Cell Type-Specific Chromatin Signatures Underline Regulatory DNA Elements in Human Induced Pluripotent Stem Cells and Somatic Cells

Ming-Tao Zhao[1,2,3], **Ning-Yi Shao**[1,2,3], **Shijun Hu**[1,2,3], **Ning Ma**[1,2,3], **Rajini Srinivasan**[4], **Fereshteh Jahanbani**[5], **Jaecheol Lee**[1,2,3], **Sophia L. Zhang**[1,2,3], **Michael P. Snyder**[5], and **Joseph C. Wu**[1,2,3]

[1]Stanford Cardiovascular Institute, Stanford, California 94305, USA

[2]Department of Medicine, Division of Cardiology, Stanford, California 94305, USA

[3]Institute of Stem Cell Biology and Regenerative Medicine, Stanford, California 94305, USA

[4]Department of Chemical and Systems Biology, Stanford, California 94305, USA

[5]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

Abstract

Rationale—Regulatory DNA elements in the human genome play important roles in determining the transcriptional abundance and spatiotemporal gene expression during embryonic heart development and somatic cell reprogramming. It is a mystery how chromatin marks in regulatory DNA elements are modulated to establish cell type-specific gene expression in the human heart.

Objective—We aimed to decipher the cell type-specific epigenetic signatures in regulatory DNA elements and how they modulate heart-specific gene expression.

Methods and Results—We profiled genome-wide transcriptional activity and a variety of epigenetic marks in the regulatory DNA elements using massive RNA-seq (n=12) and ChIP-seq (n=84) in human endothelial cells (ECs: CD31$^+$CD144$^+$), cardiac progenitor cells (CPCs: Sca1$^+$), fibroblasts (FBs: DDR2$^+$), and their respective induced pluripotent stem cells (iPSCs). We uncovered two classes of regulatory DNA elements: Class I was identified with ubiquitous enhancer (H3K4me1) and promoter (H3K4me3) marks in all cell types, whereas Class II was enriched with H3K4me1 and H3K4me3 in a cell type-specific manner. Both Class I and Class II regulatory elements exhibited stimulatory roles in nearby gene expression in a given cell type. However, Class I promoters displayed more dominant regulatory effects on transcriptional abundance regardless of distal enhancers. Transcription factor network analysis indicated that human iPSCs and somatic cells from the heart selected their preferential regulatory elements to maintain cell type-specific gene expression. In addition, we validated the function of these

enhancer elements in transgenic mouse embryos and human cells, and identified a few enhancers that could possibly regulate the cardiac-specific gene expression.

Conclusions—Given that a large number of genetic variants associated with human diseases are located in regulatory DNA elements, our study provides valuable resources for deciphering the epigenetic modulation of regulatory DNA elements that fine-tune spatiotemporal gene expression in human cardiac development and diseases.

Keywords

Human iPSCs; endothelial cells; cardiac progenitor cells; Fibroblasts; Regulatory DNA elements; stem cell; epigenetics; histone markers

Subject Terms

Cellular Reprogramming; Stem Cells; Epigenetics; Gene Expression and Regulation

INTRODUCTION

Human pluripotent stem cells (PSCs) share the dual hallmarks of self-renewal and ability to generate all cell types in the body, and thereby they hold great promise in disease modeling, drug development, and regenerative medicine.[1–3] Human induced pluripotent stem cells (iPSCs) are directly derived from somatic cells by transient overexpression of four transcription factors (OCT4/SOX2/C-MYC/KLF4),[4] and as such they are free of ethical issues associated with the use of human oocytes for nuclear reprogramming and therapeutic cloning.[5, 6] For genetically inherited cardiovascular diseases, patient-specific iPSCs have emerged as powerful tools to generate human cardiac cells for modeling disease progression and novel drug discovery. Though disease-causing mutations are frequently seen in protein-coding regions across the genome, non-coding sequences, including regulatory DNA elements, have been demonstrated to affect the susceptibility to many cardiovascular diseases.[7] However, evaluating regulatory DNA elements in cardiovascular pathogenesis has been difficult due to the lack of a catalog of cardiac cell type-specific regulatory DNA elements, particularly for enhancers and promoters.

In mammals, cell type identity is defined and maintained by specific gene expression programs. Cell type-specific gene expression is primarily driven by the proximal and distal regulatory DNA elements, including promoters and enhancers. Promoters are short DNA sequences proximal to the transcription start sites (TSSs) bound by general transcriptional machinery. Enhancers are usually distal to TSSs and contain short DNA motifs that can be recognized by lineage-determining transcription factors.[8] A large number of putative enhancers have been identified and comprise at least 12% of the human genome.[9] Cell type-specific enhancers are usually associated with histone modification and higher-order chromatin structure, which can be employed to predict putative enhancers in a given cell type.[10] Genome-wide chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-seq) studies reveal that tissue-specific enhancers are enriched with several chromatin marks.[11, 12] In human embryonic stem cells (ESCs), active enhancers are mostly

associated with p300, H3K4me1, and H3K27ac, whereas poised enhancers are enriched in H3K27me3 with the depletion of H3K27ac.[13, 14]

Somatic cell reprogramming is accompanied by resetting of cell type-specific transcriptional programs from the differentiated cell state to the pluripotent state. Cellular differentiation of patient-specific iPSCs to cardiac lineages is associated with extensive epigenetic reprogramming, which includes massive reorganization of DNA and histone modifications at regulatory DNA elements.[15, 16] Recent studies have illustrated the dynamic and coordinated epigenetic modulation of regulatory DNA elements during cardiac lineage differentiation. [17–19] Though thousands of tissue-specific enhancers have been identified in the heart using ChIP-seq prediction,[20] they are not cell type-specific. Conversely, it is still unknown how genome-wide reorganization of chromatin modifications in regulatory DNA elements is established when somatic cells from the heart are reprogrammed into pluripotent cells, which may also inform cardiac lineage dedifferentiation. Here we performed massive RNA-seq (n=12) and ChIP-seq (n=84) to profile genome-wide transcriptional activity, as well as promoter and enhancer chromatin marks (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) in human iPSCs and their parental cells (fibroblasts, endothelial cells, and cardiac progenitor cells) from the same individuals. We identified two classes of cell type-specific regulatory DNA elements in human iPSCs and somatic cells, and functionally validated the putative enhancer elements in transgenic mouse embryos and human cells.

METHODS

A detailed description of the experimental procedures is provided in the Online Data Supplement.

RESULTS

Resetting cell type-specific transcriptional program by somatic cell reprogramming

To remove potential effects of genetic composition, we generated multiple human iPSC lines from isogenic somatic cells derived from the same fetal heart: fibroblasts (FBs), endothelial cells (ECs), and Sca-1+ cardiac progenitor cells (CPCs) (Figure 1A).[21] The resulting iPSCs were denoted as FiPSCs, EiPSCs, and CiPSCs, respectively. These iPSCs highly expressed OCT4 and NANOG (Online Figure IA), with the majority of cells in the colonies being OCT4+NANOG+ (Online Figures IB–D). We confirmed the identity of somatic cells by flow cytometry using cell surface markers: CD31/CD144 for ECs, Sca-1 for CPCs, and DDR2 for FBs (Online Figures IE–I). Next we performed high-throughput RNA-seq to profile the transcriptional changes in these somatic cells and their respective iPSCs. The reprogramming process reshaped the transcriptomes of somatic cells to the pluripotent state, regardless of the parental transcriptional signatures. The transcriptional difference between somatic cells and iPSCs were apparent, with 6,151 differentially expressed genes (DEGs) identified (Figure 1B). We further divided these DEGs into five cell type-specific clusters (clusters A through E): 87% (5,353 genes, clusters A and B) of DEGs were between iPSCs and somatic cells, including 279 EC-specific genes (cluster C), 205 fibroblast signature genes (cluster D), and 314 CPC/FB-specific genes (cluster E). We also checked the cell type-specific signature gene expression, discovering that *POU5F1* (cluster A) were uniquely

expressed in human iPSCs (Figure 1D), *CDH5* (cluster B) in somatic cells, *VWF* (cluster C) in ECs, *S100A4* in FBs (cluster D), and *GDF6* (cluster E) in FBs and CPCs (Online Figures IIA–D). Gene ontology analysis showed that these DEGs were mostly associated with blood vessel morphogenesis, cardiovascular development, and focal adhesion, highlighting the fundamental transcriptional differences between iPSCs and somatic cells (Figure 1E).

In general, gene expression variation is far greater in different tissues (and derived primary cells) than in the same tissue with different genetic makeups.[22] Within iPSCs, we found that the transcriptional variance was mostly contributed by the genetic makeups. The PCA plot of global gene expression showed that iPSCs were clearly separated by the individual genetic background (Figure 1C). When compared with somatic cell types, the inter-iPSC transcriptional variation was much smaller than that between iPSCs and somatic cells (Online Figure IIE). These results were consistent with previous studies and reiterated the influence of genetic composition on the gene expression of human iPSCs.[23] Collectively, these results indicate that cell type-specific transcriptomes of somatic cells from the heart are reshaped to the unique gene expression pattern in iPSCs, the transcriptional variation of which is mostly driven by genetic makeups rather than the cell types of origin.

