Cell

Resource

Super-Enhancers in the Control of Cell Identity and Disease

Denes Hnisz,[1,3] Brian J. Abraham,[1,3] Tong Ihn Lee,[1,3] Ashley Lau,[1,2] Violaine Saint-André,[1] Alla A. Sigova,[1] Heather A. Hoke,[1,2] and Richard A. Young[1,2,*]

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]These authors contributed equally to this work

*Correspondence: young@wi.mit.edu

http://dx.doi.org/10.1016/j.cell.2013.09.053

SUMMARY

Super-enhancers are large clusters of transcriptional enhancers that drive expression of genes that define cell identity. Improved understanding of the roles that super-enhancers play in biology would be afforded by knowing the constellation of factors that constitute these domains and by identifying super-enhancers across the spectrum of human cell types. We describe here the population of transcription factors, cofactors, chromatin regulators, and transcription apparatus occupying super-enhancers in embryonic stem cells and evidence that super-enhancers are highly transcribed. We produce a catalog of super-enhancers in a broad range of human cell types and find that super-enhancers associate with genes that control and define the biology of these cells. Interestingly, disease-associated variation is especially enriched in the super-enhancers of disease-relevant cell types. Furthermore, we find that cancer cells generate super-enhancers at oncogenes and other genes important in tumor pathogenesis. Thus, super-enhancers play key roles in human cell identity in health and in disease.

INTRODUCTION

Transcription factors bind DNA regulatory elements called enhancers, which play key roles in the control of cell-type-specific gene expression programs (Bulger and Groudine, 2011; Calo and Wysocka, 2013; Carey, 1998; Lelli et al., 2012; Levine and Tjian, 2003; Maston et al., 2006; Ong and Corces, 2011; Panne, 2008; Spitz and Furlong, 2012; Xie and Ren, 2013). A typical mammalian cell contains thousands of active enhancers, and it has been estimated that there may be ~1 million enhancers active in all human cells (Bernstein et al., 2012; Heintzman et al., 2009; Thurman et al., 2012). It is important to further understand enhancers and their components because they control specific gene expression programs, and much disease-associated sequence variation occurs in these regulatory elements (Grossman et al., 2013; Hindorff et al., 2009; Lee and Young, 2013; Maurano et al., 2012).

The set of enhancers that control any one cell's gene expression program is probably best defined in murine embryonic stem cells (ESCs). Co-occupancy of murine ESC genomic sites by the master transcription factors Oct4, Sox2, and Nanog is highly predictive of enhancer activity (Chen et al., 2008), and 8,794 enhancers have been identified in ESCs by using ChIP-seq data sets for Oct4, Sox2, and Nanog (Whyte et al., 2013). A subset of these enhancers forms 231 unusual enhancer domains at most genes that control the pluripotent state; these super-enhancers consist of clusters of enhancers that are densely occupied by five key ESC transcription factors and the Mediator coactivator (Whyte et al., 2013). There are many additional transcription factors, cofactors, and chromatin regulators that contribute to the control of ESCs (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011), and it would be instructive to know how these occupy enhancers and super-enhancers in ESCs. Similarly, it would be useful to know if super-enhancers are transcribed, as enhancer RNAs (eRNAs) have been proposed to contribute to enhancer activity (Lai et al., 2013; Lam et al., 2013; Li et al., 2013; Ling et al., 2004; Mousavi et al., 2013; Ørom et al., 2010).

Super-enhancers are associated with key genes that control cell state in cells where they have been identified thus far, so identification of these domains in additional cell types could provide a valuable resource for further study of cellular control. We have generated a catalog of super-enhancers in 86 human cell and tissue types. These super-enhancers are associated with genes encoding cell-type-specific transcription factors and thus identify candidate master transcription factors for many cell types that should prove useful for further understanding transcriptional control of cell state and for reprogramming studies. Using this catalog, we find that DNA sequence variation associated with specific diseases is especially enriched in the super-enhancers of disease-relevant cells, suggesting that hypotheses regarding the role of specific cell types and genes in many diseases might be guided by knowledge of super-enhancers. Furthermore, tumor cells acquire super-enhancers at key oncogenes and at genes that function in the acquisition of hallmark capabilities in cancer, suggesting that these domains provide biomarkers for tumor-specific pathologies that may be valuable for diagnosis and therapeutic intervention.

We discuss the implications of these observations for future study of disease.

RESULTS

Transcription Factors in ESCs

Super-enhancers are clusters of enhancers—formed by binding of high levels of master transcription factors and Mediator coactivator—that drive high-level expression of genes encoding key regulators of cell identity (Figure 1A) (Whyte et al., 2013). Five ESC transcription factors were previously shown to occupy super-enhancers (Oct4, Sox2, Nanog, Klf4, and Esrrb) (Whyte et al., 2013), but there are many additional transcription factors that contribute to the control of ESCs (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011). We compiled ChIP-seq data for 15 additional transcription factors in ESCs, for which high-quality ChIP-seq data were available, and investigated whether they occupy enhancers defined by Oct4, Sox2, and Nanog (OSN) co-occupancy (Whyte et al., 2013) (Table S1 available online). The analysis showed that six additional transcription factors (Nr5a2, Prdm14, Tcfcp2l1, Smad3, Stat3, and Tcf3) occupy both typical enhancers and super-enhancers and that all of these are enriched in super-enhancers (Figures 1B–1E). Each of these factors has previously been shown to play important roles in ESC biology (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011). In contrast, nine other transcription factors (c-Myc, CTCF, Zfx, Tbx3, YY1, Tfe3, Kap1/Zfp57, Smad1, and Ronin) were not similarly enriched in enhancers (Table S1) and instead occupied other regions of the genome such as promoter-proximal sites or sites that border topological domains (Figure S1A). It is particularly interesting that Smad3, Stat3, and Tcf3 are enriched in super-enhancer domains because these are transcription factor targets of the TGF-β-, LIF-, and Wnt-signaling pathways, respectively. Previous studies have shown that these transcription factors are recruited to enhancers formed by master transcription factors (Chen et al., 2008; Mullen et al., 2011), and evidence for enrichment of these factors at super-enhancers shows how these signaling pathways can converge on key genes that control ESC identity.

To assess whether the 11 transcription factors that are enriched at super-enhancers contribute to super-enhancer formation by binding to known DNA sequence motifs, we analyzed the frequency of these binding motifs at super-enhancer regions. For all nine transcription factors for which binding motifs are available, we found that the cognate motif showed significant enrichment at super-enhancer constituents relative to background expectation, and super-enhancers were enriched for these motifs compared to typical enhancers (Figures 1F, S1B, and S1C). These results suggest that the nine transcription factors contribute to super-enhancers by binding directly to their known DNA sequence motifs.

Previous studies have described a model of core transcriptional regulatory circuitry that includes Oct4, Sox2, and Nanog (Boyer et al., 2005). The evidence that these and additional ESC transcription factors form super-enhancers that drive genes that are essential for control of cell identity suggests a revised model of transcriptional regulatory circuitry for ESCs (Figure 1G). This model contains an interconnected autoregulatory loop like that originally proposed for Oct4, Sox2, and Nanog (Boyer et al., 2005) but consists of the additional ESC transcription factors that meet three criteria: (1) their genes are driven by super-enhancers, (2) they co-occupy their own super-enhancers as well as those of the other master genes, and (3) they play important roles in regulation of ESC state and iPSC reprogramming.

RNA Polymerase II, Cofactors, and Chromatin Regulators

Super-enhancers are occupied by unusually high levels of the Mediator coactivator (Whyte et al., 2013). Previous studies have described the activities of RNA polymerase II and various cofactors, chromatin regulators, and RNA at specific enhancers (Calo and Wysocka, 2013; Kagey et al., 2010; Lai et al., 2013; Lam et al., 2013; Li et al., 2013; Ling et al., 2004; Mousavi et al., 2013; Natoli and Andrau, 2012; Ong and Corces, 2011; Ørom et al., 2010), so we used published and newly generated ChIP-seq and RNA-seq data to investigate how these components are associated with enhancers and super-enhancers across the ESC genome. The results indicate that RNA polymerase II, Mediator, cohesin, Nipbl, p300, CBP, Chd7, Brd4, and components of the esBAF (Brg1) and Lsd1-NuRD complexes are all enriched in super-enhancers relative to typical enhancers (Figures 2A–2E and Table S1). RNA polymerase II can transcribe enhancers, producing noncoding RNAs that in some cases contribute to enhancer activity (Kim et al., 2010; Lam et al., 2013; Li et al., 2013; Natoli and Andrau, 2012; Sigova et al., 2013); we found that RNA polymerase II and RNA were highly enriched at super-enhancers relative to typical enhancers (Figure 2C).

