, S.L. (1991). Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. Genes & development *5*, 1538-1552.
Cavallaro, M., Mariani, J., Lancini, C., Latorre, E., Caccia, R., Gullo, F., Valotta, M., DeBiasi, S., Spinardi, L., Ronchi, A.*, et al.* (2008). Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. Development *135*, 541-557.
Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987-1000.
Dejosez, M., Levine, S.S., Frampton, G.M., Whyte, W.A., Stratton, S.A., Barton, M.C., Gunaratne, P.H., Young, R.A., and Zwaka, T.P. (2010). Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. Genes & development *24*, 1479-1484.
Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Ferreiros-Vidal, I., Carroll, T., Taylor, B., Terry, A., Liang, Z., Bruno, L., Dharmalingam, G., Khadayate, S., Cobb, B.S., Smale, S.T.*, et al.* (2013). Genome-wide identification of Ikaros targets elucidates its contribution to mouse B-cell lineage specification and pre-B-cell differentiation. Blood *121*, 1769-1782.
Ferri, A.L., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P.P., Sala, M., DeBiasi, S.*, et al.* (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. Development *131*, 3805-3819.
Forbes, S.A., Tang, G., Bindal, N., Bamford, S., Dawson, E., Cole, C., Kok, C.Y., Jia, M., Ewing, R., Menzies, A.*, et al.* (2010). COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. Nucleic acids research *38*, D652-657.
Gherzi, R., Trabucchi, M., Ponassi, M., Gallouzi, I.E., Rosenfeld, M.G., and Briata, P. (2010). Akt2-mediated phosphorylation of Pitx2 controls Ccnd1 mRNA decay during muscle cell differentiation. Cell death and differentiation *17*, 975-983.
Grifone, R., Laclef, C., Spitz, F., Lopez, S., Demignon, J., Guidotti, J.E., Kawakami, K., Xu, P.X., Kelly, R., Petrof, B.J.*, et al.* (2004). Six1 and Eya1 expression can reprogram adult muscle from the slow-twitch phenotype into the fast-twitch phenotype. Molecular and cellular biology *24*, 6253-6267.
Gstaiger, M., Georgiev, O., van Leeuwen, H., van der Vliet, P., and Schaffner, W. (1996). The B cell coactivator Bob1 shows DNA sequence-dependent complex formation with Oct-1/Oct-2 factors, leading to differential promoter activation. The EMBO journal *15*, 2781-2790.
Han, D.W., Tapia, N., Hermann, A., Hemmer, K., Hoing, S., Arauzo-Bravo, M.J., Zaehres, H., Wu, G., Frank, S., Moritz, S.*, et al.* (2012). Direct reprogramming of fibroblasts into neural stem cells by defined factors. Cell stem cell *10*, 465-472.
Hsu, Y.C., Osinski, J., Campbell, C.E., Litwack, E.D., Wang, D., Liu, S., Bachurski, C.J., and Gronostajski, R.M. (2011). Mesenchymal nuclear factor I B regulates cell proliferation and epithelial differentiation during lung maturation. Developmental biology *354*, 242-252.
Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375-386.
Iwakawa, R., Takenaka, M., Kohno, T., Shimada, Y., Totoki, Y., Shibata, T., Tsuta, K., Nishikawa, R., Noguchi, M., Sato-Otsubo, A.*, et al.* (2013). Genome-wide identification of genes with amplification and/or fusion in small cell lung cancer. Genes, chromosomes & cancer *52*, 802-816.
Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.
Kajimura, S., Seale, P., Kubota, K., Lunsford, E., Frangioni, J.V., Gygi, S.P., and Spiegelman, B.M. (2009). Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex. Nature *460*, 1154-1158.
