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

Ming-Tao Zhao,* Ning-Yi Shao,* Shijun Hu,* Ning Ma, Rajini Srinivasan, Fereshteh Jahanbani, Jaecheol Lee, Sophia L. Zhang, Michael P. Snyder, Joseph C. Wu

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 not well known 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 (chromatin immunoprecipitation combined with high-throughput sequencing; n=84) in human endothelial cells (CD31+CD144+), cardiac progenitor cells (Sca-1+), fibroblasts (DDR2+), and their respective induced pluripotent stem cells. We uncovered 2 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 induced pluripotent stem cells 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. (Circ Res. 2017;121:1237-1250. DOI: 10.1161/CIRCRESAHA.117.311367.)

Key Words: cardiac progenitor cells ■ endothelial cells ■ epigenetics ■ fibroblasts ■ human iPSCs ■ humans ■ stem cells

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.1-5 Human induced pluripotent stem cells (iPSCs) are directly derived from somatic cells by transient overexpression of 4 transcription factors (OCT4/SOX2/C-MYC/KLF4 [POU domain, class 5, transcription factor 1/transcription factor SOX-2/Myc proto-oncogene protein/Kruppel-like factor 4]),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...
Novelty and Significance

What Is Known?
- The majority of human genome is comprised of noncoding elements, and protein-coding genes account for <2%.
- Genome-wide association studies reveal that >3 quarters of disease-associated single nucleotide polymorphisms 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 [chromatin immunoprecipitation combined with high-throughput sequencing]) to profile transcriptional and epigenetic changes in regulatory DNA elements of human induced pluripotent stem cells 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 2 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.

Nonstandard Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPC</td>
<td>cardiac progenitor cell</td>
</tr>
<tr>
<td>EC</td>
<td>endothelial cell</td>
</tr>
<tr>
<td>ESC</td>
<td>embryonic stem cell</td>
</tr>
<tr>
<td>iPSC</td>
<td>induced pluripotent stem cell</td>
</tr>
<tr>
<td>Pol II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>PSC</td>
<td>pluripotent stem cell</td>
</tr>
<tr>
<td>TSS</td>
<td>transcription start site</td>
</tr>
</tbody>
</table>

have emerged as powerful tools to generate human cardiac cells for modeling disease progression and novel drug discovery. Although disease-causing mutations are frequently seen in protein-coding regions across the genome, noncoding sequences, including regulatory DNA elements, have been demonstrated to affect the susceptibility to many cardiovascular diseases. However, evaluating regulatory DNA elements in cardiovascular pathogenesis has been difficult because of 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. A large number of putative enhancers have been identified and comprise at least 12% of the human genome. Cell type–specific enhancers are usually associated with histone modification and higher-order chromatin structure, which can be used to predict putative enhancers in a given cell type. Genome-wide ChIP-seq (chromatin immunoprecipitation combined with high-throughput sequencing) studies reveal that tissue-specific enhancers are enriched with several chromatin marks. 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.