It was notable that a broad spectrum of cofactors and chromatin regulators that are responsible for gene activation, enhancer looping, histone modification, and nucleosome remodeling are especially enriched in ESC super-enhancers. The Mediator coactivator binds Nipbl, which loads cohesin, thus facilitating looping of enhancers to the promoters of their target genes (Kagey et al., 2010). The coactivator p300 is a histone acetyltransferase, which is generally found at enhancer regions (Heintzman et al., 2007; Visel et al., 2009). CBP is a transcriptional coactivator that interacts with p300 and promotes synergy between enhancer components (Merika et al., 1998). Chd7 is a chromatin remodeler that also interacts with p300 and is often found at enhancers (Schnetz et al., 2010). Brd4, a member of the bromodomain protein family, binds to Mediator and acetylated histones and is involved in regulation of transcriptional elongation by RNA polymerase II (Jang et al., 2005; Jiang et al., 1998). Brg1 is a subunit of the mammalian esBAF (SWI/SNF) complex, an ATP-dependent chromatin remodeler, which contributes to maintenance of pluripotency and self-renewal in ESCs (Ho et al., 2009a; Ho et al., 2009b). Lsd1, Hdac1, Hdac2, Mi-2b, and Mbd3 are subunits of the Lsd1-NuRD complex, which possesses histone deacetylase-, demethylase-, and nucleosome-remodeling activities and contributes to enhancer decommissioning during differentiation (Denslow and Wade, 2007; Foster et al., 2010; Kaji et al., 2006, 2007; Reynolds et al., 2012a, 2012b; Shi et al., 2004; Whyte et al., 2012).

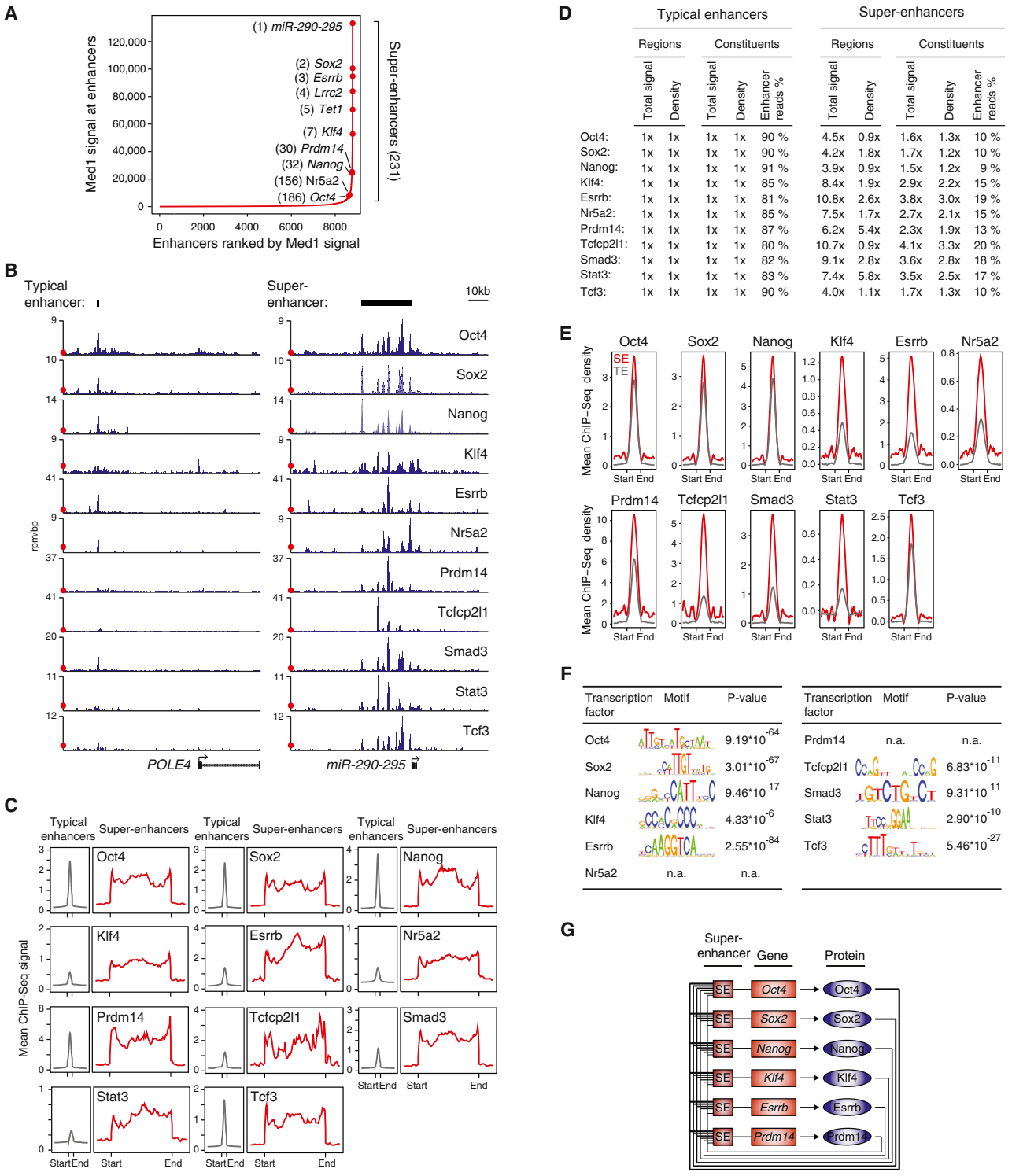


Figure 1. Transcription Factors at Super-Enhancers

(A) Distribution of Med1 ChIP-seq signal at enhancers reveals two classes of enhancers in ESCs. Enhancer regions are plotted in an increasing order based on their input-normalized Med1 ChIP-seq signal. Super-enhancers are defined as the population of enhancers above the inflection point of the curve. Example super-enhancers are highlighted along with their respective ranks and their associated genes.

(B) ChIP-seq binding profiles for the indicated transcription factors at the *POLE4* and *miR-290–295* loci in ESCs. Red dots indicate the median enrichment of all bound regions in the respective ChIP-seq data sets and are positioned at maximum 20% of the axis height. rpm/bp, reads per million per base pair.

(legend continued on next page)

Super-enhancers are occupied by an unusually large portion of the enhancer-associated RNA polymerase II and its associated cofactors and chromatin regulators. As measured by ChIP-seq reads, between 12% and 36% of RNA polymerase II and the cofactors associated with all 8,794 enhancers were found within the 231 super-enhancers (Figure 2C). The evidence that a large fraction of these enhancer cofactors are associated with super-enhancers helps to explain why these large domains produce relatively high levels of RNA (Figure 2C) and drive high-level expression of their associated genes when compared to typical enhancers (Whyte et al., 2013). The presence of high levels of RNA at super-enhancers is especially interesting in light of recent studies suggesting that enhancer RNA contributes to gene activation (Lai et al., 2013; Lam et al., 2013; Li et al., 2013; Ling et al., 2004; Mousavi et al., 2013; Ørom et al., 2010) and evidence that the *MYOD1* super-enhancer is transcribed into eRNAs that contribute to the transcriptional activation of *MYOD1* in muscle cells (Mousavi et al., 2013).

ESC differentiation causes preferential loss of expression of super-enhancer-associated genes, which may be a consequence of the unusual vulnerability of super-enhancers to perturbation of their components (Dowen et al, 2013; Lovén et al., 2013; Whyte et al., 2013) (Figures S2A–S2C). We speculate that this dual feature of super-enhancers—their ability to drive high-level expression of key regulators of cell identity and their vulnerability to perturbation of their components—may facilitate cell state transitions during development.

Super-Enhancers in Many Cell Types

Because super-enhancers drive expression of genes that control and define cell identity, it would be useful to identify these elements and their associated genes in all human cells. However, the master transcription factors that might form super-enhancers are not known for most cell types, and genome-wide binding data are limited for those that are known. We therefore explored the ability of various surrogate marks of enhancers (histone H3K27ac, H3K4me1, DNase hypersensitivity, and p300) to identify super-enhancers in ESCs (Creyghton et al., 2010; Heintzman et al., 2007; Neph et al., 2012; Rada-Iglesias et al., 2011; Shen et al., 2012; Visel et al., 2009). Of the marks available for a broad range of human samples, the histone H3K27ac modification was superior to the others in that it identified a large fraction of OSN-Mediator super-enhancers while minimizing excess sites (Figure S3A).