Identification of two classes of cell type-specific enhancers in iPSCs and somatic cells

To identify prospective enhancers, we next performed ChIP-seq experiments (n=84) using antibodies against several histone marks (H3K4me1, H3K4me3, H3K27ac, and H3K27me3), co-factor (p300), and a component of transcriptional machinery (RNA polymerase II, Pol II). Overall, these chromatin marks and co-factors showed a genome-wide cell type-specific distribution, and iPSCs were obviously separated from their parental somatic cells in the t-SNE plot (Online Figure III). H3K27ac and H3K4me1 have been widely used to identify active (H3K4me1+/H3K27ac+) and poised (H3K4me1+/H3K27ac-) enhancers.[13, 24] Because we had a variety of conditions (six cell types) with multiple sets of chromatin marks, we first used H3K27ac to predict all potential enhancers outside of ±3kb regions of annotated transcription start sites. In total, we identified 46,261 potential enhancer elements using significantly enriched H3K27ac peaks in at least one of our 12 samples. We further divided these putative enhancers into two categories based on the pattern of H3K4me1 enrichment.[25] Class I enhancers were enriched with H3K4me1 in all cell types, whereas Class II enhancers exhibited cell type-specific H3K4me1 enrichment. Class I enhancers (2,700) comprised of 5.8% of the total, whereas Class II enhancers (43,561) were dominant in all putative enhancers (Online Table I). These putative enhancers were active (H3K4me1+/H3K27ac+) in at least one cell type and were poised or silenced in other cell types.

Ubiquitous H3K4me1 enhancers are mostly active in human iPSCs

Class I enhancers showed cell type-specific distribution of H3K27ac and ubiquitous enrichment of H3K4me1 in both somatic cells and iPSCs (Figures 2A–B). Because H3K27ac is enriched in active enhancers, most Class I enhancers displayed high activity in iPSCs, but were selectively active in some somatic cell types (Figure 2A). We also examined the p300 and Pol II distribution on the same genomic loci enriched by H3K27ac. H3K27ac enrichment was positively correlated with the binding of co-factor p300 and the component

of transcriptional machinery Pol II (Figures 2C–D), indicating synergized chromatin modifications for active gene transcription. Furthermore, we observed positively correlated H3K4me3 and negatively correlated H3K27me3 enrichment across these genomic regions shared with H3K27ac (Online Figures IVA–B). We performed correlation analysis and found H3K27ac was positively correlated with H3K4me1, p300, and Pol II, but was negatively correlated with H3K27me3 in Class I enhancers (Online Figure V). There were two conditions for Class I enhancers: active enhancers with both H3K27ac and H3K4me1 enrichment (H3K27ac+/H3K4me1+) in one cell type versus poised enhancers with only H3K4me1 enrichment (H3K27ac-/H3K4me1+) in all cell types (Figure 2E). Class I enhancers were mostly located in the gene bodies (>60%) and intergenic regions (Online Figure IVD). In contrast, a substantial number of Class II enhancers were located in the gene deserts (Online Figure IVE).

Class I enhancers were further grouped into 7 clusters (1 to 7, from the top to the bottom, Figure 2A) depending on the dynamic profiles of H3K27ac enrichment in somatic cells and iPSCs. Cluster 1 and 2 enhancers were mostly active in human iPSCs (80% of Class I enhancers) (Online Table I), as shown by the enrichment of active enhancer mark H3K27ac, co-factor p300, and RNA Pol II. Class I enhancers were typically enriched with a high density of H3K4me1 and cell type-specific distribution of H3K27ac across a large genomic region (Figure 2F). These enhancers can be activated in one cell type but were poised in another cell type, with most of them being active in human iPSCs but poised in somatic cells (Figure 2A). We then interrogated the nearby gene expression profiles of these clusters. We found the average transcription levels of these genes were affected by Class I enhancers (Figure 2G and Online Figure IVC). The gene expression level was well correlated with the enrichment of H3K27ac; higher levels of H3K27ac enrichment corresponded to higher levels of cell type-specific gene expression. We then looked into the significantly enriched gene ontology terms of the nearby genes in these clusters. We found that cluster 1 genes were associated with signal transduction, cell communication, and endocytosis, whereas Cluster 4 genes were related to blood vessel development and endothelial cell function (Online Figures VIA–B). Furthermore, motif enrichment analysis of these clusters in Class I enhancers revealed that they were possibly bound by lineage-determining transcription factors, such as ETV1, ETV2, and ERG (Online Figures VIC–D). Together, these results indicate that Class I enhancers are mostly active in iPSCs and possibly modulate the establishment of iPSC-specific gene expression during somatic cell reprogramming.

Cell type-specific H3K4me1 enhancers reflect cell type-specific gene expression

Class II enhancers were the major part (94.2%) of cell type-specific enhancers identified between human iPSCs and their parental somatic cells. These enhancers showed cell type-specific enrichment of H3K27ac and H3K4me1, positively correlating with the enrichment of co-factor p300 and RNA Pol II (Figures 3A–D). Additionally, H3K4me3 displayed a similar cell type-specific distribution pattern, whereas the repressive mark H3K27me3 was not significantly enriched in a cell type-specific manner (Online Figures VIIA–B). Next we grouped Class II enhancers into seven clusters based on their activation patterns in human iPSCs and somatic cells. Class II enhancers were highly cell type-specific with iPSC-specific enhancers (active in iPSCs) accounting for only 28.2%, in contrast to the fact that

79.8% of Class I enhancers were active in iPSCs (Online Table I). Class II enhancers were either active (H3K4me1+/H3K27ac+) or silenced (H3K4me1-/H3K27ac-), not counting the poised enhancers that were prevalent in Class I enhancers (Figure 2E). Most of the Class II enhancers were located in gene bodies and intergenic regions. However, about 10% of Class II enhancers were situated in gene deserts devoid of protein-coding genes (Online Figure IVE), indicating the possible functional differences between Class I and Class II enhancers.

We then investigated the effects of Class II enhancers on the cell type-specific gene expression by examining the expression levels of the nearby genes. Active Class II enhancers were separated from silent enhancers by H3K27ac enrichment, though H3K4me1 could cover a broader genomic locus (Figure 3E). We observed consistently higher gene expression activities in active Class II enhancers than those in silent Class II enhancers across all clusters (Figure 3F and Online Figure VIIC), suggesting the former had greater functional activity in regulating gene expression.

We next looked into the functions of genes that were putatively regulated by Class II enhancers. Interestingly, the genes regulated by Class II enhancers were different from those regulated by Class I enhancers. For example, nearby genes targeted by cluster I (iPSC-specific) enhancers were mostly associated with chromatin modification, anti-apoptosis, and organelle organization (Online Figure VIID), whereas flanking genes affected by cluster 5 (EC-specific) enhancers were related to chemokine production, inflammatory response, and immune system process (Online Figure VIIE). These results indicate that compared to gene functions of Class I enhancers that are mostly associated with cell type identity, Class II enhancers appear to regulate the biological function of specific cell types. We next examined the transcription factor motifs enriched by Class II enhancers. Interestingly, the top enriched TF motifs in cluster 1 were pluripotent stem cell transcription factors (OCT4, SOX2, and NANOG), whereas those in cluster 4 were relevant to endothelial cell lineage determination (Figures 3G–H), which were distinct from motifs bound by Class I enhancers (Online Figures VIC–D). Additionally, distinct gene ontology terms were associated with the genes that were potentially regulated by Class I and Class II enhancers, respectively (Online Figure VIII). Taken together, these results suggest that Class II enhancers reflect cell type-specific expression by regulating cell identity determining TFs, and modulate different biological functions compared to Class I enhancers.

Ubiquitous H3K4me3 promoters are prevalent in human iPSCs and somatic cells

As promoters are usually marked by H3K4me3 and located adjacent to the TSSs,[26] we next probed the epigenetic signatures of promoters using H3Kme3, H3K27ac (active), H3K27me3 (repressive), and RNA Pol II (transcription) (Figures 4A–D). We also interrogated other histone marks (H3K4me1 and p300), but did not find a significant enrichment in the promoter regions (Online Figures IXA–B). To exclude any potential enhancers, we only looked into ±3 kb within transcription start sites. We identified 5,230 promoter regions with differential enrichment of H3K27ac activity between human iPSCs and their parental cells (FBs, ECs, and CPCs). We further divided them into two distinct groups according to the distribution of general promoter mark H3K4me3: Class I promoters with ubiquitous H3K4me3 distribution versus Class II promoters with cell-type specific

H3K4me3 enrichment (Figure 4E). Promoters with both H3K4me3 and H3K27ac were considered active; promoters with H3K4me3 but without H3K27ac were poised; and promoters with neither H3K4me3 nor H3K27ac were inactive in a given cell type. Surprisingly, about 75% of these promoters (3,925) were Class I promoters with ubiquitous H3K4me3 enrichment and cell type-specific distribution of H3K27ac (Figures 4A–D and Online Table II). In contrast, the repression mark H3K27me3 was negatively correlated with active mark H3K27ac (Figure 4C). The genes driven by active Class I promoters (H3K27ac +) showed a higher transcriptional activity than those with low H3K27ac enrichment in any given cell types (Figures 4F–G and Online Figure IXC). Although only a small percentage (5.8%) of Class I enhancers showed ubiquitous H3K4me1 enrichment, Class I promoters were much more prevalent (75%) and constituted the majority of promoters driving strong cell type-specific gene expression.