Kallies, A., and Nutt, S.L. (2010). Bach2: plasma-cell differentiation takes a break. The EMBO journal *29*, 3896-3897.
Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome research *12*, 996-1006.
Kioussis, D. (2007). Aiolos: an ungrateful member of the Ikaros family. Immunity *26*, 275-277.
Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C.*, et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nature structural & molecular biology *18*, 956-963.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.
Lee, K.E., Nam, S., Cho, E.A., Seong, I., Limb, J.K., Lee, S., and Kim, J. (2008). Identification of direct regulatory targets of the transcription factor Sox10 based on function and conservation. BMC genomics *9*, 408.
Lehrke, M., and Lazar, M.A. (2005). The many faces of PPARgamma. Cell *123*, 993-999.
Lin, C.R., Kioussi, C., O'Connell, S., Briata, P., Szeto, D., Liu, F., Izpisua-Belmonte, J.C., and Rosenfeld, M.G. (1999). Pitx2 regulates lung asymmetry, cardiac positioning and pituitary and tooth morphogenesis. Nature *401*, 279-282.
Lin, C.Y., Loven, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. Cell *151*, 56-67.
Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome biology *12*, R83.
Loven, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320-334.
Ludtke, T.H., Farin, H.F., Rudat, C., Schuster-Gossler, K., Petry, M., Barnett, P., Christoffels, V.M., and Kispert, A. (2013). Tbx2 controls lung growth by direct repression of the cell cycle inhibitor genes Cdkn1a and Cdkn1b. PLoS genetics *9*, e1003189.
Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L., and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. Genes & development *9*, 1654-1666.
Ma, S., Pathak, S., Mandal, M., Trinh, L., Clark, M.R., and Lu, R. (2010). Ikaros and Aiolos inhibit pre-B-cell proliferation by directly suppressing c-Myc expression. Molecular and cellular biology *30*, 4149-4158.
Martis, P.C., Whitsett, J.A., Xu, Y., Perl, A.K., Wan, H., and Ikegami, M. (2006). C/EBPalpha is required for lung maturation at birth. Development *133*, 1155-1164.
Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J.*, et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190-1195.
McEvilly, R.J., de Diaz, M.O., Schonemann, M.D., Hooshmand, F., and Rosenfeld, M.G. (2002). Transcriptional regulation of cortical neuron migration by POU domain factors. Science *295*, 1528-1532.
Medvedovic, J., Ebert, A., Tagoh, H., and Busslinger, M. (2011). Pax5: a master regulator of B cell development and leukemogenesis. Advances in immunology *111*, 179-206.
Mullen, A.C., Orlando, D.A., Newman, J.J., Loven, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master transcription factors determine cell-type-specific responses to TGF-beta signaling. Cell *147*, 565-576.
Nadeau, M., Georges, R.O., Laforest, B., Yamak, A., Lefebvre, C., Beauregard, J., Paradis, P., Bruneau, B.G., Andelfinger, G., and Nemer, M. (2010). An endocardial pathway involving Tbx5, Gata4, and Nos3 required for atrial septum formation. Proceedings of the National Academy of Sciences of the United States of America *107*, 19356-19361.
Naya, F.J., Black, B.L., Wu, H., Bassel-Duby, R., Richardson, J.A., Hill, J.A., and Olson, E.N. (2002). Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. Nature medicine *8*, 1303-1309.