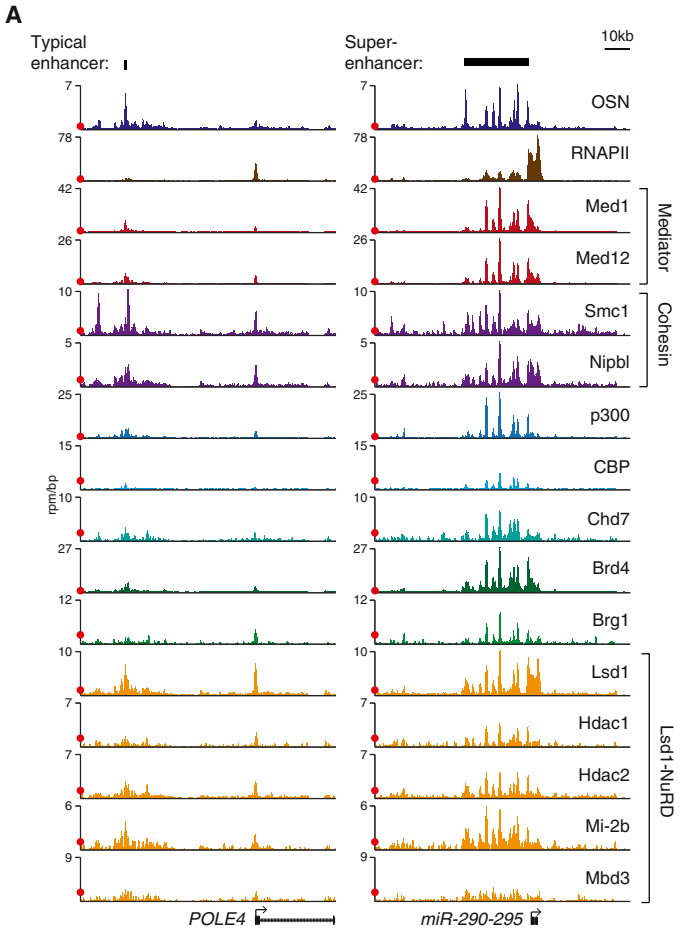
We used H3K27ac ChIP-seq data to create a catalog of super-enhancers for 86 human cell and tissue samples (Figure 3 and Table S2). A substantial portion of these super-enhancers and their associated genes are cell type specific (Figures 3A and S3B). In contrast, typical enhancer-associated genes are less cell type specific (Figure S3C). Characterization of super-enhancer-associated genes by Gene Ontology analysis revealed that they are linked to biological processes that largely define the identities of the respective cell and tissue types (Figure 3B). Some of the super-enhancer domains overlap previously described locus control regions (LCRs), transcription initiation platforms (TIPs), and DNA methylation valleys (DMV) (Figure S3D) (Bonifer, 2000; Forrester et al., 1990; Grosveld et al., 1987; Koch et al., 2011; Tuan et al., 1985; Xie and Ren, 2013).

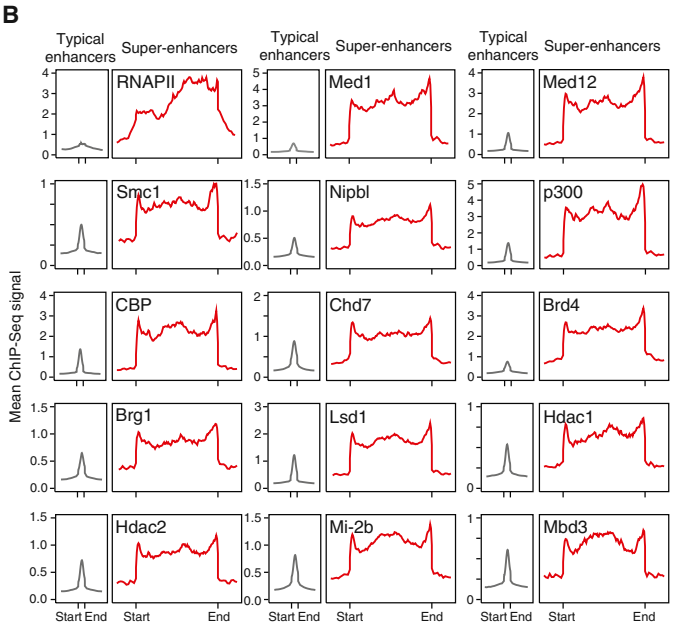
To gain further understanding of the transcriptional regulatory circuitry of cells and to facilitate efforts to reprogram cells for regenerative medicine, it would be valuable to identify the master transcription factors that control all cell states (Cherry and Daley, 2012; Graf and Enver, 2009; Lee and Young, 2013; Zhou et al., 2008). Super-enhancers were previously identified in five murine cell types (ESC, myotubes, pro-B cells, Th cells, and macrophages), and the genes encoding known master transcription factors in these cells were found to have associated super-enhancers (Whyte et al., 2013). We reasoned that candidate master transcription factors could be identified in most cells by identifying genes associated with super-enhancers that encode transcription factors and carried out this analysis in all of the cells in this study. For those cells where master transcription factors have already been identified, this exercise captured the vast majority of these factors (Figure 3C). A catalog of candidate master transcription factors for other cell types can be found in Table S3. Prior studies of key transcriptional regulators suggest that these candidates should be useful for deducing the transcriptional regulatory circuitry of many different human cells and for reprogramming studies.

Disease-Associated DNA Sequence Variation in Super-Enhancers

Several recent studies suggest that much of disease-associated DNA sequence variation occurs in transcriptional regulatory regions defined by DNase hypersensitivity (Maurano et al., 2012;

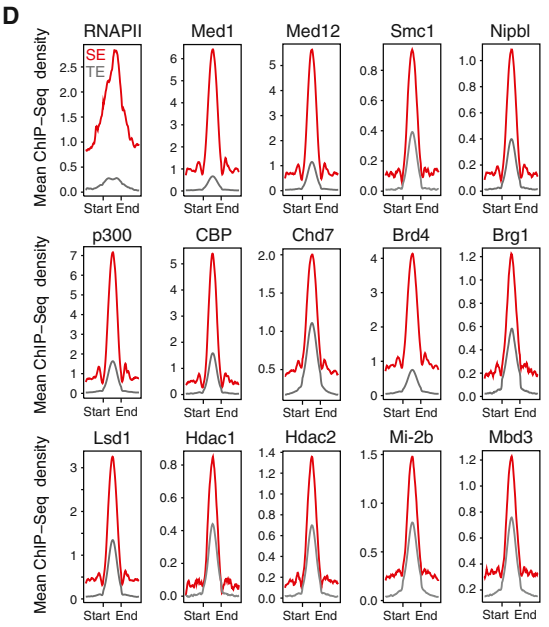
(C) Metagene representations of the mean ChIP-seq signal for the indicated transcription factors across typical enhancers and super-enhancer domains. Metagenes are centered on the enhancer region, and the length of the enhancer reflects the difference in median lengths (703 bp for typical enhancers, 8,667 bp for super-enhancers). Additional 3 kb surrounding each enhancer region is also shown.
(D) Fold difference values of ChIP-seq signal between typical enhancers and super-enhancers for the indicated transcription factors. Total signal indicates the mean ChIP-seq signal (total reads) at typical enhancers and super-enhancers normalized to the mean value at typical enhancers. Density indicates the mean ChIP-seq density at constituent enhancers (rpm/bp) of typical enhancers and super-enhancers normalized to the mean value at typical enhancers. Enhancer read % indicates the percentage of all reads mapped to enhancer regions that fall in the constituents of typical enhancer or super-enhancer regions.
(E) Metagene representations of the mean ChIP-seq density for the indicated transcription factors across the constituent enhancers within typical enhancers and super-enhancers. Each metagene is centered on enhancer constituents. Additional 2.5 kb surrounding the constituent enhancer regions is also shown.
(F) Table depicting transcription factor binding motifs enriched at constituent enhancers within super-enhancer regions and associated p values.
(G) Revised model of the core transcriptional regulatory circuitry of ESCs. The model contains an interconnected autoregulatory loop consisting of transcription factors that meet three criteria: (1) their genes are driven by super-enhancers, (2) they co-occupy their own super-enhancers as well as those of the other transcription factor genes in the circuit, and (3) they play essential roles in regulation of ESC state and iPSC reprogramming. The layout of the circuit model was adapted from Whyte et al. (2013).
See also Figure S1 and Table S1.





C

	Typical enhancers					Super-enhancers				
	Regions		Constituents			Regions		Constituents		
	Total signal	Density	Total signal	Density	Enhancer reads %	Total signal	Density	Total signal	Density	Enhancer reads %
RNAPII:	1x	1x	1x	1x	68 %	28.8x	8.1x	8.2x	7.1x	32 %
Med1:	1x	1x	1x	1x	64 %	28.2x	6.8x	10.3x	8.1x	36 %
Med12:	1x	1x	1x	1x	75 %	16.0x	3.5x	5.8x	4.5x	25 %
Smc1:	1x	1x	1x	1x	86 %	8.4x	2.2x	2.9x	2.3x	14 %
Nipbl:	1x	1x	1x	1x	80 %	9.4x	2.1x	3.2x	2.5x	15 %
p300:	1x	1x	1x	1x	86 %	13.3x	0.8x	2.0x	4.2x	14 %
CBP:	1x	1x	1x	1x	80 %	10.7x	2.4x	4.0x	3.1x	20 %
Chd7:	1x	1x	1x	1x	87 %	7.6x	1.5x	2.3x	1.8x	13 %
Brd4:	1x	1x	1x	1x	74 %	19.7x	4.2x	6.2x	5.0x	26 %
Brg1:	1x	1x	1x	1x	85 %	8.6x	1.5x	2.7x	2.0x	15 %
Lsd1:	1x	1x	1x	1x	85 %	9.0x	1.9x	2.9x	2.3x	15 %
Hdac1:	1x	1x	1x	1x	88 %	6.5x	1.3x	2.1x	1.6x	12 %
Hdac2:	1x	1x	1x	1x	87 %	6.9x	1.4x	2.2x	1.7x	13 %
Mi-2b:	1x	1x	1x	1x	88 %	7.4x	1.5x	2.2x	1.7x	12 %
Mbd3:	1x	1x	1x	1x	88 %	8.3x	1.5x	2.1x	1.7x	12 %
RNA:	1x	1x	1x	1x	74 %	24.3x	6.0x	5.4x	4.6x	26 %