Class II promoters with cell-type specific H3K4me3 enrichment are weaker in driving gene expression

About a quarter of cell type-specific promoters (1,305) were enriched with cell type-specific H4K3me3 and H3K27ac and termed as Class II promoters (Figure 4E). Class II promoters were marked with cell type-specific H3K27ac, H3K4me3, and RNA Pol II but negatively correlated with the repressive mark H3K27me3 (Figures 5A–D). This cell type-specific enrichment pattern was also observed for histone mark H3K4me1 and co-factor p300 (Online Figures XA–B). The genes associated with Class II promoters displayed dramatic cell type-specific expression patterns: active promoters (H3K4me3+/H3K27ac+) drove higher gene expression than inactive promoters (H3K4me3-/H3K27ac-) in any given cell types (Figures 5E–F). Class II promoters were further divided into 7 cell type-specific clusters (Online Table II). For example, cluster 1 promoters were active in iPSCs whereas cluster 2 promoters were active in somatic cells. The cell type-specific gene expression regulated by these clusters was correlated with chromatin mark (H3K27ac) enrichment (Figures 5E–F and Online Figure XC). However, the average levels of gene expression driven by Class II promoters were much lower than those driven by Class I promoters (Figures 4F–G), suggesting a stronger promoter activity with consistent H3K4me3 presence in all cell types. We then surveyed the potential TF motifs enriched by Class I and Class II promoters. Compared to enhancers, the TF motif enrichment scores for promoters were much lower, though stem cells factor POU5F1 motif was enriched in cluster 1 of Class I promoters (Figure 5G). TF motifs enriched by Class II promoters were distinct from those by Class I promoters (Figure 5H), suggesting that different biological functions are modulated by these two types of promoters. In addition, the biological functions of genes regulated by Class I and Class II promoters were clearly separated. Class I promoters were mostly associated with cellular development and gene expression regulation, whereas Class II promoters regulated genes relevant to cellular and molecular functions and metabolic processes (Online Figure XI). In summary, we identified two classes of cell type-specific promoters with distinct gene regulatory functions that were primed by histone chromatin marks (H3K4me3 and H3K27ac).

Cell type-specific gene expression regulated by promoters and enhancers

Cell type-specific gene expression is regulated by distal enhancers and driven by proximal promoters.[8] To illustrate the combinatorial influence of promoters and enhancers on cell type-specific transcriptional activity, we analyzed the common genes regulated by both Class I and II promoters and enhancers activated in a given cell type. These genes were divided into four groups: Class I enhancers/Class I promoters (E1_P1, 497 genes), Class I enhancers/Class II promoters (E1_P2, 162 genes), Class II enhancers/Class I promoters (E2_P1, 2,245 genes), and Class II enhancers/Class II promoters (E2_P2, 882 genes) (Online Figure XIIA). These overlapped genes were determined by the genomic locations near the regulatory DNA elements so the overlaps within promoters and enhancers were also observed (Figure 6A). We next examined the expression of these common genes regulated by the combination of promoters and enhancers. Regardless of the presence of enhancers, transcription activity executed by Class I promoters was much stronger than those driven by Class II promoters, though the average transcripts were different among individual cell types (Online Figures XIIB–C). In particular, gene expression was predominately affected by the activity of promoters, with Class I promoters showing a higher gene expression than Class II promoters in any combinations with enhancers in both human iPSCs and somatic cells (Figure 6B). Accordingly, the gene ontology (GO) analysis also showed higher enrichment scores (lower P-values) associated with common genes mediated by Class I promoters than those by Class II promoters, independent of Class I or Class II enhancer activity (Online Figure XIID). Finally, we constructed gene regulatory networks associated with cell type-specific transcription factors, regulatory DNA elements, and gene expression (mRNA transcripts). For iPSCs, all classes of regulatory elements (promoters and enhancers) were potentially targeted by stem cell factors NANOG, OCT4, STAT3, and SOX2 (Figure 6C). However, in endothelial cells, Class II enhancers preferentially interacted with endothelial TFs such as ETV2, NR2F2, and GATA2 (Figure 6D), suggesting that human iPSCs and somatic cells from the heart exhibit distinct preferences in selecting regulatory DNA elements to maintain their cell type-specific transcriptional program. Taken together, these results demonstrate that promoters determine the transcriptional activity and highlight the role of enhancers on the cell type-specific transcriptional activation.

Functional validation of putative regulatory DNA elements

To functionally validate the cell type-specific regulatory DNA elements, we first performed data mining to locate the identified enhancers in the Vista Enhancer Browser (https://enhancer.lbl.gov). We found many of these cell type-specific human enhancers could modulate the tissue-specific gene expression in transgenic mouse embryos (Figure 7A), highlighting the evolutionary conservation of these regulatory elements between human and mouse.[27] We retrieved a number of human enhancer elements that could drive the tissue-specific expression of the reporter, particularly in the heart, blood vessel and somite of mouse transgenic embryos (Figures 7B–D). To further confirm the activity of these enhancer elements in human cells, we made enhancer reporter constructs with a basal promoter driving a firefly luciferase reporter (Figure 7E). We transfected multiple types of human cells, including iPSCs, iPSC-derived cardiomyocytes (CMs), ECs (fetal aorta), and FBs (fetal heart) to test the cell type-specific activation of these enhancer elements. As predicted, the basal construct pGL3-promoter lacking any enhancer elements did not show cell type-

specific reporter activity (Figure 7F). In contrast, the vector including a SV40 enhancer could drive more preferential expression of reporter genes in HEK293T cells than any other cell types (Figure 7G), indicating the cell type-specific activation of enhancer elements. Using the human enhancer reporter vectors for transfection, we observed a cell type-specific enhancement of reporter luciferase activity, with most of these enhancers highly active in CMs and ECs (Figure 7H). These results were consistent with the tissue-specific gene expression in the transgenic embryos, as these enhancers could modulate the reporter genes specifically in the heart (CMs) and blood vessel (ECs) (Figures 7A–D). To further illustrate the target genes that are possibly modulated by these cell type-specific enhancers, we surveyed the expression of genes adjacent to these enhancer elements. HS2205 was a 4.8 kb enhancer element residing in the *GATA4* locus. *GATA4* was highly expressed in cardiomyocytes compared to other cell types (Figure 7I), coinciding with a high level of H3K27ac enrichment in this region (Figure 7J). Simultaneously, HS2205 could exogenously drive the heart-specific mRNA expression in transgenic mouse embryos (Figure 7D), indicating that *GATA4* is likely regulated by this enhancer. In addition, we found that HS1887 could potentially regulate heart (cardiomyocyte)-specific expression of *TEAD3* and HS2027 would possibly modulate the expression of *TANC1* in cardiomyocytes and endothelial cells (Online Figures XIIIA–B). This prediction was further consolidated with heart-specific reporter gene expression driven by these enhancer elements (Online Figures XIIIC–D). The activation histone mark H3K27ac was also enriched in the element HS2027 in a cell type-specific manner, which was positively correlated with the cell type-specific gene expression in somatic cells (Online Figures XIIIE). In summary, we validated the cardiac-specific enhancer elements in human cells and transgenic mouse embryos, and identified the target genes that could be potentially regulated by these enhancers in the heart.

DISCUSSION

In this study, we identified two classes of cell type-specific enhancers and promoters based on chromatin histone marks (H3K27ac, H3K4me1, and H3K4me3) enrichment in human iPSCs and somatic cells (Figure 6E). We found that ubiquitous H3K4me1 enhancers (Class I) were mostly active in human iPSCs whereas cell type-specific H3K4me1 enhancers (Class II) reflected cell type-specific gene expression. Likewise, we discovered two types of promoters with ubiquitous (Class I) H3K4me3 and cell type-specific (Class II) H3K4me3 enrichment in multiple cell types. Moreover, we validated the function of these human enhancer elements in both transgenic mouse embryos and human cells, and identified a number of enhancers that could potentially modulate cardiac cell-specific gene expression. We conclude that promoters determine the transcriptional activity, whereas enhancers confer cell type-specific gene expression in a particular cell type. Collectively, our data may prove valuable for future efforts to understand the epigenetic chromatin remodeling of regulatory DNA elements in cardiac development and heart diseases.

Previous studies identified poised enhancers marked with H3K4me1 and H3K27me3 but depleted of H3K27ac as underlying developmental enhancers in human ESCs.[13] Though later studies profiled chromatin mark dynamics during human ESC differentiation, they did not investigate the cell type-specific epigenetic features of regulatory DNA elements (promoters and enhancers) between human PSCs and tissue-derived primary cells.[28, 29]

During the process of somatic cell reprogramming, DNA regulatory elements must be epigenetically remodeled to establish stem cell signatures associated with transcription factor binding redistribution.[30] Our study uncovers two classes of cell type-specific DNA regulatory elements in human iPSCs and somatic cells. While other studies have focused on cellular differentiation, particularly ESC/iPSCs and their differentiated progeny, here we used tissue-derived somatic cells to benchmark the regulatory DNA elements for two reasons. First, iPSC-derived differentiated cells are usually immature and more like fetal-stage cells. Second, differentiated stem cell progeny display global epigenetic profiles closer to their parental iPSCs than tissue-derived primary cells.[31] The epigenetic difference between human iPSCs and somatic cells will be informative for improving *in vitro* cardiac lineage differentiation. In addition, transcriptional variation among iPSCs derived from different cell types is mostly contributed by genetic compositions among individuals,[32] and cell type-specific gene expression is completely remodeled to iPSC-specific transcriptional profiles. Therefore, our genomic data pave the way to understanding how cell type-specific transcriptional program is modulated by the interactions between regulatory DNA elements, chromatin marks, and transcription factors during somatic cell reprogramming and cardiac lineage differentiation.