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. Recent studies have illustrated the dynamic and coordinated epigenetic modulation of regulatory DNA elements during cardiac lineage differentiation. Although thousands of tissue-specific enhancers have been identified in the heart using ChIP-seq prediction, 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 [ECs], and cardiac progenitor cells [CPCs]) from the same individuals. We identified 2 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...
Figure 1. Reprogramming of cell type–specific gene expression into induced pluripotent stem cells (iPSC)-specific transcriptional program. A, Schematic diagram of overall experimental design in this study. B, Unsupervised hierarchical clustering of 6151 differentially expressed genes 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 (3140); cluster B, common genes highly expressed in somatic cells but not in iPSCs (2213); cluster C, endothelial cell (EC)-specific genes (279); cluster D, fibroblast (FB)-specific genes (205); cluster E, genes highly expressed in both FBs and cardiac progenitor cells (CPCs; 314). C, Principal component analysis 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 terms associated with DEGs (differentially expressed genes) between iPSCs and somatic cells. ChIP-seq indicates chromatin immunoprecipitation combined with high-throughput sequencing; and OSKM, OCT4/SOX2/C-MYC/KLF4.
from the same fetal heart: fibroblasts, ECs, and Sca-1+ CPCs (Figure 1A).21 The resulting iPSCs were denoted as FiPSCs (fibroblast-derived iPSCs), EiPSCs (endothelial cell-derived iPSCs), and CiPSCs (cardiac progenitor cell-derived iPSCs), respectively. These iPSCs highly expressed OCT4 and NANOG (homeobox protein NANOG; Online Figure IA), with the majority of cells in the colonies being OCT4+NANOG+ (Online Figure IB through ID). 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 fibroblasts (Online Figure IE through IF). 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 transcriptionomes 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 identified (Figure 1B). We further divided these differentially expressed genes into 5 cell type–specific clusters (clusters A through E): 87% (5353 genes, clusters A and B) of differentially expressed genes were between iPSCs and somatic cells, including 279 EC-specific genes (cluster C), 205 fibroblast signature genes (cluster D), and 314 CPC/fibroblast-specific genes (cluster E). We also checked the cell type–specific signature gene expression, discovering that POU5F1 (cluster A) was uniquely expressed in human iPSCs (Figure 1D), CDH5 (cluster B) in somatic cells, VWF (cluster C) in ECs, S100A4 in fibroblasts (cluster D), and GDF6 (cluster E) in fibroblasts and CPCs (Online Figure IIA through IID). Gene ontology analysis showed that these differentially expressed genes 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 principal component analysis 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 III). 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 transcriptionomes 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 2 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), cofactor (p300), and a component of transcriptional machinery (RNA polymerase II [Pol II]). Overall, these chromatin marks and cofactors showed a genomewide 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.23,24 Because we had a variety of conditions (6 cell types) with multiple sets of chromatin marks, we first used H3K27ac to predict all potential enhancers outside of ±3 kb regions of annotated TSSs. In total, we identified 46,261 potential enhancer elements using significantly enriched H3K27ac peaks in at least 1 of our 12 samples. We further divided these potential enhancers into 2 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 (2700) 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 1 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 (Figure 2A and 2B). 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 cofactor p300 and the component of transcriptional machinery Pol II (Figure 2C and 2D), 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 Figure IVA and IVB). We performed correlation analysis and found that 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 2 conditions for class I enhancers: active enhancers with both H3K27ac and H3K4me1 enrichment (H3K27ac+/H3K4me1+) in 1 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 (1–7, from top to 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, cofactor 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 1 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...
Figure 2. Identification of class I enhancers in human induced pluripotent stem cells (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 and 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 2 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. CPC indicates cardiac progenitor cell; EC, endothelial cell; and FB, fibroblast.
profiles of these clusters. We found the average transcription levels of these genes were affected by class I enhancers (Figure 2G; 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 EC function (Online Figure VIA and VIB). 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 (ETS translocation variant), ETV2, and ERG (transcriptional regulator ERG; Online Figure VIC and VID). 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 cofactor p300 and RNA Pol II (Figure 3A through 3D). 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 Figure VIA and VIIB). Next we grouped class II enhancers into 7 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, ≈10% of class II enhancers were situated in gene deserts devoid of protein-coding genes (Online Figure IVΕ), 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, although 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; 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, antiapoptosis, 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 VIIΕ). These results indicate that compared with gene functions of class I enhancers that are mostly associated with cell type identity, Class II enhancers seem 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 transcription factor (TF) motifs in cluster 1 were PSC transcription factors (OCT4, SOX2, and NANOG), whereas those in cluster 4 were relevant to EC lineage determination (Figure 3G and 3H), which were distinct from motifs bound by class I enhancers (Online Figure VIC and VID). Additionally, distinct gene ontology terms were associated with the genes that were potentially regulated by class I and class II enhancers, respectively (Online Figure VIIΙ). 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 with class I enhancers.