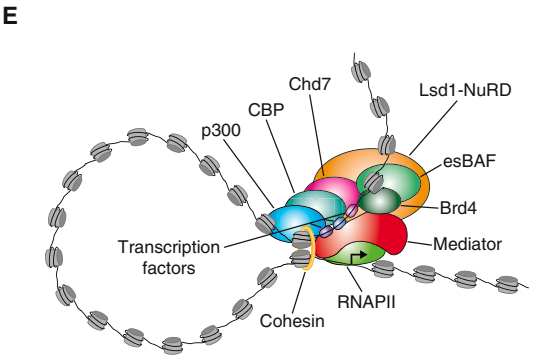


Figure 2. RNA Polymerase II, Cofactors, and Chromatin Regulators at Super-Enhancers

(A) ChIP-seq binding profiles for RNA polymerase II (RNAPII) and the indicated transcriptional cofactors and chromatin regulators at the *POLE4* and *miR-290-295* loci in ESCs. OSN denotes the merged ChIP-seq binding profiles of the transcription factors Oct4, Sox2, and Nanog and serves as a reference. Red dots indicate

(legend continued on next page)

Vernot et al., 2012). We investigated the extent to which disease-associated DNA sequence variation occurs in enhancers and super-enhancers defined by histone H3K27ac. We compiled a list of 5,303 single-nucleotide polymorphisms (SNP) linked to diverse phenotypic traits and diseases in 1,675 genome-wide association studies (GWAS) and investigated their distribution within enhancers and super-enhancers identified in the 86 human cell and tissue samples (Figure 4A and Table S4). We found that the majority of trait-associated SNPs occur in noncoding regions and that 64% of these occur within enhancer regions defined by H3K27ac (Figure 4A). Thus, 64% of trait-associated noncoding SNPs occur in the ~33% of the genome covered by all enhancer regions defined by H3K27ac (permutation test, $p < 10^{-4}$). The trait-associated SNPs were more enriched in super-enhancers than in typical enhancers (χ^2 test, $p < 10^{-12}$) (Figure S4A), and for certain diseases, the enrichment in super-enhancers was particularly striking (Figure S4B). These results confirm that much of disease-associated DNA sequence variation occurs in transcriptional regulatory regions of the genome, indicate that most of this variation occurs in enhancers, and reveal that variation disproportionately impacts super-enhancer domains.

If disease-associated SNPs occur disproportionately in super-enhancer domains, we would expect that SNPs associated with specific diseases would tend to occur in the super-enhancers of disease-relevant cells and not in those of disease-irrelevant cells. Indeed, for a broad spectrum of diseases, we found that disease-associated SNPs tend to occur in the super-enhancers of disease-relevant cells (Figure 4B and Table S4). This relationship was more pronounced for super-enhancers than for typical enhancers (Figure S4C). Because super-enhancers drive the expression of genes that control and define cell identity, these results suggest that altered expression of cell identity genes may often contribute to these diseases.

Examples of Disease-Associated SNPs in Super-Enhancers

We focused further study on several diseases in which SNPs occur in super-enhancers of disease-relevant cell types in order to gain further insights into the relationship between these SNPs, specific super-enhancers, and their associated genes. The diseases that we selected for further study included Alzheimer's disease, type 1 diabetes, and systemic lupus erythematosus (Figure 5).

Alzheimer's disease is a common form of dementia characterized by progressive neurodegeneration and loss of cognitive functions of the brain, and much of the genetic variation implicated in Alzheimer's disease is associated with amyloid precursor protein, transmembrane proteins, and apolipoprotein E4 (Bertram and Tanzi, 2008; Tanzi, 2012). The SNP catalog contains 27 SNPs linked to Alzheimer's disease, and five of these occur in the super-enhancers of brain tissue (Figure 5A). Thus, ~19% (5/27) of all of the Alzheimer's disease SNPs occur in the 1.4% of the genome encompassed by brain tissue super-enhancers (permutation test, $p < 10^{-2}$). Two SNPs occur in the super-enhancer associated with the gene *BIN1* (Figure 5A), whose expression has recently been shown to be associated with Alzheimer's disease risk (Chapuis et al., 2013). Additional variation in the *BIN1* super-enhancer, involving a small insertion, was shown to be associated with Alzheimer's disease (Chapuis et al., 2013).

Type 1 diabetes is caused by T-cell-mediated autoimmune destruction of pancreatic β cells, and much of the genetic variation implicated in type 1 diabetes is associated with major histocompatibility antigens, interleukin-2 signaling, T cell receptor signaling, and interferon signaling (Bluestone et al., 2010; Noble and Erlich, 2012). The SNP catalog contains 76 SNPs linked to type 1 diabetes, and 67 of these occur in noncoding sequences. The noncoding SNPs were particularly enriched in the super-enhancers of primary Th cells, with 13 occurring in the super-enhancer regions of genes with prominent roles in Th cell biology (Figure 5B). It was striking that 19% (13/67) of all of the type 1 diabetes SNPs in noncoding regions occur in the 1.3% of the genome encompassed by Th cell super-enhancers (permutation test, $p < 10^{-4}$).

Systemic lupus erythematosus (SLE) is a systemic autoimmune disease characterized by the loss of tolerance for self-antigens and production of excess amounts of serum autoantibodies. Most genetic variation associated with SLE involves major histocompatibility antigens and lymphocyte-signaling pathways (Costa-Reis and Sullivan, 2013; Deng and Tsao, 2010). The SNP catalog contains 72 SNPs linked to SLE, and 67 of these occur in noncoding regions. Among the cell types examined here, the noncoding SNPs occur most frequently in

the median enrichment of all bound regions in the respective ChIP-seq data sets and are positioned at maximum 20% of the axis height. rpm/bp, reads per million per base pair.

(B) Metagene representations of the mean ChIP-seq signal for RNAPII and the indicated transcriptional cofactors and chromatin regulators across typical enhancers and super-enhancer domains. Metagenes are centered on the enhancer region, and the length of the enhancer reflects the difference in median lengths (703 bp for typical enhancers, 8,667 bp for super-enhancers). Additional 3 kb surrounding each enhancer region is also shown.

(C) Fold difference values of ChIP-seq signal between typical enhancers and super-enhancers for RNAPII and the indicated transcriptional cofactors and chromatin regulators and RNA-seq. Total signal indicates the mean ChIP-seq signal (total reads) at typical enhancers and super-enhancers normalized to the mean value at typical enhancers. Density indicates the mean ChIP-seq density at constituent enhancers (rpm/bp) of typical enhancers and super-enhancers normalized to the mean value at typical enhancers. Enhancer read % indicates the percentage of all reads mapped to enhancer regions that fall in the constituents of typical enhancer or super-enhancer regions. Reads mapped to exons were removed for the RNA-seq analysis.

(D) Metagene representations of the mean ChIP-seq density for RNAPII and the indicated transcriptional cofactors and chromatin regulators across the constituent enhancers within typical enhancers and super-enhancers. Each metagene is centered on enhancer constituents. Additional 2.5 kb surrounding the constituent enhancer regions is also shown.

(E) Model showing RNAPII, transcriptional cofactors, and chromatin regulators that are found in ESC super-enhancers. The indicated proteins are responsible for diverse enhancer-related functions, such as enhancer looping, gene activation, nucleosome remodeling, and histone modification.

See also Figure S2 and Table S1.