The reciprocal interactions between promoters and enhancers determine the spatiotemporal gene expression during embryonic development. Promoters typically ensure the accurate transcriptional initiation of a gene, whereas enhancers are primarily responsible for the precise regulation of gene expression in a spatial and temporal manner.[33] In this study, we demonstrate the combinatorial effects of promoters and enhancers on cell type-specific gene expression. For genes that are presumptively regulated by both promoters and enhancers, promoters tend to control the quantity of mRNA transcripts, whereas enhancers execute cell type-specific gene expression, though RNA Pol II can bind both of these regulatory regions and initiate transcription. The long-distance interaction of promoters and enhancers mediated by the mediator and cohesin complex may account for their functional control of gene expression in a cell type-specific manner. Recent studies on higher-order chromatin organization in human ESCs and differentiated cells also suggest that enhancers are actively involved in the looping interactions with genes and promoters.[34]

Functional validation of human regulatory DNA elements is crucial for understanding the roles of regulatory elements during embryonic development and disease pathology. Recent genome-wide association studies (GWAS) have identified thousands of human DNA variants associated with complex diseases, the majority of which are noncoding DNA elements.[35] However, the molecular mechanisms of disease-associated loci are rarely illustrated due to lack of systematic annotation of functional noncoding elements. Epigenomic annotation of cardiac specific regulatory DNA elements has facilitated the understanding of the functional roles of previously identified non-coding DNA variants in the contribution to the pathogenesis of cardiovascular diseases.[36] In this respect, our study identified several enhancer elements that could regulate the gene expression of cardiac specific genes (such as *GATA4*) associated with congenital heart disease.[37] This is important because future interrogation of such disease-associated genetic variants in the regulatory DNA elements may generate novel insights on personalized diagnosis and treatment of cardiovascular diseases.

In summary, we have identified two classes of cell type-specific enhancers and promoters in human iPSCs and somatic cells. Class I and Class II regulatory DNA elements exhibit distinct regulatory roles on cell type-specific gene expression in a given cell type. Our study provides invaluable resources for understanding how cell type-specific gene expression is maintained and modulated by regulatory DNA elements, as well as how the cell identity is epigenetically preserved by chromatin modifications in human PSCs and cardiac cells.[19] Given that a large number of genetic variants associated with human diseases are located in regulatory DNA elements, our data will also shed light on the potential genetic and epigenetic interventions to correct abnormal gene expression in a given cell type under disease conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Nonstandard Abbreviations and Acronyms

BFGF	Basic fibroblast growth factor
ChIP	Chromatin immunoprecipitation
CPC	Cardiac progenitor cell
EC	Endothelial cell
EGF	Epithelial growth factor
ESC	Embryonic stem cell
FB	Fibroblast
GO	Gene ontology
HGF	Hepatocyte growth factor
IGF	Insulin-like growth factor
iPSC	Induced pluripotent stem cells
Pol II	RNA polymerase II

TSS   Transcription start site

References

1. Tabar V, Studer L. Pluripotent stem cells in regenerative medicine: Challenges and recent progress. Nature Reviews Genetics. 2014; 15:82–92.

2. De Los Angeles A, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, Hochedlinger K, Jaenisch R, Lee S, Leitch HG, Lensch MW, Lujan E, Pei D, Rossant J, Wernig M, Park PJ, Daley GQ. Hallmarks of pluripotency. Nature. 2015; 525:469–478. [PubMed: 26399828]

3. Wilson KD, Wu JC. Induced pluripotent stem cells. JAMA. 2015; 313:1613–1614. [PubMed: 25919522]

4. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007; 131:861–872. [PubMed: 18035408]

5. Tachibana M, Amato P, Sparman M, Gutierrez NM, Tippner-Hedges R, Ma H, Kang E, Fulati A, Lee HS, Sritanaudomchai H, Masterson K, Larson J, Eaton D, Sadler-Fredd K, Battaglia D, Lee D, Wu D, Jensen J, Patton P, Gokhale S, Stouffer RL, Wolf D, Mitalipov S. Human embryonic stem cells derived by somatic cell nuclear transfer. Cell. 2013; 153:1228–1238. [PubMed: 23683578]

6. Yang X, Smith SL, Tian XC, Lewin HA, Renard JP, Wakayama T. Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. Nature Genetics. 2007; 39:295–302. [PubMed: 17325680]

7. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009; 461:199–205. [PubMed: 19741700]

8. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nature Reviews Molecular Cell Biology. 2015; 16:144–154. [PubMed: 25650801]

9. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. [PubMed: 25693563]

10. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: From properties to genome-wide predictions. Nature Reviews Genetics. 2014; 15:272–286.

11. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–112. [PubMed: 19295514]

12. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. Chip-seq accurately predicts tissue-specific activity of enhancers. Nature. 2009; 457:854–858. [PubMed: 19212405]

13. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–283. [PubMed: 21160473]

14. Calo E, Wysocka J. Modification of enhancer chromatin: What, how, and why? Molecular Cell. 2013; 49:825–837. [PubMed: 23473601]

15. Apostolou E, Hochedlinger K. Chromatin dynamics during cellular reprogramming. Nature. 2013; 502:462–471. [PubMed: 24153299]

16. Papp B, Plath K. Epigenetics of reprogramming to induced pluripotency. Cell. 2013; 152:1324–1343. [PubMed: 23498940]

17. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR, Capra JA, Erwin G, Kattman SJ, Keller GM, Srivastava D, Levine SS, Pollard KS, Holloway AK, Boyer LA, Bruneau BG. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell. 2012; 151:206–220. [PubMed: 22981692]

18. Paige SL, Thomas S, Stoick-Cooper CL, Wang H, Maves L, Sandstrom R, Pabon L, Reinecke H, Pratt G, Keller G, Moon RT, Stamatoyannopoulos J, Murry CE. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell. 2012; 151:221–232. [PubMed: 22981225]

19. Burridge PW, Sharma A, Wu JC. Genetic and epigenetic regulation of human cardiac reprogramming and differentiation in regenerative medicine. Annual Review of Genetics. 2015; 49:461–484.

20. May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Afzal V, Simpson PC, Rubin EM, Black BL, Bristow J, Pennacchio LA, Visel A. Large-scale discovery of enhancers from human heart tissue. Nature Genetics. 2012; 44:89–93.

21. Hu S, Zhao MT, Jahanbani F, Shao NY, Lee WH, Chen H, Snyder MP, Wu JC. Effects of cellular origin on differentiation of human induced pluripotent stem cell-derived endothelial cells. JCI Insight. 2016; 1:e85558. [PubMed: 27398408]

22. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segre AV, Djebali S, Niarchou A, Consortium GT. Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigo R. Human genomics. The human transcriptome across tissues and individuals. Science. 2015; 348:660–665. [PubMed: 25954002]

23. Choi J, Lee S, Mallard W, Clement K, Tagliazucchi GM, Lim H, Choi IY, Ferrari F, Tsankov AM, Pop R, Lee G, Rinn JL, Meissner A, Park PJ, Hochedlinger K. A comparison of genetically matched cell lines reveals the equivalence of human ipscs and escs. Nature Biotechnology. 2015; 33:1173–1181.

24. Coppola CJ, R CR, Mendenhall EM. Identification and function of enhancers in the human genome. Human Molecular Genetics. 2016; 25:R190–R197. [PubMed: 27402881]

25. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics. 2007; 39:311–318. [PubMed: 17277777]

26. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007; 130:77–88. [PubMed: 17632057]

27. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006; 444:499–502. [PubMed: 17086198]

28. Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, Zhang X, Coyne M, Fostel JL, Holmes L, Meldrim J, Guttman M, Epstein C, Park H, Kohlbacher O, Rinn J, Gnirke A, Lander ES, Bernstein BE, Meissner A. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell. 2013; 153:1149–1163. [PubMed: 23664763]

29. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, Yang H, Wang T, Lee AY, Swanson SA, Zhang J, Zhu Y, Kim A, Nery JR, Urich MA, Kuan S, Yen CA, Klugman S, Yu P, Suknuntha K, Propson NE, Chen H, Edsall LE, Wagner U, Li Y, Ye Z, Kulkarni A, Xuan Z, Chung WY, Chi NC, Antosiewicz-Bourget JE, Slukvin I, Stewart R, Zhang MQ, Wang W, Thomson JA, Ecker JR, Ren B. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013; 153:1134–1148. [PubMed: 23664764]

30. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K. Cooperative binding of transcription factors orchestrates reprogramming. Cell. 2017; 168:442–459 e420. [PubMed: 28111071]

31. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011; 471:68–73. [PubMed: 21289626]

32. Burrows CK, Banovich NE, Pavlovic BJ, Patterson K, Gallego Romero I, Pritchard JK, Gilad Y. Genetic variation, not cell type of origin, underlies the majority of identifiable regulatory differences in ipscs. PLoS Genetics. 2016; 12:e1005793. [PubMed: 26812582]

33. Kim TK, Shiekhattar R. Architectural and functional commonalities between enhancers and promoters. Cell. 2015; 162:948–959. [PubMed: 26317464]

34. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W, Diao Y, Liang J, Zhao H, Lobanenkov VV, Ecker JR, Thomson JA, Ren B. Chromatin architecture reorganization during stem cell differentiation. Nature. 2015; 518:331–336. [PubMed: 25693564]

35. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutyavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337:1190–1195. [PubMed: 22955828]

36. Gupta RM, Hadaya J, Trehan A, Zekavat SM, Roselli C, Klarin D, Emdin CA, Hilvering CRE, Bianchi V, Mueller C, Khera AV, Ryan RJH, Engreitz JM, Issner R, Shoresh N, Epstein CB, de Laat W, Brown JD, Schnabel RB, Bernstein BE, Kathiresan S. A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. Cell. 2017; 170:522–533 e515. [PubMed: 28753427]

37. Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D. Gata4 mutations cause human congenital heart defects and reveal an interaction with tbx5. Nature. 2003; 424:443–447. [PubMed: 12845333]

NOVELTY AND SIGNIFICANCE

What Is Known?