Panman, L., Andersson, E., Alekseenko, Z., Hedlund, E., Kee, N., Mong, J., Uhde, C.W., Deng, Q., Sandberg, R., Stanton, L.W.*, et al.* (2011). Transcription factor-induced lineage selection of stem-cell-derived neural progenitor cells. Cell stem cell *8*, 663-675.
Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., Nelander, J., Dufour, A., Bjorklund, A., Lindvall, O., Jakobsson, J., and Parmar, M. (2011). Direct conversion of human fibroblasts to dopaminergic neurons. Proceedings of the National Academy of Sciences of the United States of America *108*, 10343-10348.
Pomerantz, M.M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M.P., Doddapaneni, H., Beckwith, C.A., Chan, J.A., Hills, A., Davis, M.*, et al.* (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nature genetics *41*, 882-884.
Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research *35*, D61-65.
Qian, L., Huang, Y., Spencer, C.I., Foley, A., Vedantham, V., Liu, L., Conway, S.J., Fu, J.D., and Srivastava, D. (2012). In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. Nature *485*, 593-598.
Reusch, J.E., Colton, L.A., and Klemm, D.J. (2000). CREB activation induces adipogenesis in 3T3-L1 cells. Molecular and cellular biology *20*, 1008-1020.
Ring, K.L., Tong, L.M., Balestra, M.E., Javier, R., Andrews-Zwilling, Y., Li, G., Walker, D., Zhang, W.R., Kreitzer, A.C., and Huang, Y. (2012). Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor. Cell stem cell *11*, 100-109.
Rosen, E.D., Sarraf, P., Troy, A.E., Bradwin, G., Moore, K., Milstone, D.S., Spiegelman, B.M., and Mortensen, R.M. (1999). PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. Molecular cell *4*, 611-617.
Rosen, E.D., Walkey, C.J., Puigserver, P., and Spiegelman, B.M. (2000). Transcriptional regulation of adipogenesis. Genes & development *14*, 1293-1307.
Rudnicki, M.A., Schnegelsberg, P.N., Stead, R.H., Braun, T., Arnold, H.H., and Jaenisch, R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. Cell *75*, 1351-1359.
Sanda, T., Lawton, L.N., Barrasa, M.I., Fan, Z.P., Kohlhammer, H., Gutierrez, A., Ma, W., Tatarek, J., Ahn, Y., Kelliher, M.A.*, et al.* (2012). Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer cell *22*, 209-221.
Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., and Cooper, C.S. (2010). A census of amplified and overexpressed human cancer genes. Nature reviews Cancer *10*, 59-64.
Schaefer, U., Schmeier, S., and Bajic, V.B. (2011). TcoF-DB: dragon database for human transcription co-factors and transcription factor interacting proteins. Nucleic acids research *39*, D106-110.
Schlesinger, J., Schueler, M., Grunert, M., Fischer, J.J., Zhang, Q., Krueger, T., Lange, M., Tonjes, M., Dunkel, I., and Sperling, S.R. (2011). The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. PLoS genetics *7*, e1001313.
Schupp, M., Cristancho, A.G., Lefterova, M.I., Hanniman, E.A., Briggs, E.R., Steger, D.J., Qatanani, M., Curtin, J.C., Schug, J., Ochsner, S.A.*, et al.* (2009). Re-expression of GATA2 cooperates with peroxisome proliferator-activated receptor-gamma depletion to revert the adipocyte phenotype. The Journal of biological chemistry *284*, 9458-9464.
Shou, Y., Martelli, M.L., Gabrea, A., Qi, Y., Brents, L.A., Roschke, A., Dewald, G., Kirsch, I.R., Bergsagel, P.L., and Kuehl, W.M. (2000). Diverse karyotypic abnormalities of the c-myc locus associated with c-myc dysregulation and tumor