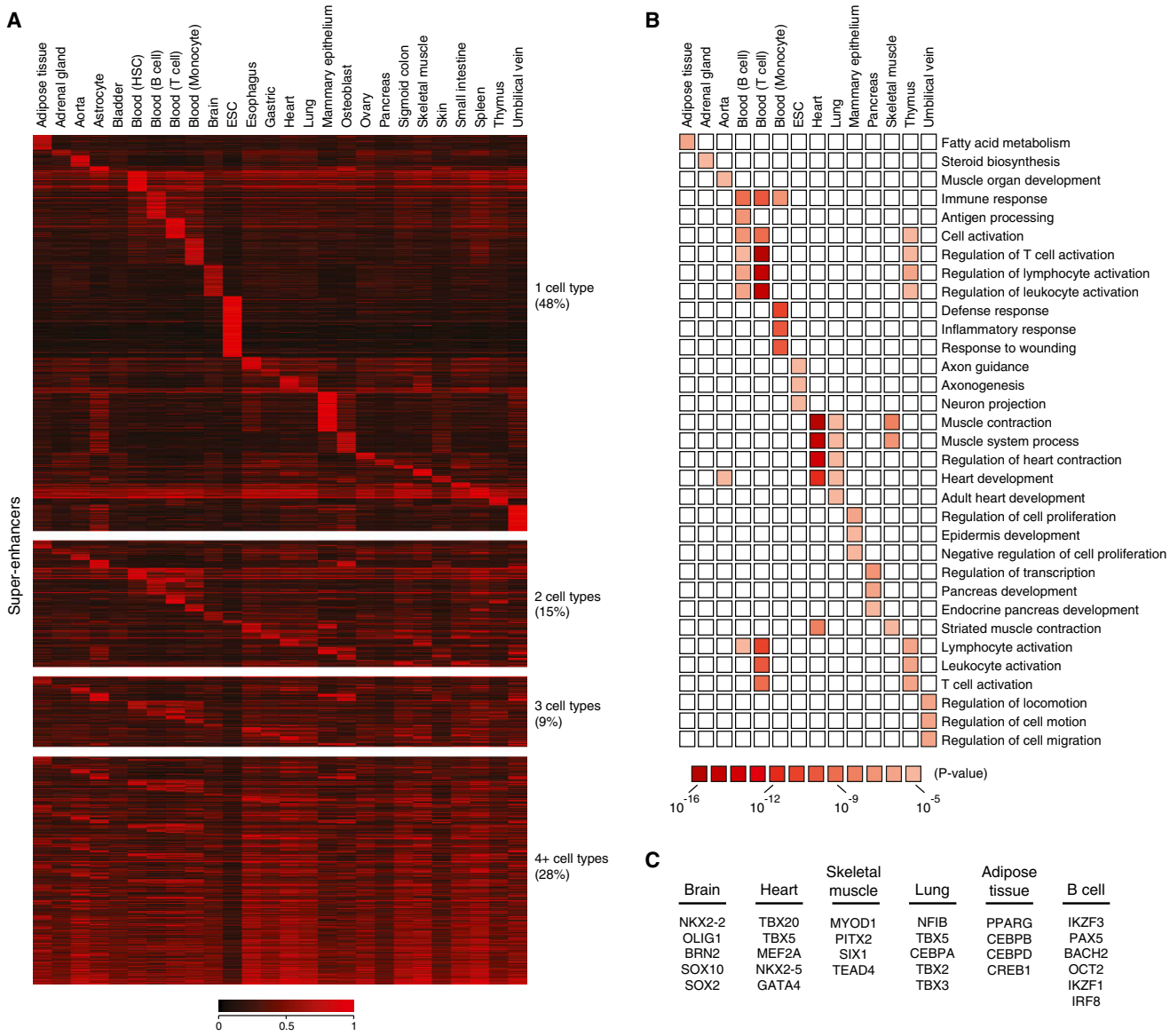


Figure 3. Super-Enhancers and Candidate Master Transcription Factors in Many Cell Types

(A) Heatmap showing the classification of super-enhancer domains across 26 human cell and tissue types. Color scale reflects the density of H3K27ac signal at the super-enhancer regions.

(B) Gene Ontology terms for super-enhancer-associated genes in 14 human cell and tissue types with corresponding p values.

(C) Candidate master transcription factors identified in six cell types. All of these transcription factors were previously demonstrated to play key roles in the biology of the respective cell type or facilitate reprogramming to the respective cell type.

See also Figure S3 and Tables S2 and S3.

B cell super-enhancers, with 22 SNPs occurring in the super-enhancer regions of 16 genes that play key roles in B cell biology (Figure 5C). Thus, 33% (22/67) of the SLE SNPs in noncoding regions occur in the 1.5% of the genome encompassed by B cell super-enhancers (permutation test, $p < 10^{-4}$).

Similar enrichment of disease-associated variation in super-enhancers was observed for many additional diseases, including rheumatoid arthritis, multiple sclerosis, systemic scleroderma, primary biliary cirrhosis, Crohn's disease, Graves disease, vitiligo, and atrial fibrillation (Table S4). This suggests that hypotheses regarding the role of specific cell types and genes in many diseases might be guided by knowledge of super-enhancers.

Super-Enhancers in Cancer

Super-enhancers associate with key oncogenes in several cancer cells (Lovén et al., 2013). To gain further insights into the relationship between super-enhancers and cancer cell states, we

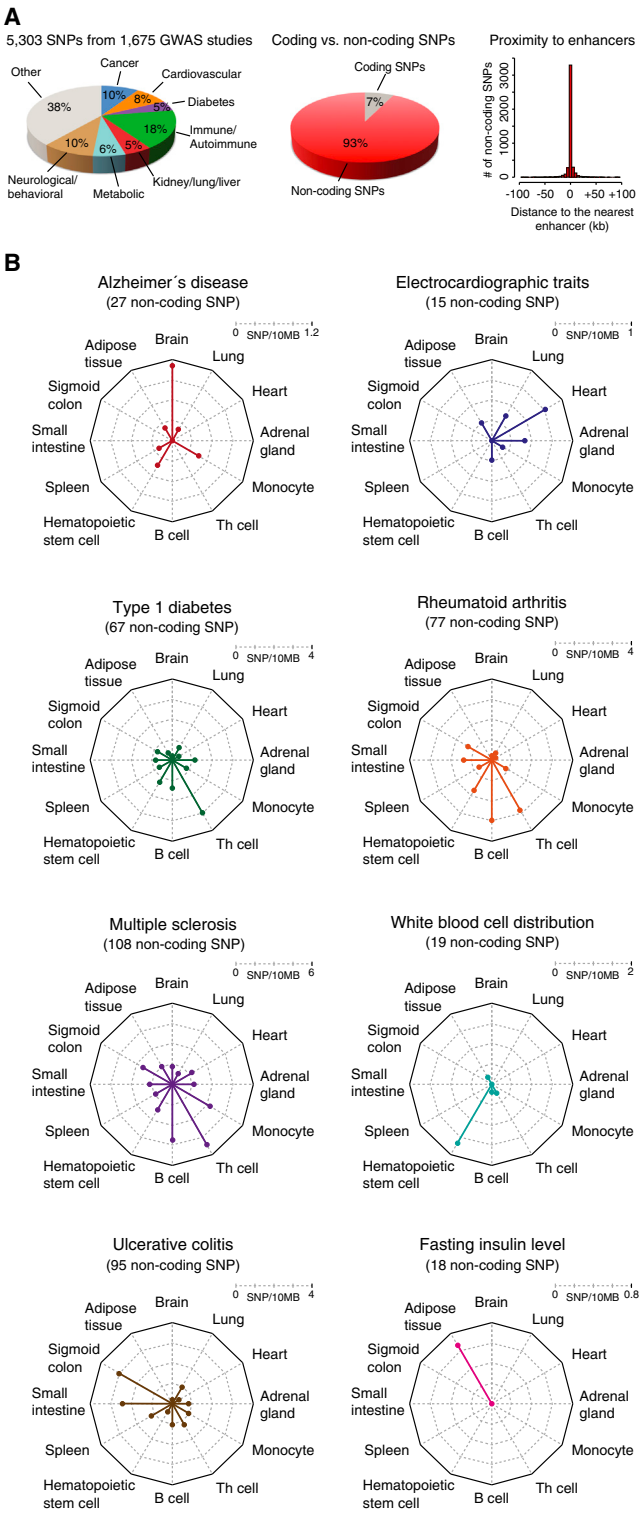


Figure 4. Disease-Associated DNA Sequence Variation in Super-Enhancers

(A) Catalog of single-nucleotide polymorphisms (SNP) linked to phenotypic traits and diseases in genome-wide association studies (GWAS). (Left) Pie chart showing percentage of SNPs associated with the highlighted classes of traits and diseases. (Middle) Distribution of trait-associated SNPs in coding and noncoding regions of the genome. (Right) Location of all noncoding trait-associated SNPs relative to all enhancers identified in 86 human cell and tissue samples. x axis reflects binned distances of each SNP to the nearest enhancer. SNPs located within enhancers are assigned to the 0 bin.

(B) Radar plots showing the density of trait-associated noncoding SNPs linked to the highlighted traits and diseases in the super-enhancer domains 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/10 MB sequence) in the super-enhancer domains of each cell and tissue type. Lines connecting the density values to the origin of the plot are added to improve visualization.

See also Figure S4 and Tables S2 and S4.

used H3K27ac ChIP-seq data to identify super-enhancers in 18 human cancer cells and identified their associated genes (Table S2). The data revealed that a remarkable spectrum of known oncogene drivers have associated super-enhancers in this set of cancer cells (Figure 6A and Table S2). These results suggest that super-enhancers may be useful for identifying key oncogenes in specific cancers.

Further analysis of super-enhancers in tumor cells and related healthy cells suggests that cancer cells acquire super-enhancers at oncogene drivers during the process of tumor pathogenesis (Figure 6B). For example, for multiple cancer cells, large super-enhancers were found in the gene desert surrounding the *c-MYC* gene in the cancer cells, but not in their healthy counterparts (Figure 6B). Furthermore, the super-enhancers formed in the *MYC* locus were tumor type specific (Figures 6B and S5). These results are consistent with the model that cancer cells acquire cancer-specific super-enhancers at key oncogenes that are not present in their healthy counterparts.