- The majority of human genome is comprised of noncoding elements, and protein-coding genes account for less than 2%.

- Genome-wide association studies (GWAS) reveal that over 3 quarters of disease-associated single nucleotide polymorphisms (SNPs) are located in regulatory DNA elements.

- Regulatory DNA elements, such as promoters and enhancers, play a pivotal role in modulating the spatial and temporal gene expression during cardiac development.

What New Information Does This Article Contribute?

- The use of state-of-the-art next-generation sequencing technology (RNA-seq and ChIP-seq) to profile transcriptional and epigenetic changes in regulatory DNA elements of human induced pluripotent stem cells (iPSCs) and somatic cells.

- Cell type-specific gene expression and epigenetic signatures in regulatory elements are dramatically remodeled during cellular reprogramming.

- Cardiac-specific enhancer elements were experimentally validated in transgenic mouse embryos and human cells.

Regulatory DNA elements (promoters and enhancers) mediate cell type-specific gene expression during cardiac development. However, cell type-specific regulatory DNA elements have been largely unknown in the human heart. Here we decipher the epigenetic signatures of regulatory DNA elements in cardiac progenitor cells, endothelial cells, and fibroblasts derived from human fetal heart. We reveal two classes of regulatory DNA elements according to epigenetic marks (H3K4me1 and H3K4me3) enrichment: Class I elements are enriched with these marks in all cell types, whereas Class II elements are labeled in a cell type-specific manner. Class I promoters exhibit stronger transcriptional regulation of the nearby genes, regardless of the presence of distal enhancers. We validate the functions of human enhancer elements in mouse embryos and human cells, and identify enhancer elements that could mediate the cardiac-specific gene expression.
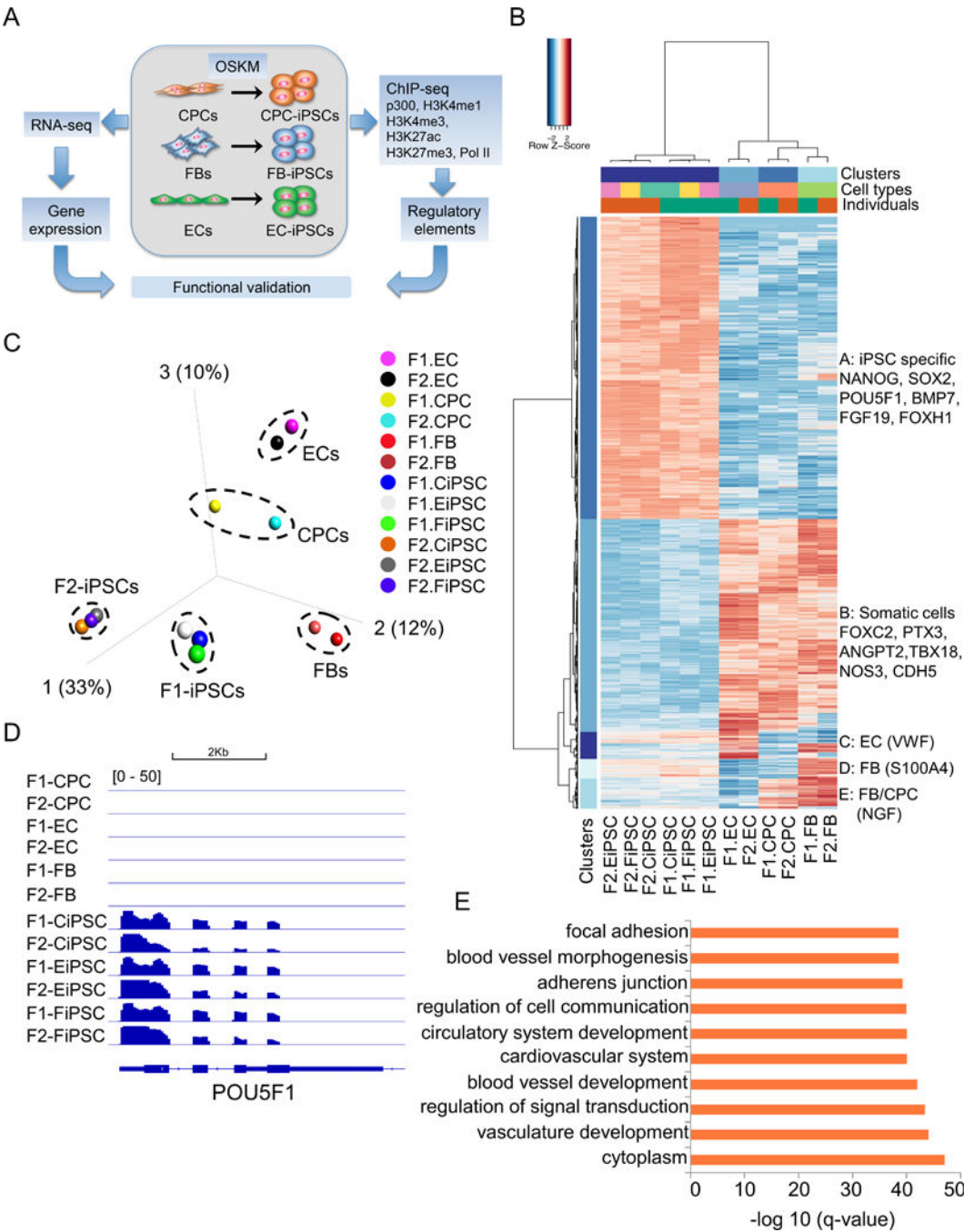
Figure 1. Reprogramming of cell type-specific gene expression into iPSC-specific transcriptional program

(**A**) Schematic diagram of overall experimental design in this study. (**B**) Unsupervised hierarchical clustering of 6,151 differentially expressed genes (DEGs) in human iPSCs and their parental somatic cells (q<0.0001). Cell type-specific gene expression patterns were classified into 5 clusters. Cluster A: iPSC signature genes (3,140); Cluster B: common genes highly expressed in somatic cells but not in iPSCs (2,213); Cluster C: EC-specific genes (279); Cluster D: FB-specific genes (205); Cluster E: genes highly expressed in both FBs