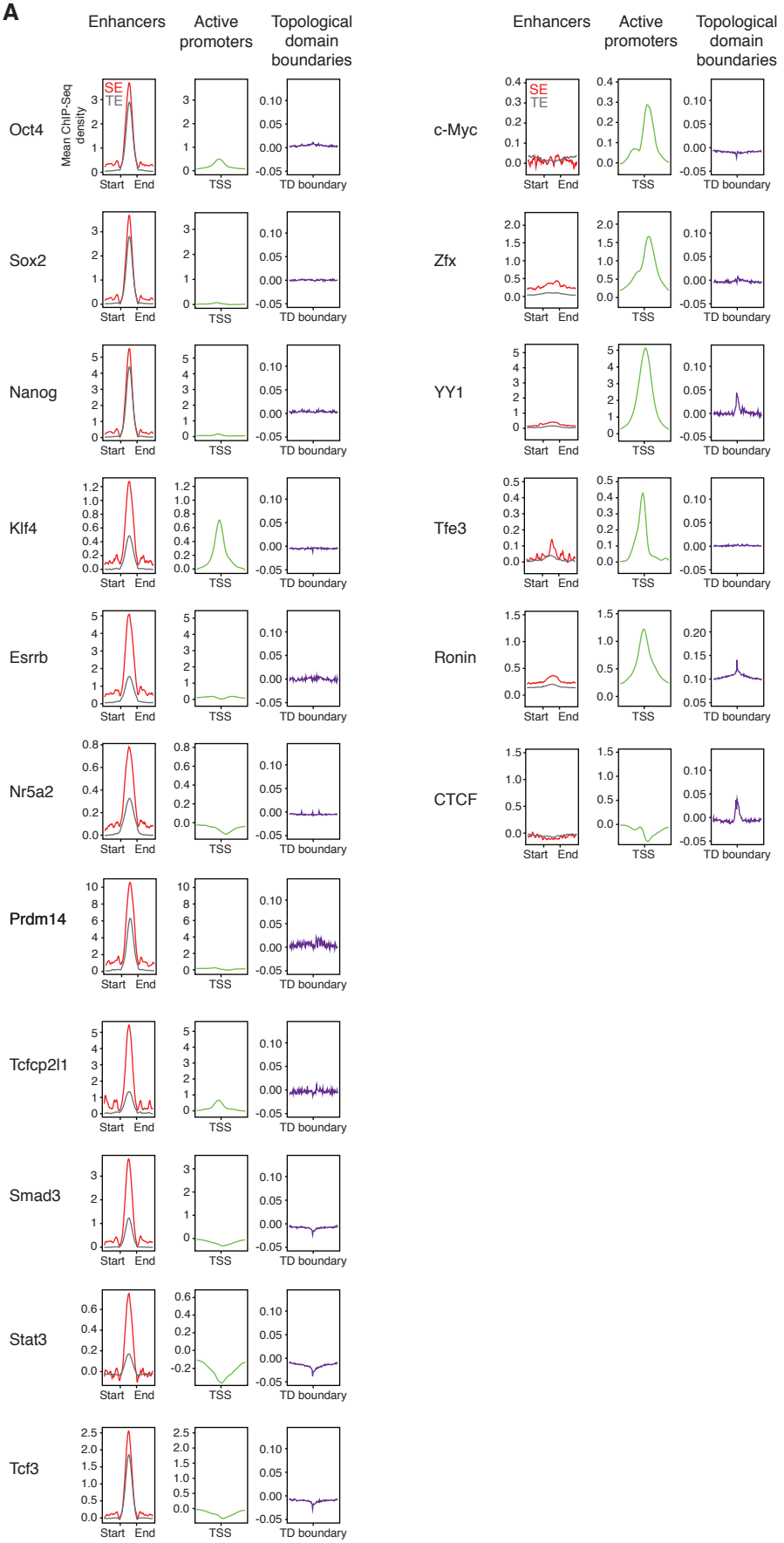
progression in multiple myeloma. Proceedings of the National Academy of Sciences of the United States of America *97*, 228-233.
Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C.*, et al.* (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America *110*, 2876-2881.
Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. Cell stem cell *9*, 205-218.
Song, K., Nam, Y.J., Luo, X., Qi, X., Tan, W., Huang, G.N., Acharya, A., Smith, C.L., Tallquist, M.D., Neilson, E.G.*, et al.* (2012). Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature *485*, 599-604.
Sugitani, Y., Nakai, S., Minowa, O., Nishi, M., Jishage, K., Kawano, H., Mori, K., Ogawa, M., and Noda, T. (2002). Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons. Genes & development *16*, 1760-1765.
Tajbakhsh, S., Rocancourt, D., Cossu, G., and Buckingham, M. (1997). Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. Cell *89*, 127-138.
Takeuchi, J.K., Mileikovskaia, M., Koshiba-Takeuchi, K., Heidt, A.B., Mori, A.D., Arruda, E.P., Gertsenstein, M., Georges, R., Davidson, L., Mo, R.*, et al.* (2005). Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. Development *132*, 2463-2474.
Tanaka, T., Yoshida, N., Kishimoto, T., and Akira, S. (1997). Defective adipocyte differentiation in mice lacking the C/EBPbeta and/or C/EBPdelta gene. The EMBO journal *16*, 7432-7443.
Thomas-Chollier, M., Defrance, M., Medina-Rivera, A., Sand, O., Herrmann, C., Thieffry, D., and van Helden, J. (2011). RSAT 2011: regulatory sequence analysis tools. Nucleic acids research *39*, W86-91.
Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Turbendian, H.K., Gordillo, M., Tsai, S.Y., Lu, J., Kang, G., Liu, T.C., Tang, A., Liu, S., Fishman, G.I., and Evans, T. (2013). GATA factors efficiently direct cardiac fate from embryonic stem cells. Development *140*, 1639-1644.
Watt, A.J., Battle, M.A., Li, J., and Duncan, S.A. (2004). GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. Proceedings of the National Academy of Sciences of the United States of America *101*, 12573-12578.
Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.
Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307-319.
Wirth, T., Pfisterer, P., Annweiler, A., Zwilling, S., and Konig, H. (1995). Molecular principles of Oct2-mediated gene activation in B cells. Immunobiology *193*, 161-170.
Xie, W., and Ren, B. (2013). Developmental biology. Enhancing pluripotency and lineage specification. Science *341*, 245-247.
Yajima, H., Motohashi, N., Ono, Y., Sato, S., Ikeda, K., Masuda, S., Yada, E., Kanesaki, H., Miyagoe-Suzuki, Y., Takeda, S.*, et al.* (2010). Six family genes control the proliferation and differentiation of muscle satellite cells. Experimental cell research *316*, 2932-2944.
Zhang, H.M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., and Guo, A.Y. (2012). AnimalTFDB: a comprehensive animal transcription factor database. Nucleic acids research *40*, D144-149.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome biology *9*, R137.