DNA translocation, transcription factor overexpression, and focal amplification occur frequently in cancer, and these mechanisms can account for the ability of cancer cells to acquire super-enhancers (Figure 6C). In multiple myeloma, for example, tumor cells often have a translocation that places the 3′ IgH super-enhancer adjacent to the *MYC* gene (Figure 6C). Overexpression of the TAL1 transcription factor in acute lymphoblastic leukemia (T-ALL) is associated with super-enhancer formation at another site in the *MYC* locus (Figure 6C). Focal amplification in lung cancer involves a large super-enhancer that spans the *MYC* gene and its normal regulatory elements (Figure 6C); tandem repeats of DNA segments can lead to the formation of clusters of enhancers.

Hanahan and Weinberg (2011) have proposed that cancer cells acquire a number of hallmark biological capabilities during the multistep process of tumor pathogenesis (Hanahan and Weinberg, 2011). We used these hallmarks as an organizing principle to investigate whether genes that acquire super-enhancers are associated with these biological capabilities in tumor cells. We identified super-enhancers that were acquired by cancer cells (not present in a healthy counterpart) and determined how their associated genes fit into the hallmarks. The results of such analysis with a colorectal cancer line revealed that more than one-third of the super-enhancer genes have functions that are associated with a cancer hallmark (Figures 6D, 6E, and Table S5). A similar analysis of two additional cancer lines confirmed that a large fraction of genes that acquire super-enhancers have hallmark functions (Figure 6E and Table S5).

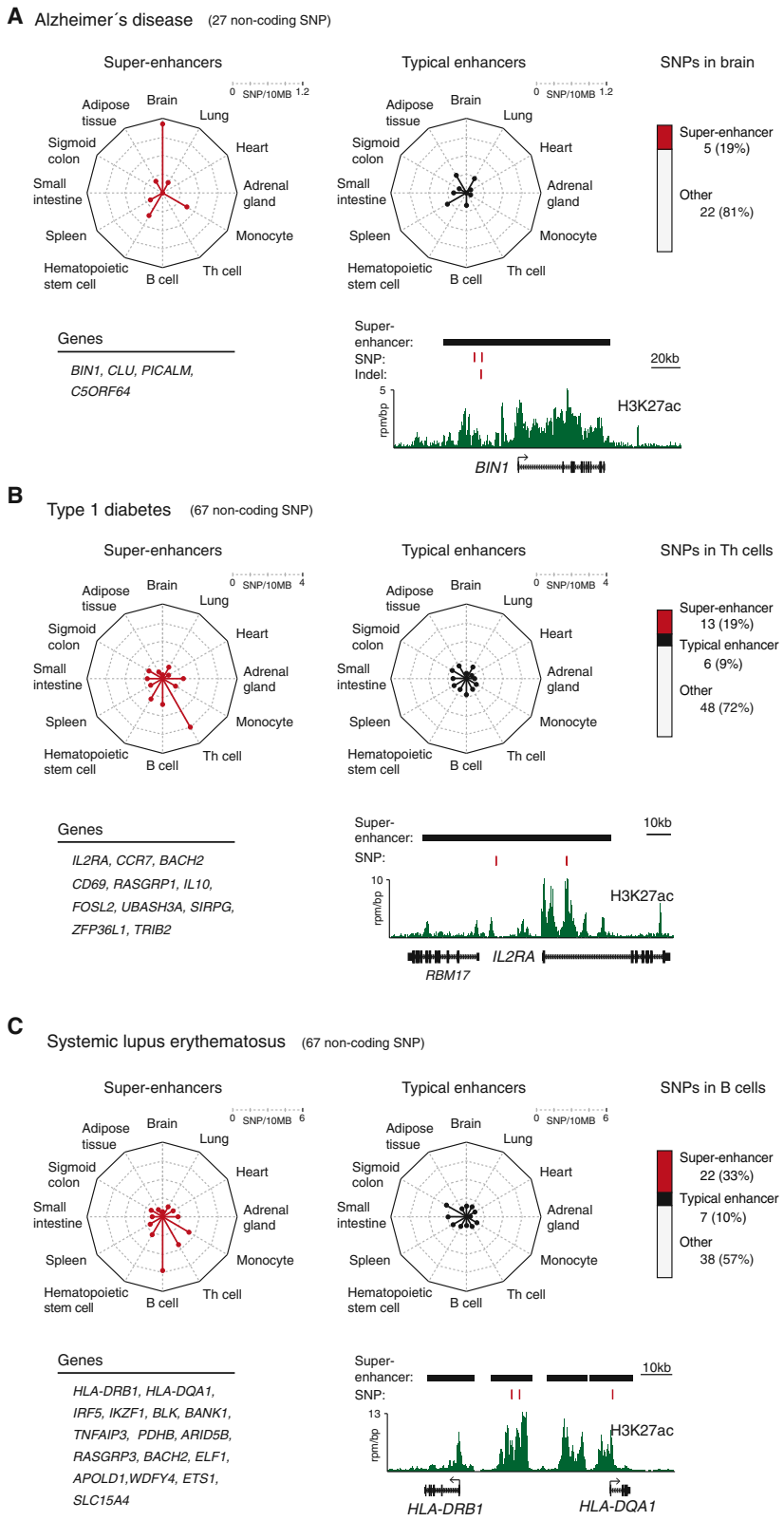


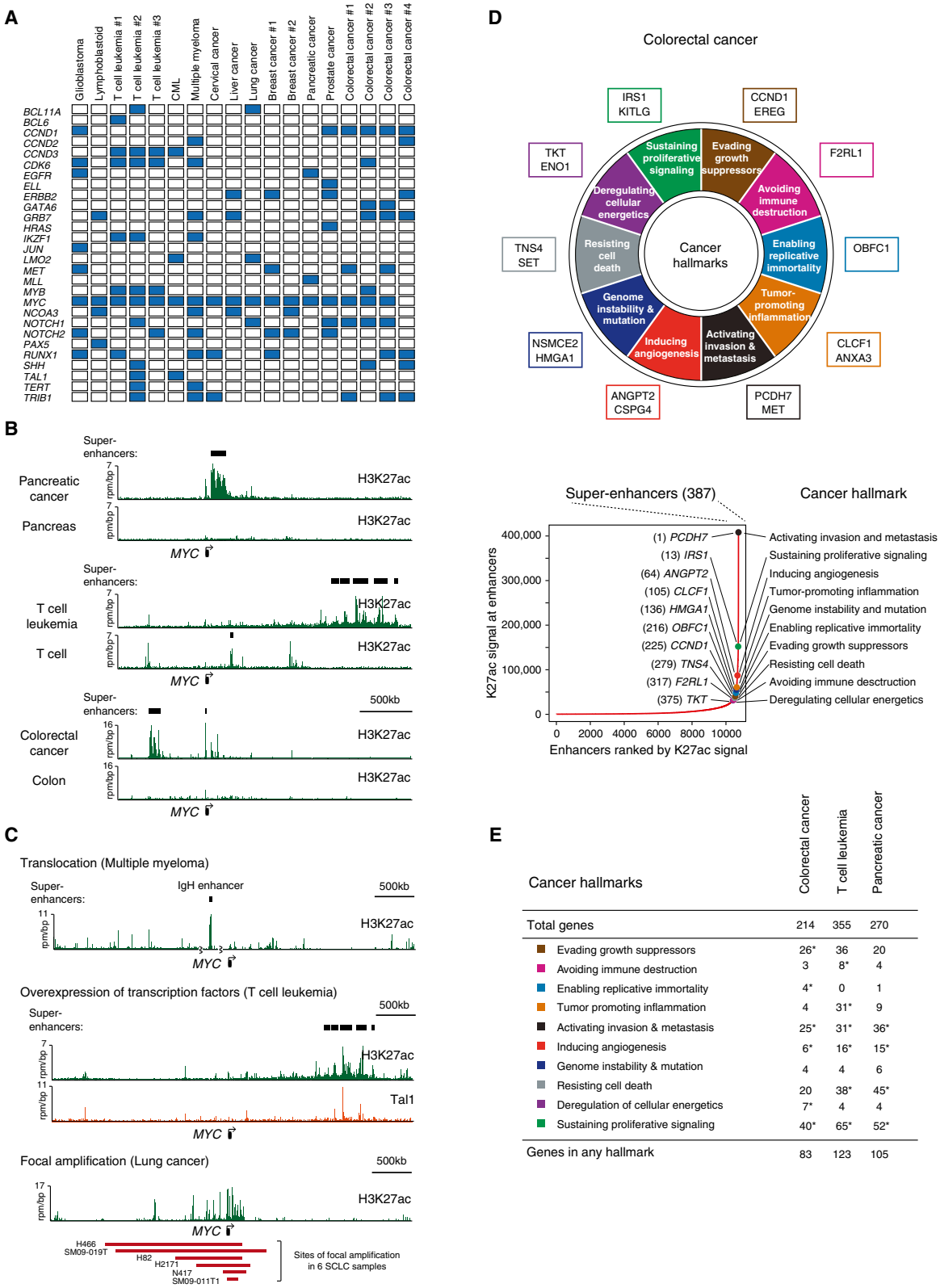
Figure 5. Examples of Disease-Associated SNPs in Super-Enhancers

(A) (Upper-left) Radar plots showing the density of noncoding SNPs linked to Alzheimer's disease (AD) 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/10 MB 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. (Upper-right) Distribution of noncoding SNPs linked to AD in the typical enhancers and super-enhancers of brain tissue. (Lower-left) List of genes associated with AD SNP-containing super-enhancers in brain tissue. (Lower-right) ChIP-seq binding profile for H3K27ac at the *BIN1* locus in brain tissue. The positions of AD SNPs are highlighted as red bars, and the super-enhancers are highlighted as black bars above the binding profile. Indel rs59335482 (a three base pair insertion) is also highlighted. rpm/bp, reads per million per base pair.