and CPCs (314). (**C**) Principal component analysis (PCA) of somatic cells and their respective iPSCs according to global gene expression profiles. (**D**) *POU5F1* was expressed in all iPSC lines but not in somatic cells. (**E**) Top enriched gene ontology (GO) terms associated with DECs between iPSCs and somatic cells.
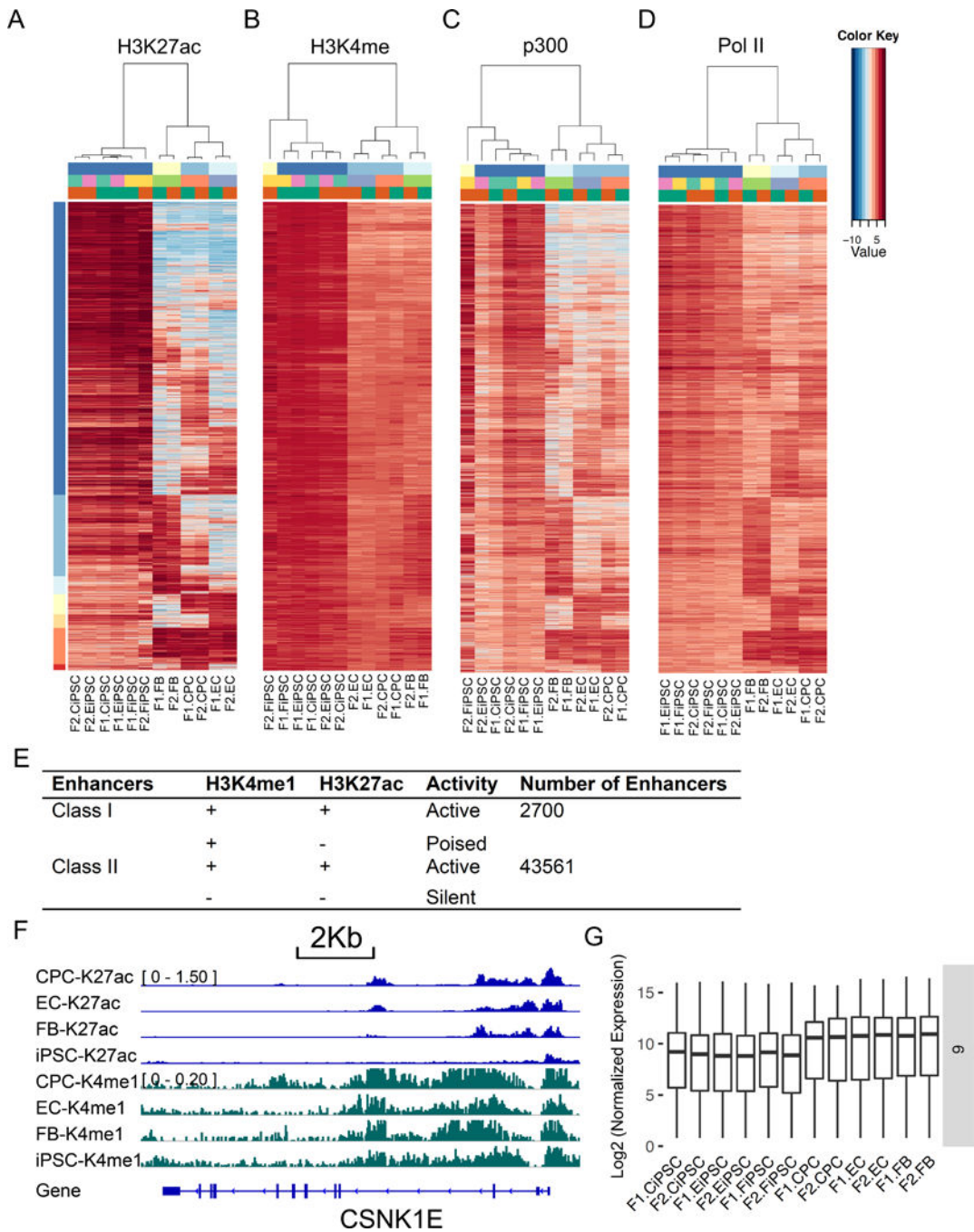
Figure 2. Identification of Class I enhancers in human iPSCs and somatic cells by chromatin marks (H3K4me1 and H3K27ac)

(**A**) Cell type-specific enrichment of active enhancer mark H3K27ac. (**B**) Ubiquitous H3K4me1 enrichment in the same genomic loci surveyed in (A). (**C–D**) Histone acetyltransferase p300 and RNA Pol II were enriched in a similar pattern as active enhancer mark H3K27ac across multiple cell types. (**E**) Categorization of two classes of enhancers according to H3K4me1 and H3K27ac enrichment. (**F**) An example of a putative Class I

enhancer in *CSNK1E* locus. (**G**) Impact of putative Class I enhancers on cell type-specific expression of the nearest genes in cluster 6.
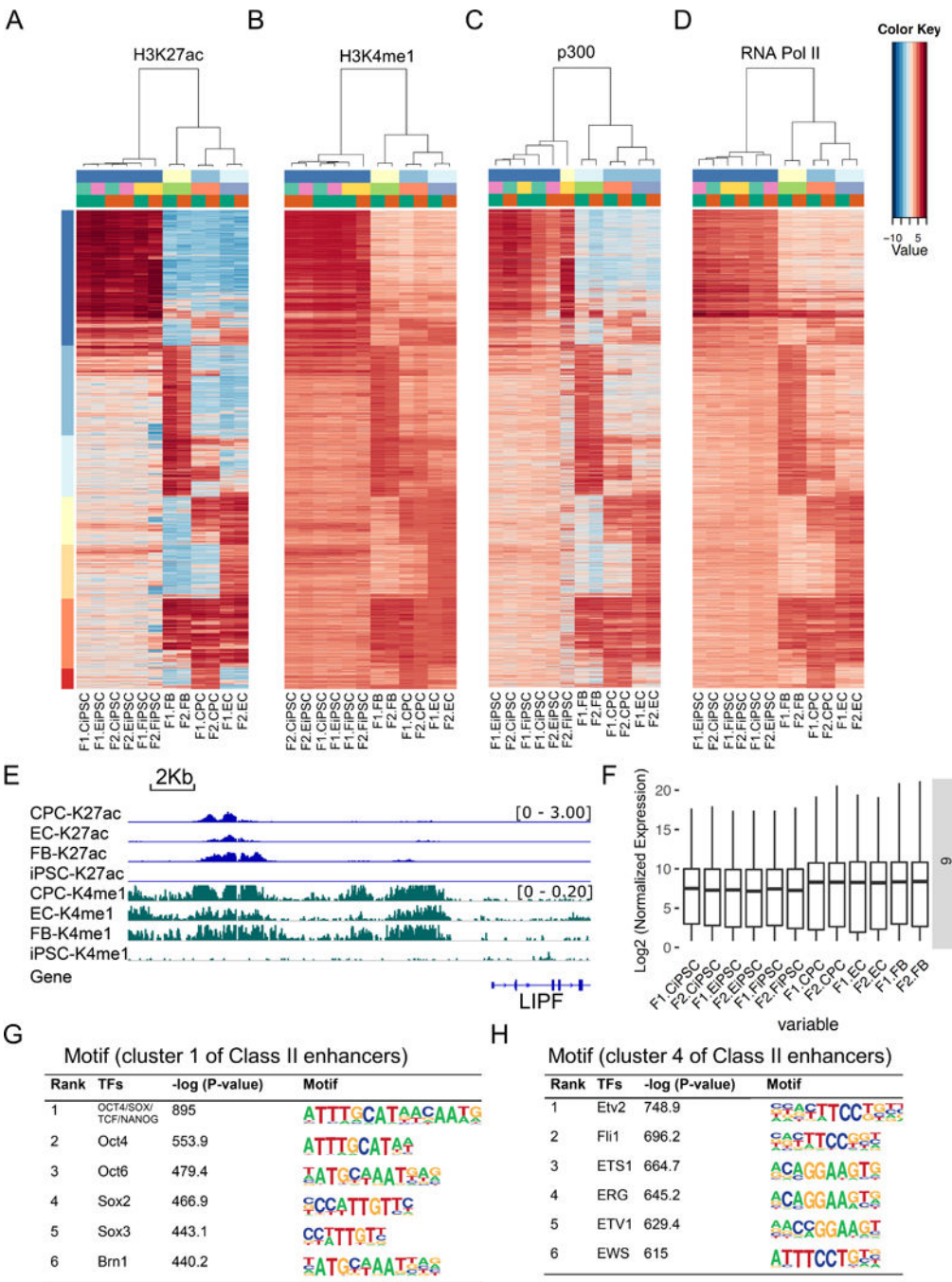
Figure 3. Class II enhancers are identified by cell type-specific H3K4me1 enrichment
(**A**–**B**) Cell type-specific enrichment of H3K27ac and H3K4me1 in Class II enhancers. (**C**–**D**) In Class II enhancers, the genomic distribution of p300 and RNA Pol II was positively correlated with those of H3K27ac and H3K4me1. (**E**) Cell type-specific enrichment of H3K4me1 and H3K27ac in the upstream region of *LIPF*. (**F**) Transcriptional abundance was influenced by the cell type-specific enhancer activity in the cluster 6 genes. (**G**–**H**) Motif enrichment analysis identified binding sites for transcription factors relevant to stem cell