Supplemental Figure1

Supplemental Figure 1

A



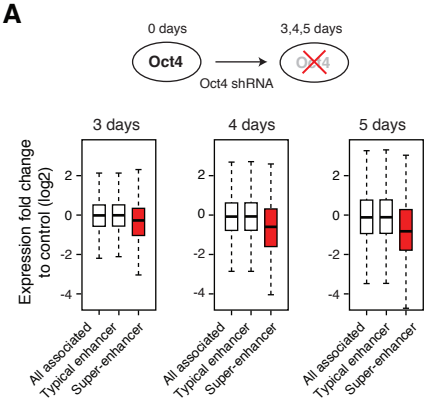
B

Transcription factor	Motif	P-value
Oct4		0
Sox2		0
Nanog		$1.75*10^{-55}$
Klf4		1
Esrrb		$6.27*10^{-60}$
Nr5a2	n.a.	n.a.
Prdm14	n.a.	n.a.
Tcfcp2l1		$6.54*10^{-4}$
Smad3		$3.24*10^{-10}$
Stat3		$3.62*10^{-29}$
Tcf3		$2.95*10^{-159}$

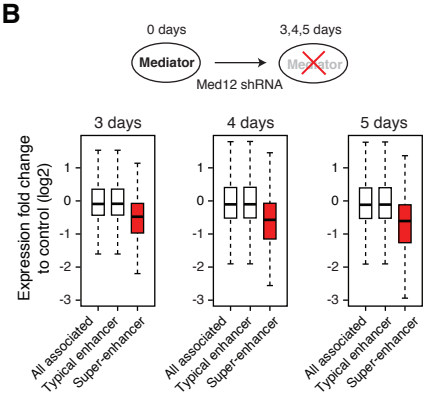
Supplemental Figure2

Supplemental Figure 2

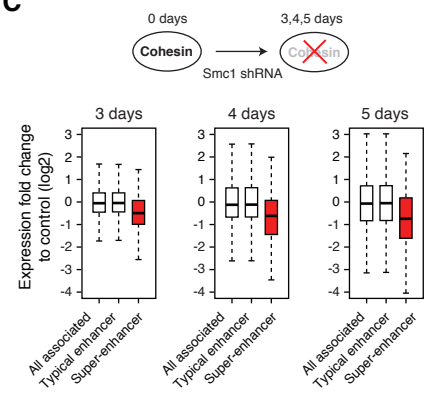
A



B

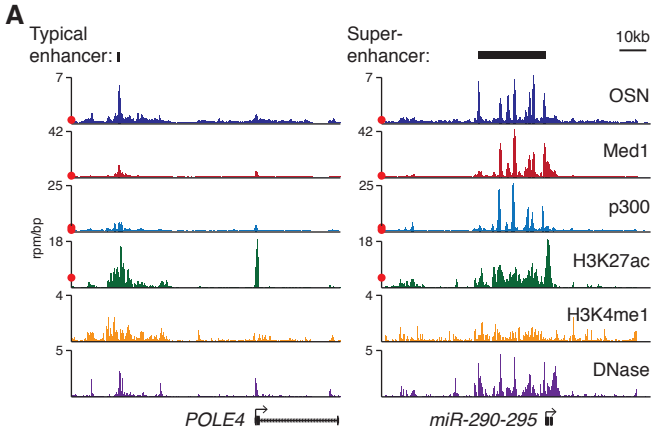


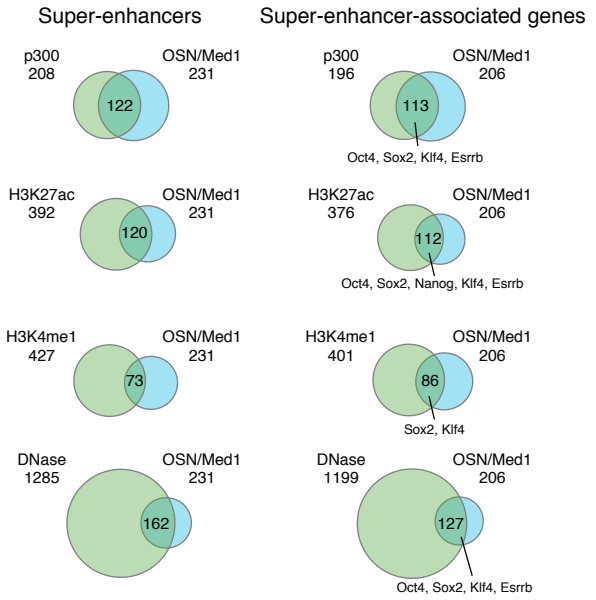
C

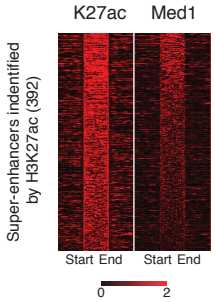


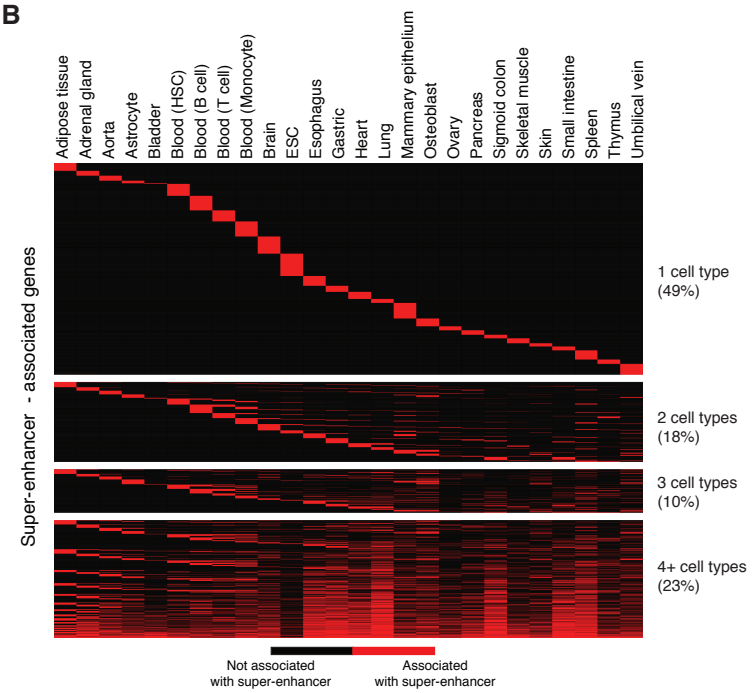
Supplemental Figure3

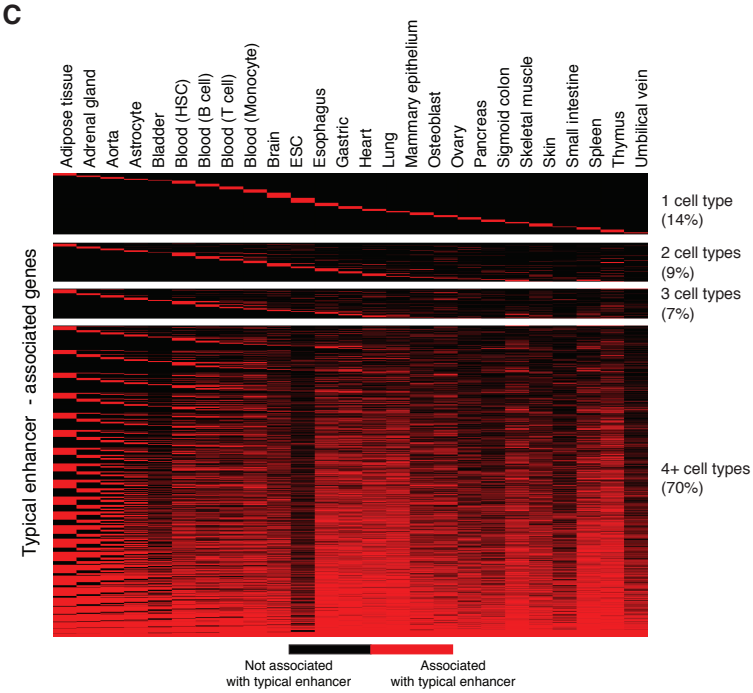
Supplemental Figure 3









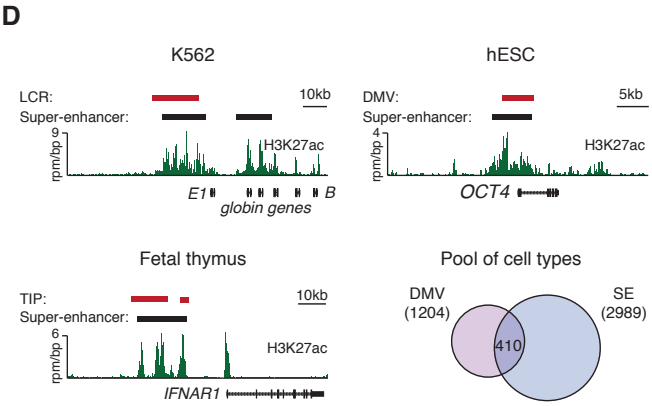


Genes associated with super-enhancers identified by H3K27ac but not OSN/Med1

Biological Process	P-value
Regulation of transcription	$1.2*10^{-8}$
Transcription	$8.9*10^{-7}$
Positive regulation of nitrogen compound metabolic process	$9.3*10^{-7}$
Positive regulation of nucleobase metabolic process	$2.0*10^{-6}$
Positive regulation of transcription	$2.4*10^{-6}$
Positive regulation of gene expression	$3.7*10^{-6}$
Regulation of transcription from RNAPII promoter	$4.0*10^{-6}$
Negative regulation of macromolecule biosynthetic process	$5.0*10^{-6}$
Positive regulation of cellular biosynthetic process	$7.3*10^{-6}$
Negative regulation of cellular biosynthetic process	$7.6*10^{-6}$

Genes associated with super-enhancers identified by OSN/Med1 but not H3K27ac

Biological Process	P-value
Regulation of cell proliferation	$2.6*10^{-6}$
Cellular component morphogenesis	$3.9*10^{-4}$
Positive regulation of cell proliferation	$5.8*10^{-4}$
Cell morphogenesis involved in differentiation	$7.1*10^{-4}$
Neural tube development	$8.3*10^{-4}$
Cell morphogenesis	$9.5*10^{-4}$
Vasculature development	$1.7*10^{-3}$
Tube morphogenesis	$1.7*10^{-3}$
Tube development	$2.2*10^{-3}$
Embryonic morphogenesis	$2.2*10^{-3}$