**Ubiquitous H3K4me3 Promoters Are Prevalent in Human iPSCs and Somatic Cells**

Because promoters are usually marked by H3K4me3 and located adjacent to the TSSs, we next probed the epigenetic signatures of promoters using H3Kme3, H3K27ac (active), H3K27me3 (repressive), and RNA Pol II (transcription; Figure 4A through 4D). We also interrogated other histone marks (H3K4me1 and p300), but did not find a significant enrichment in the promoter regions (Online Figure IXA and IXB). To exclude any potential enhancers, we only looked into ±3 kb within TSSs. We identified 5,230 promoter regions with differential enrichment of H3K27ac activity between human iPSCs and their parental cells (fibroblasts, ECs, and CPCs). We further divided them into 2 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, ≈75% of these promoters (3925) were class I promoters with ubiquitous H3K4me3 enrichment and cell type–specific distribution of H3K27ac (Figure 4A through 4D; 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 (Figure 4F and 4G; 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 (1305) were enriched with cell type–specific H4K3me3 and H3K27ac and
Class II enhancers are identified by cell type–specific H3K4me1 enrichment. A and B, Cell type–specific enrichment of H3K27ac and H3K4me1 in class II enhancers. C and 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 and H, Motif enrichment analysis identified binding sites for transcription factors relevant to stem cell pluripotency (OCT4/SOX2/NANOG [POU domain, class 5, transcription factor 1/transcription factor SOX-2/homeobox protein NANOG]) in cluster 1 and endothelial development (ETV2/FLI1/ETS1 [ETS translocation variant 2/friend leukemia integration 1 transcription factor/protein C-ets-1]) in cluster 4 of class II enhancers. CPC indicates cardiac progenitor cell; EC, endothelial cell; FB, fibroblast; and iPSCs, induced pluripotent stem cells.

Figure 3. Class II enhancers are identified by cell type–specific H3K4me1 enrichment. A and B, Cell type–specific enrichment of H3K27ac and H3K4me1 in class II enhancers. C and 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 and H, Motif enrichment analysis identified binding sites for transcription factors relevant to stem cell pluripotency (OCT4/SOX2/NANOG [POU domain, class 5, transcription factor 1/transcription factor SOX-2/homeobox protein NANOG]) in cluster 1 and endothelial development (ETV2/FLI1/ETS1 [ETS translocation variant 2/friend leukemia integration 1 transcription factor/protein C-ets-1]) in cluster 4 of class II enhancers. CPC indicates cardiac progenitor cell; EC, endothelial cell; FB, fibroblast; and iPSCs, induced pluripotent stem cells.

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 (Figure 5A through 5D). This cell type–specific enrichment pattern was also observed for histone mark H3K4me1 and cofactor p300 (Online Figure XA and XB).
Figure 4. Ubiquitous H3K4me3 promoters were prevalent in human induced pluripotent stem cells (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 (3925, ubiquitous H3K4me3) and class II promoters (1305, tissue-specific H3K4me3) according to the distribution of H3K4me3 and H3K27ac. F and G, Cell type–specific gene expression controlled by class I promoters: iPSC-specific (cluster 1) and somatic cell-specific (cluster 3). CPC indicates cardiac progenitor cell; EC, endothelial cell; and FB, fibroblast.
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 (Figure 5E and 5F). 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 (Figure 5E and 5F; 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 (Figure 4F and 4G), 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 with enhancers, the TF motif enrichment scores for promoters were much lower, although 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 2 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 2 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 4 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, 2245 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, although the average transcripts were different among individual cell types (Online Figure XIIIB and XIIIC). 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 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 XIIID). 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 (signal transducer and activator of transcription 3), and SOX2 (Figure 6C). However, in ECs, class II enhancers preferentially interacted with endothelial TFs, such as ETV2, NR2F2 (COUP transcription factor 2), and GATA2 (endothelial transcription factor GATA-2; 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 that 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 several 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 (Figure 7B through 7D). 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, ECs (fetal aorta), and fibroblasts (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 cardiomyocytes and ECs (Figure 7H). These results were consistent with the tissue-specific gene expression in the transgenic embryos because these enhancers could modulate the reporter genes specifically in the heart (cardiomyocytes) and blood vessels (ECs; Figure 7A through 7D). 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 with 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 ECs (Online Figure XIIIB and XIIIBIIIB). This prediction was further consolidated with heart-specific reporter gene expression driven by these enhancer elements (Online Figure XIIIB and XIIIBD). The activation of 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 Figure XIIID). 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 2 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
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 through D, Representative mouse embryos with heart-specific (HS1951 and 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 induced pluripotent stem cells (iPSCs), iPSC-derived cardiomyocytes (CMs), endothelial cells (ECs; fetal aorta), fibroblasts (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 (D) potentially regulates the cell type–specific mRNA expression of the flanking gene GATA4 (endothelial transcription factor GATA-4) 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.
iPSCs, whereas cell type–specific H3K4me1 enhancers (class II) reflected cell type–specific gene expression. Likewise, we discovered 2 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 many 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. Although 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. 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. Our study uncovers 2 classes of cell type–specific DNA regulatory elements in human iPSCs and somatic cells. Although 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 2 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. 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, 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. 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, although 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.