pluripotency (OCT4/SOX2/NANOG) in cluster 1 and endothelial development (ETV2/FLI1/ ETS1) in cluster 4 of Class II enhancers.
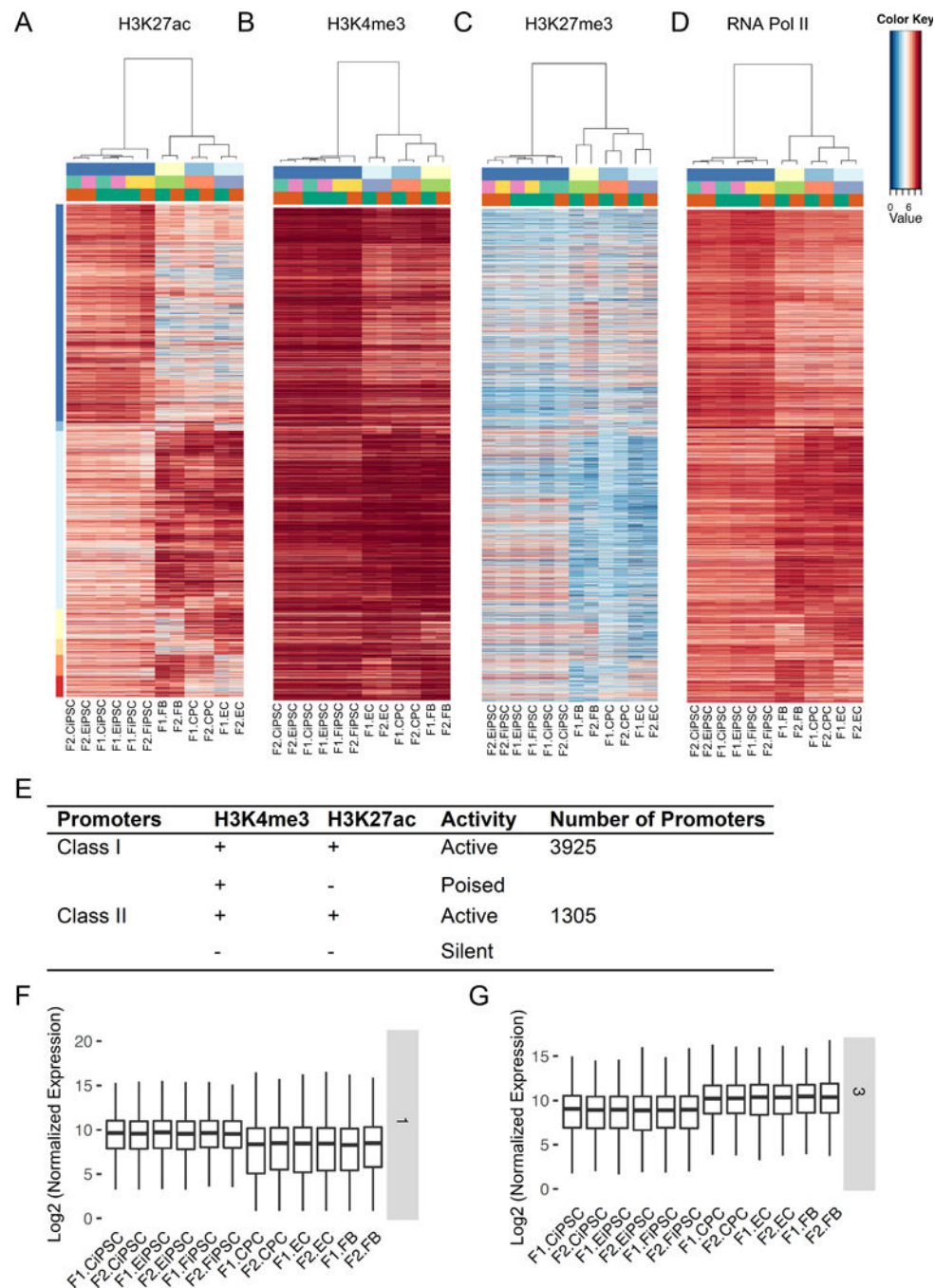
Figure 4. Ubiquitous H3K4me3 promoters were prevalent in human iPSCs and somatic cells
(**A**) Active promoter mark H3K27ac was distributed in a cell type-specific manner. (**B**) The general promoter mark H3K4me3 was omnipresent in all cell types in Class I promoters. (**C**) In contrast, the repressive mark H3K27me3 was negatively correlated with H3K27ac. (**D**) RNA Pol II was positively correlated with active mark H3K27ac. Each line represents the same genomic region enriched by these histone marks and RNA Pol II across multiple cell types. (**E**) Categorization of Class I (3,925, ubiquitous H3K4me3) and Class II promoters (1,305, tissue-specific H3K4me3) according to the distribution of H3K4me3 and H3K27ac.

(**F**–**G**) Cell type-specific gene expression controlled by Class I promoters: iPSC-specific (cluster 1) and somatic cell-specific (cluster 3).
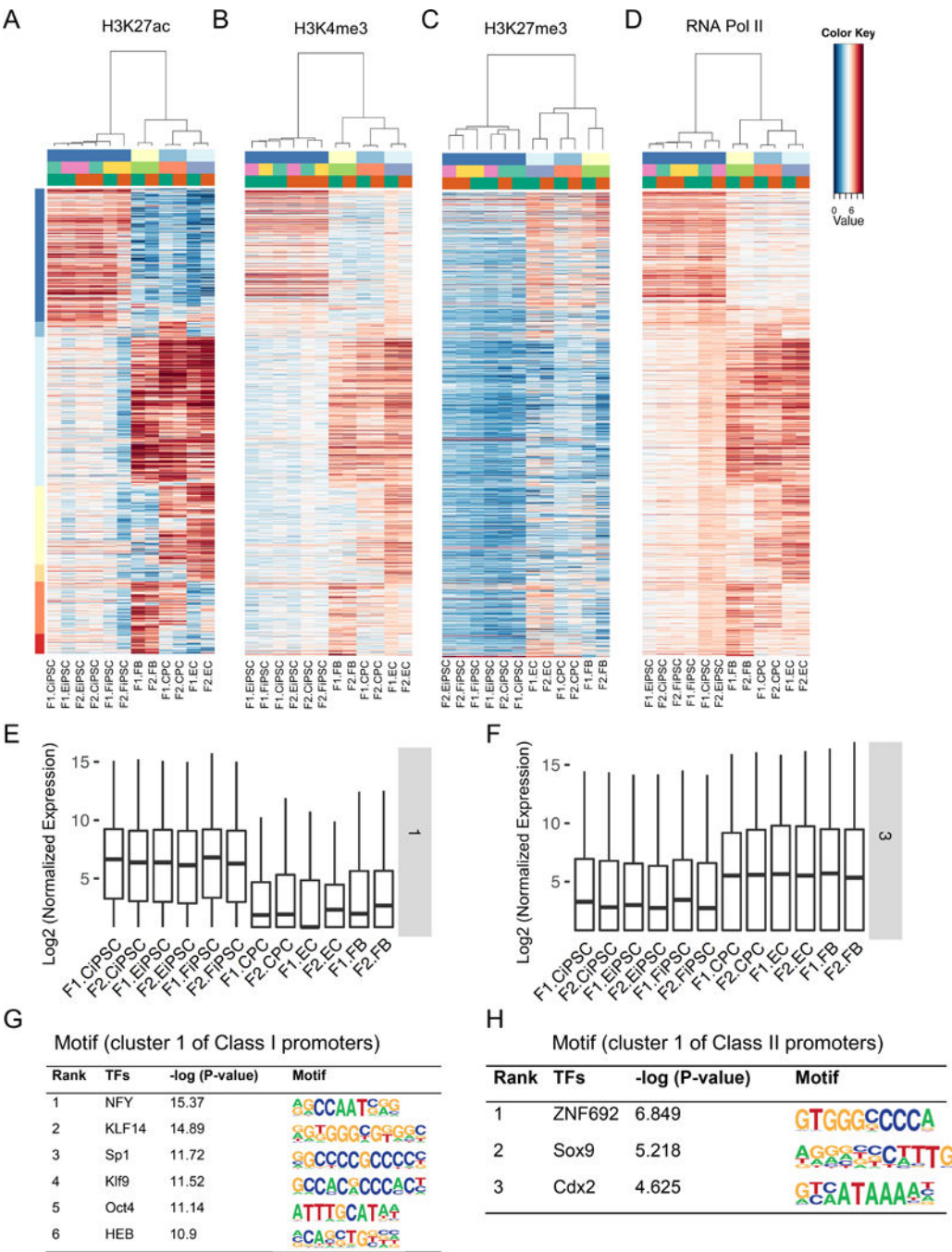
Figure 5. Cell type-specific H3K4me3 promoters and their influence on gene expression
(**A**–**B**) Coordinated enrichment of H3K27ac and H3K4me3 distinguished different cell types by Class II promoters. (**C**) Negative correlation of repressive mark H3K27me3 in these regulatory regions. (**D**) RNA Pol II was positively correlated with the active histone mark H3K27ac. (**E**–**F**) Cell type-specific gene expression controlled by the Class II promoters. Cluster 1 genes were preferentially expressed in iPSCs whereas cluster 3 genes were highly expressed in somatic cells. (**G**–**H**) Differential top enriched TF binding motifs by cluster 1 (iPSC-specific) of Class I and Class II promoters.
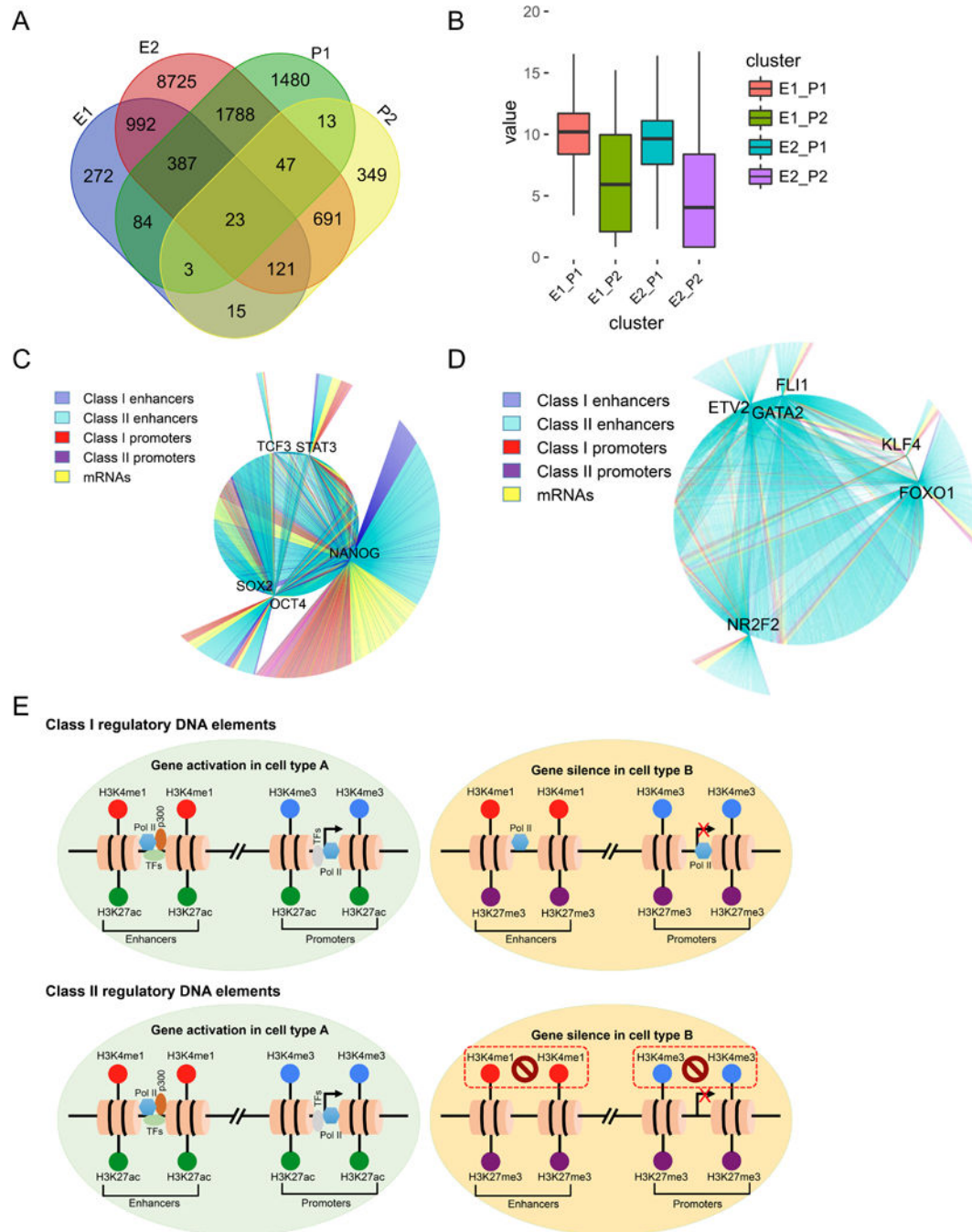
Figure 6. Cell type-specific gene expression modulated by both enhancers and promoters
(**A**) Overlapped genes regulated by both Class I and Class II promoters and enhancers. (**B**) Gene expression levels regulated by the Class I enhancers and Class I promoters (E1_P1), Class I enhancers and Class II promoters (E1_P2), Class II enhancers and Class I promoters (E2_P1), and Class II enhancers and Class II promoters (E2_P2). (**C-D**) Pluripotency-associated transcription factors (OCT4/SOX2/NANOG) tended to interact with all classes of regulatory DNA elements (**C**), whereas endothelial lineage-specific TFs (ETV2/GATA2/NR2F2) were preferentially bound by Class II enhancers (**D**). (**E**) Proposed models for Class