(B) (Upper-left) Radar plots showing the density of noncoding SNPs linked to type 1 diabetes (T1D) 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/10 MB 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. (Upper-right) Distribution of noncoding SNPs linked to T1D in the typical enhancers and super-enhancers of Th cells. (Lower-left) List of genes associated with T1D SNP-containing super-enhancers in Th cells. (Lower-right) ChIP-seq binding profile for H3K27ac at the *IL2RA* locus in Th cells. The positions of T1D SNPs are highlighted as red bars, and the super-enhancers are highlighted as black bars above the binding profile.

(C) (Upper-left) Radar plots showing the density of noncoding SNPs linked to systemic lupus erythematosus (SLE) 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/10 MB 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. (Upper-right) Distribution of noncoding SNPs linked to SLE in the typical enhancers and super-enhancers of B cells. (Lower-left) List of genes associated with SLE SNP-containing super-enhancers in B cells. (Lower-right) ChIP-seq binding profile for H3K27ac at the *HLA-DRB1* and *HLA-DQA1* loci in B cells. The positions of SLE SNPs are highlighted as red bars, and the super-enhancers are highlighted as black bars above the binding profile.

See also Tables S2 and S4.



E

Cancer hallmarks	Colorectal cancer	T cell leukemia	Pancreatic cancer
Total genes	214	355	270
Evading growth suppressors	26*	36	20
Avoiding immune destruction	3	8*	4
Enabling replicative immortality	4*	0	1
Tumor promoting inflammation	4	31*	9
Activating invasion & metastasis	25*	31*	36*
Inducing angiogenesis	6*	16*	15*
Genome instability & mutation	4	4	6
Resisting cell death	20	38*	45*
Deregulation of cellular energetics	7*	4	4
Sustaining proliferative signaling	40*	65*	52*
Genes in any hallmark	83	123	105

Figure 6. Super-Enhancers in Cancer

(A) Selected genes associated with super-enhancers in the indicated cancers. Blue box indicates the gene being associated with a super-enhancer in the respective cancer. CML stands for chronic myelogenous leukemia.

(legend continued on next page)

We conclude that cancer cells acquire cancer-specific super-enhancers at genes whose functions are associated with these hallmarks of cancer.

DISCUSSION

Super-enhancers were previously identified in a small number of cells, where they were shown to consist of large clusters of transcriptional enhancers formed by binding of master transcription factors and to be associated with genes that control and define cell identity (Lovén et al., 2013; Whyte et al., 2013). We have extended our understanding of super-enhancers by identifying the population of transcription factors, cofactors, chromatin regulators, and core transcription apparatus that occupy these domains in embryonic stem cells and by demonstrating that super-enhancers are highly transcribed. We have created a catalog of super-enhancers for 86 different human cell and tissue types and have shown that these are associated with genes encoding cell-type-specific transcription factors and other components that play important roles in cell-type-specific biology. Most importantly, we find that sequence variation associated with a broad spectrum of diseases is especially enriched in the super-enhancers of disease-relevant cell types and that cancer cells generally acquire super-enhancers at oncogenes and other genes that play important roles in cancer pathogenesis.

The enhancers and transcription factors that control embryonic stem cell state are probably better understood than those for any other cell type, making ESCs an excellent model for identifying components of super-enhancers (Ng and Surani, 2011; Orkin and Hochedlinger, 2011; Young, 2011). Several important insights were gained by studying how >35 transcription factors, cofactors, chromatin regulators, and components of the core transcription apparatus occupy enhancers and super-enhancers in ESCs. All of the enhancer-binding transcription factors are enriched at super-enhancers, with some so highly enriched that they distinguish super-enhancers from typical enhancers. The transcription factor targets of the TGF-β-, LIF-, and Wnt-signaling pathways are enriched in super-enhancers, suggesting how these signaling pathways converge on key genes that control ESC identity. Super-enhancers are occupied by a large portion of the enhancer-associated RNA polymerase II and its associated cofactors and chromatin regulators, which can explain how they contribute to high-level transcription of associated genes. Furthermore, the levels of RNA detected at super-enhancers vastly exceed those at typical enhancers, and recent evidence suggests that these eRNAs may contribute to gene activation (Lai et al., 2013; Lam et al., 2013; Li et al., 2013; Ling et al., 2004; Mousavi et al., 2013; Ørom et al., 2010).

We have generated a catalog of super-enhancers and their associated genes in a broad spectrum of human cell and tissue types. The super-enhancers tend to be cell type specific, and the genes associated with these elements tend to be cell type specific in their expression and linked to biological processes that largely define the identities of the respective cell and tissue types. Genes that encode candidate master transcription factors and noncoding RNAs such as miRNAs are among those associated with super-enhancers. Thus, the super-enhancer catalog should provide a valuable resource for further study of transcriptional control of cell identity and for reprogramming (Cherry and Daley, 2012; Graf and Enver, 2009; Lee and Young, 2013; Zhou et al., 2008).

Several recent studies suggest that much of disease-associated DNA sequence variation occurs in transcriptional regulatory regions defined by DNase hypersensitivity (Maurano et al., 2012; Vernot et al., 2012). We found that disease-associated SNPs occur in super-enhancers of disease-relevant cells and that this occurs more frequently for super-enhancers than for typical enhancers. Because super-enhancers drive the expression of genes that control and define cell identity, these results suggest that altered expression of cell identity genes may often contribute to these diseases. These observations also suggest that hypotheses regarding the role of specific cell types and genes in many diseases might be guided in the future by knowledge of super-enhancers.

Cancer cells acquire super-enhancers at oncogene drivers during the process of tumor pathogenesis. Cancer cells appear to acquire super-enhancers through a variety of mechanisms,

(B) Cancer cells acquire super-enhancers. ChIP-seq binding profiles for H3K27ac are shown at the gene desert surrounding *MYC* in pancreatic cancer, T cell leukemia, colorectal cancer, and healthy counterparts. In colorectal cancer, several regions in the 1 MB window upstream of *MYC* were shown to interact with the *MYC* gene in colorectal cancer (Ahmadiyeh et al., 2010; Pomerantz et al., 2009). rpm/bp, reads per million per base pair.

(C) Chromosomal translocation, overexpression of transcription factors, and focal amplification may contribute to super-enhancer formation in cancer. Displayed are ChIP-seq binding profiles for H3K27ac and indicated transcription factors at the gene desert surrounding *MYC* in the indicated cancers. (Top) A translocation event places the *MYC* gene proximal to an inserted IgH super-enhancer in multiple myeloma. (Middle) Tal1 binding is observed at a distal super-enhancer in T cell leukemia. (Bottom) Large H3K27ac domains are observed at the site of focal amplification in lung cancer. The red bars below the binding profiles indicate the genomic positions of focal amplification in six different samples, two of which (SM09-019T and SM09-11T1) are primary patient samples (Iwakawa et al., 2013).

(D) Tumor-specific super-enhancers associate with hallmark cancer genes in colorectal cancer. (Top) Diagram of the ten hallmarks of cancer adapted from Hanahan and Weinberg, 2011. Genes associated with super-enhancers in colorectal cancer, but not in healthy colon samples, were assigned to hallmark categories based on their functions and their previous implication in tumorigenesis. Prominent genes that associate with tumor-specific super-enhancers are highlighted at each cancer hallmark. (Bottom) Distribution of H3K27ac signal across enhancers identified in colorectal cancer. Uneven distribution of signal allows the identification of 387 super-enhancers. Prominent genes associated with super-enhancers in colorectal cancer, but not in healthy colon, are highlighted with their respective super-enhancer ranks and cancer hallmarks that they were assigned to.

(E) Super-enhancers acquired by cancer cells associate with hallmark genes. Each cancer hallmark was assigned a Gene Ontology term, and the number of genes that are associated with acquired super-enhancers and are included in that GO term is displayed for each cancer. Asterisk denotes statistical significance above genomic expectation (hypergeometric test, $p < 0.05$).

See also Figure S5 and Tables S2 and S5.

including chromosomal translocation of super-enhancers normally associated with other genes, focal amplification, or overexpression of an oncogenic transcription factor. The super-enhancers acquired by cancer cells are associated with a remarkably broad spectrum of oncogenes that have been described thus far in cancer (Bishop, 1987; Fearon and Vogelstein, 1990; Forbes et al., 2010; Futreal et al., 2004; Garraway and Lander, 2013; Hanahan and Weinberg, 2011; Vogelstein et al., 2013). They are also associated with genes that function in the acquisition of hallmark capabilities in cancer (Hanahan and Weinberg, 2011). These results suggest that super-enhancers can provide biomarkers for cancer-specific pathologies that may be valuable for further understanding cancer biology, diagnosis, and therapy.