Supplemental Figure 4

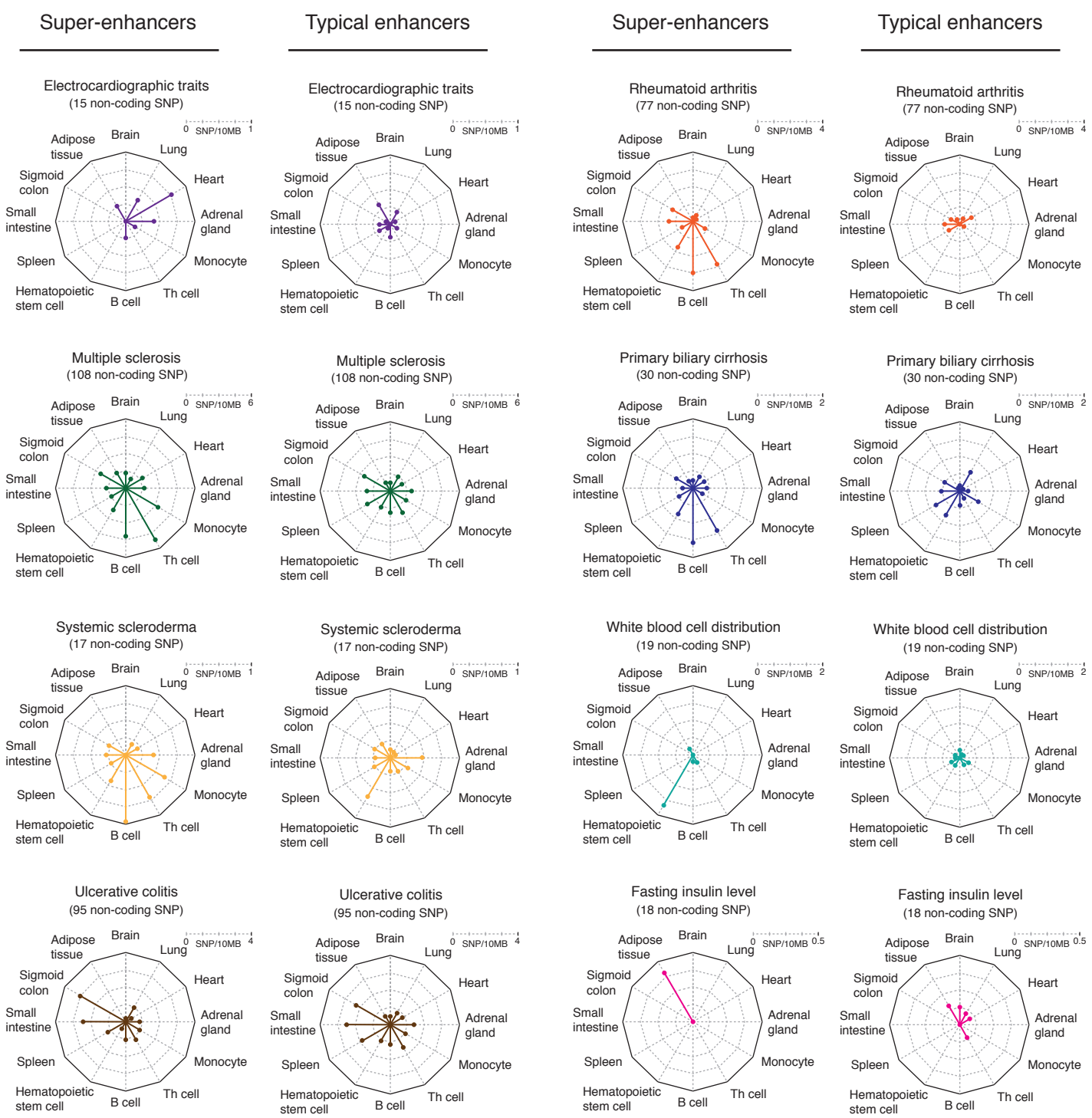
A

	Super-enhancers	Typical enhancers	H3K27ac peaks
Trait-associated SNPs	35%	60%	49%
% genome contained	14%	31%	22%
SNP enrichment	2.5x	1.9x	2.2x

B

SNP enrichment for listed traits:	Super-enhancers	Typical enhancers
Multiple sclerosis	4.6x	2.6x
Celiac disease	4.3x	2.5x
Type 1 diabetes	4.1x	2.3x
Systemic lupus ertythomatosus	4.1x	2.3x
Inflammatory bowel disease	3.6x	2.5x
Crohn´s disease	3.3x	2.1x
HDL cholesterol	3.3x	2.6x
Coronary heart disease	3.1x	2.2x
Rheumatoid arthritis	3.1x	2.3x

Supplemental Figure 5



Supplemental Figure6

Supplemental Figure 6