I and Class II regulatory DNA elements. For Class I regulatory elements, H3K4me1 and H3K4me3 are ubiquitously present in distal enhancers and promoters in all cell types, respectively. H3K27ac is enriched in the regulatory elements together with co-factor p300 and RNA Pol II if nearby genes are actively transcribed, whereas H3K27me3 is recruited when target genes are silent in one cell type but are poised for activation upon cellular differentiation or somatic cell reprogramming. In Class II regulatory elements, H3K4me1 and H3K4me3 are only enriched when the target genes are activated in a given cell type. In contrast, these genes are usually repressed in other cell types, where H3K4me1 and H3K4me3 are absent in the regulatory DNA elements that are instead covered by the repressive mark H3K27me3.
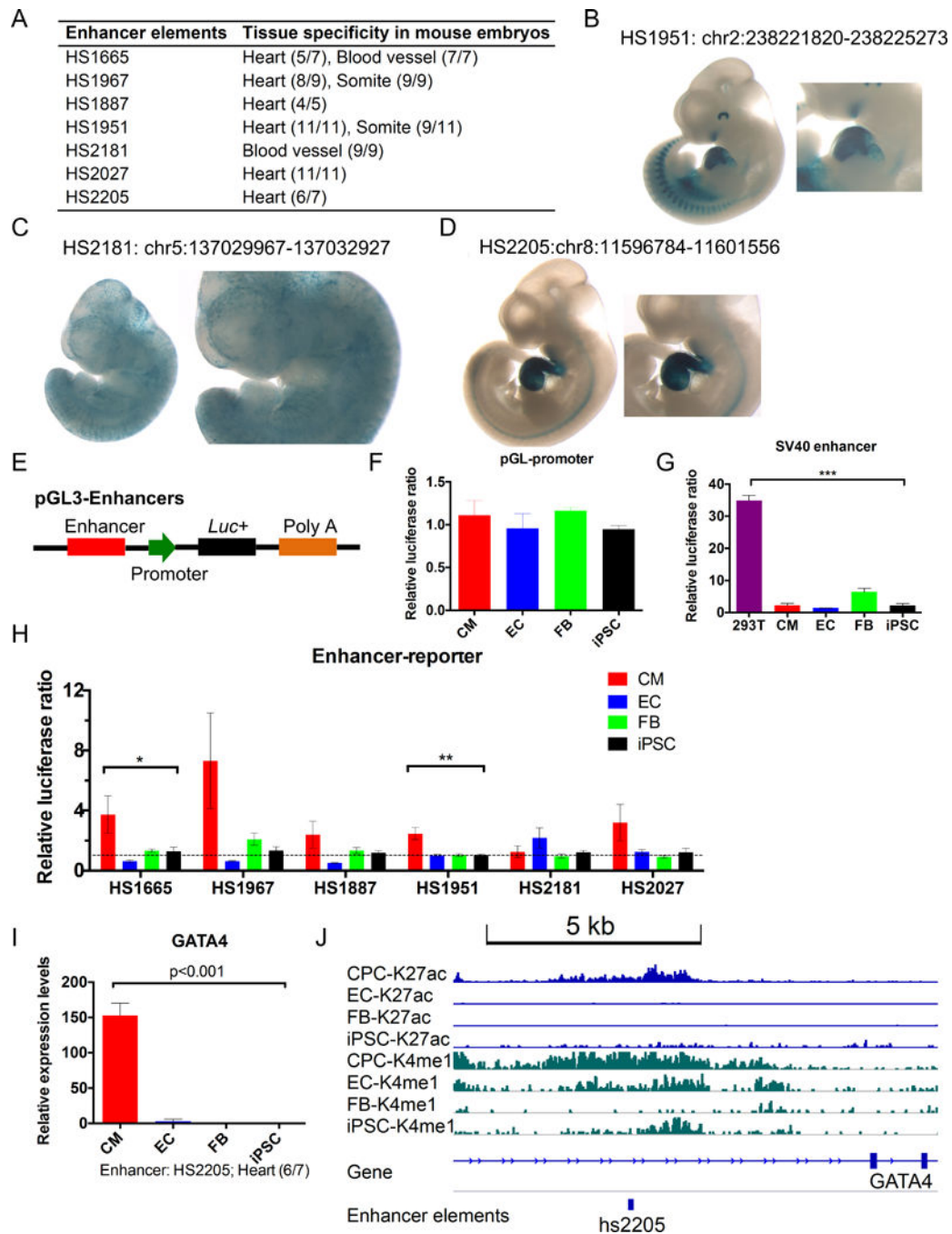
Figure 7. Function validation of putative enhancer elements in human cells and transgenic mouse embryos

(**A**) Tissue specificity of the human enhancer elements measured in transgenic mouse embryos. The numbers in parentheses indicate the reproducibility of *LacZ* reporter staining in mouse embryos. Enhancer elements HS1665 (Cluster 1), HS1967 (Cluster 2), and HS1887 (Cluster 2) are in Class I enhancers, whereas HS1951 (Cluster 3), HS2181 (Cluster 4), HS2027 (Cluster 4), and HS2205 (Cluster 7) are in Class II enhancers. (**B**–**D**) Representative mouse embryos with heart-specific (HS1951, HS2205) and blood vessel-

specific gene expression (HS2181) regulated by evolutionarily conserved human enhancer elements. Additional information with each enhancer can be viewed in the Vista Enhancer Browser. (**E**) Schematic diagram of enhancer reporter constructs using the firefly luciferase (backbone: pGL3-Promoter). (**F**) Relative luciferase activity of pGL3-promoter transfected in multiple human cell types. (**G**) Relative luciferase activity of the SV40 enhancer in human iPSCs, iPSC-derived cardiomyocytes (CMs), ECs (fetal aorta), FBs (fetal heart), and HEK293T cells transfected with pGL3-Control vector containing a SV40 enhancer. (**H**) Relative luciferase activity of 6 enhancer elements in human CMs, ECs, FBs, and iPSCs. Most of these enhancer elements played a cell type-specific regulatory role. (**I**) A representative human enhancer element HS2205 (panel **D**) potentially regulates the cell type-specific mRNA expression of the flanking gene *GATA4* that is specifically expressed in cardiomyocytes. (**J**) Cell type-specific enrichment of H3K27ac and H3K4me1 in human enhancer element HS2205. Error bars show SEM. **$p<0.05$, **$p<0.01$, ANOVA.