EXPERIMENTAL PROCEDURES

Data Analysis

ChIP-seq data sets were aligned using Bowtie (version 0.12.9) to build version mm9 of the mouse genome or version hg19 of the human genome. The GEO accession IDs for all analyzed data sets are listed in Table S6.

Normalized read density of a ChIP-seq data set in any region was calculated as described (Whyte et al., 2013). ChIP-seq reads aligning to the region were extended by 200 bp, and the density of reads per base pair (bp) was calculated. The density of reads in each region was normalized to the total number of million mapped reads producing read density in units of reads per million mapped reads per base pair (rpm/bp).

We used the MACS version 1.4.1 (model-based analysis of ChIP-seq) (Zhang et al., 2008) peak-finding algorithm to identify regions of ChIP-seq enrichment over background. A p value threshold of enrichment of 1×10^{-9} was used for all data sets.

Enhancers were defined as regions of ChIP-seq enrichment for transcription factors in murine ESCs and H3K27ac in human cells. To accurately capture dense clusters of enhancers, we allowed enhancer regions within 12.5 kb of one another to be stitched together.

The methods for identifying super-enhancers and the assignment of enhancers to genes are fully described in the Supplemental Information.

ACCESSION NUMBERS

The CBP, Mbd3, and Ronin ChIP-seq data in mESCs; the H3K27ac ChIP-seq data in RPMI-8402; and the RNA-seq data in mESCs have been deposited with the Gene Expression Omnibus under the accession ID GSE51522.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, six tables, and two data files and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.09.053.

ACKNOWLEDGMENTS

We thank David Orlando and Charles Lin for bioinformatics support; Lee Lawton, Jamie Newman, Peter Rahl, and Warren Whyte for sharing data; Nancy Hannett for assistance with data acquisition; Tom Volkert, Jennifer Love, Sumeet Gupta, and Jeong-Ah Kwon at the Whitehead Genome Technologies Core for Solexa sequencing; and members of the Young lab for helpful discussions. We acknowledge the Genome Centers at the Broad Institute and the University of California and San Diego for ChIP-seq data sets produced in the framework of the ENCODE and NIH Roadmap Epigenome Projects and the Rush Alzheimer's Disease Center at Rush University Medical Center, Chicago (supported by P30AG10161, R01AG17917, R01AG36042, and RC2AG36547). This work was supported by the National Institutes of Health grants HG002668 (R.A.Y.), CA109901 (R.A.Y.) and CA146445 (R.A.Y. and T.I.L.). R.A.Y. is a founder of Syros Pharmaceuticals.

Received: September 9, 2013
Revised: September 27, 2013
Accepted: September 30, 2013
Published: October 10, 2013

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. Proc. Natl. Acad. Sci. USA *107*, 9742–9746.

Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Bertram, L., and Tanzi, R.E. (2008). Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat. Rev. Neurosci. *9*, 768–778.

Bishop, J.M. (1987). The molecular genetics of cancer. Science *235*, 305–311.

Bluestone, J.A., Herold, K., and Eisenbarth, G. (2010). Genetics, pathogenesis and clinical interventions in type 1 diabetes. Nature *464*, 1293–1300.

Bonifer, C. (2000). Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? Trends Genet. *16*, 310–315.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Carey, M. (1998). The enhanceosome and transcriptional synergy. Cell *92*, 5–8.

Chapuis, J., Hansmannel, F., Gistelinck, M., Mounier, A., Van Cauwenberghe, C., Kolen, K.V., Geller, F., Sottejeau, Y., Harold, D., Dourlen, P., et al. (2013). Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. Mol. Psychiatry. Published online February 12, 2013. http://dx.doi.org/10.1038/mp.2013.1.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Cherry, A.B., and Daley, G.Q. (2012). Reprogramming cellular identity for regenerative medicine. Cell *148*, 1110–1122.

Costa-Reis, P., and Sullivan, K.E. (2013). Genetics and epigenetics of systemic lupus erythematosus. Curr. Rheumatol. Rep. *15*, 369.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.

Deng, Y., and Tsao, B.P. (2010). Genetic susceptibility to systemic lupus erythematosus in the genomic era. Nat. Rev. Rheumatol. *6*, 683–692.

Denslow, S.A., and Wade, P.A. (2007). The human Mi-2/NuRD complex and gene regulation. Oncogene *26*, 5433–5438.

Dowen, J.M., Bilodeau, S., Orlando, D.A., Hubner, M.R., Abraham, B.J., Spector, D.L., and Young, R.A. (2013). Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. Stem Cell Rep. *1*. Published online October 24, 2013. http://dx.doi.org/10.1016/j.stemcr.2013.09.002.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell *61*, 759–767.

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 Res. *38*(Database issue), D652–D657.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990). A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. Genes Dev. *4*, 1637–1649.

Foster, C.T., Dovey, O.M., Lezina, L., Luo, J.L., Gant, T.W., Barlev, N., Bradley, A., and Cowley, S.M. (2010). Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability. Mol. Cell. Biol. *30*, 4851–4863.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. Cell *153*, 17–37.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D.J., Griesemer, D., Karlsson, E.K., Wong, S.H., et al.; 1000 Genomes Project. (2013). Identifying recent adaptations in large-scale genomic data. Cell *152*, 703–713.

Grosveld, F., van Assendelft, G.B., Greaves, D.R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell *51*, 975–985.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA *106*, 9362–9367.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009a). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc. Natl. Acad. Sci. USA *106*, 5187–5191.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009b). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc. Natl. Acad. Sci. USA *106*, 5181–5186.

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.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Mol. Cell *19*, 523–534.

Jiang, Y.W., Veschambre, P., Erdjument-Bromage, H., Tempst, P., Conaway, J.W., Conaway, R.C., and Kornberg, R.D. (1998). Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. Proc. Natl. Acad. Sci. USA *95*, 8538–8543.

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.

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat. Cell Biol. *8*, 285–292.

Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development *134*, 1123–1132.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

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. Nat. Struct. Mol. Biol. *18*, 956–963.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497–501.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature *498*, 511–515.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237–1251.

Lelli, K.M., Slattery, M., and Mann, R.S. (2012). Disentangling the many layers of eukaryotic transcriptional regulation. Annu. Rev. Genet. *46*, 43–68.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147–151.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516–520.

Ling, J., Ainol, L., Zhang, L., Yu, X., Pi, W., and Tuan, D. (2004). HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. J. Biol. Chem. *279*, 51704–51713.

Lovén, 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.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. Annu. Rev. Genomics Hum. Genet. *7*, 29–59.

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.

Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Mol. Cell *1*, 277–287.

Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. Mol. Cell *51*, 606–617.

Mullen, A.C., Orlando, D.A., Newman, J.J., Lovén, 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-β signaling. Cell *147*, 565–576.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annu. Rev. Genet. *46*, 1–19.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Ng, H.H., and Surani, M.A. (2011). The transcriptional and signalling networks of pluripotency. Nat. Cell Biol. *13*, 490–496.

Noble, J.A., and Erlich, H.A. (2012). Genetics of type 1 diabetes. Cold Spring Harb. Med. *2*, a007732.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. *12*, 283–293.

Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. Cell *145*, 835–850.

Ørom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46–58.

Panne, D. (2008). The enhanceosome. Curr. Opin. Struct. Biol. *18*, 236–242.

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. Nat. Genet. *41*, 882–884.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012a). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell *10*, 583–594.

Reynolds, N., Salmon-Divon, M., Dvinge, H., Hynes-Allen, A., Balasooriya, G., Leaford, D., Behrens, A., Bertone, P., and Hendrich, B. (2012b). NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. EMBO J. *31*, 593–605.

Schnetz, M.P., Handoko, L., Akhtar-Zaidi, B., Bartels, C.F., Pereira, C.F., Fisher, A.G., Adams, D.J., Flicek, P., Crawford, G.E., Laframboise, T., et al. (2010). CHD7 targets active gene enhancer elements to modulate ES cell-specific gene expression. PLoS Genet. *6*, e1001023.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

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., and Young, R.A. (2013). Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. Proc. Natl. Acad. Sci. USA *110*, 2876–2881.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Tanzi, R.E. (2012). The genetics of Alzheimer disease. Cold Spring Harb. Perspect. Med. *2*, a006296.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Tuan, D., Solomon, W., Li, Q., and London, I.M. (1985). The "beta-like-globin" gene domain in human erythroid cells. Proc. Natl. Acad. Sci. USA *82*, 6384–6388.

Vernot, B., Stergachis, A.B., Maurano, M.T., Vierstra, J., Neph, S., Thurman, R.E., Stamatoyannopoulos, J.A., and Akey, J.M. (2012). Personal and population genomics of human regulatory variation. Genome Res. *22*, 1689–1697.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

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.

Xie, W., and Ren, B. (2013). Developmental biology. Enhancing pluripotency and lineage specification. Science *341*, 245–247.

Young, R.A. (2011). Control of the embryonic stem cell state. Cell *144*, 940–954.

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

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627–632.