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 have identified thousands of human DNA variants associated with complex diseases, the majority of which are noncoding DNA elements. However, the molecular mechanisms of disease-associated loci are rarely illustrated because of the 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 noncoding DNA variants in the contribution to the pathogenesis of cardiovascular diseases. 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. 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 2 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. 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.

Acknowledgments

We thank Larry Bowen, Blake Wu, and Angela Zhang for critical editing of this article. We would like to thank Drs Joanna Wysocka and Tomek Swigut for their suggestions on ChIP-seq experiment and data analysis. We thank the Stanford Center for Genomics and Personalized Medicine for assistance with high-throughput DNA sequencing (supported by the NIH grant S10OD020141).
References


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

Ming-Tao Zhao, Ning-Yi Shao, Shijun Hu, Ning Ma, Rajini Srinivasan, Fereshteh Jahanbani, Jaecheol Lee, Sophia L. Zhang, Michael P. Snyder and Joseph C. Wu

Circ Res. 2017;121:1237-1250; originally published online October 13, 2017; doi: 10.1161/CIRCRESAHA.117.311367

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2017 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/121/11/1237

Data Supplement (unedited) at:
http://circres.ahajournals.org/content/suppl/2017/10/12/CIRCRESAHA.117.311367.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org/subscriptions/
Supplemental Material

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

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

\textsuperscript{1}Stanford Cardiovascular Institute; \textsuperscript{2}Department of Medicine, Division of Cardiology; \textsuperscript{3}Institute of Stem Cell Biology and Regenerative Medicine; \textsuperscript{4}Department of Chemical and Systems Biology; \textsuperscript{5}Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

* Co-first authors. ** Co-corresponding authors.
Methods

Human iPSC derivation. The protocols used in this study were approved by the Institutional Review Board (IRB) at Stanford University. Briefly, human fibroblasts (FBs) were isolated from fetal heart; endothelial cells (ECs) were derived from the aorta;\(^1\) and Sca-1+ cardiac progenitor cells (CPCs) were isolated from the heart of the same human fetuses.\(^2\) These cells (FBs, ECs, and CPCs) were characterized for cell type-specific markers and maintained for about 3 passages. FBs were cultured in DMEM medium (Gibco) supplemented with 10% FBS (Hyclone) and GlutaMax (Gibco). ECs were maintained on 0.1% gelatin-coated (Sigma) plates with EGM-2 medium (Lonza). CPCs were enriched using anti-Sca-1 microbeads (Miltenyi Biotec) and then cultured on 0.1% gelatin-coated dishes in M199 (Gibco)/EGM-2 (3:1) medium supplemented with 10% FBS, 10 ng/ml basic fibroblast growth factor (bFGF, PeproTech), 5 ng/ml epithelial growth factor (EGF, PeproTech), 5 ng/ml insulin-like growth factor (IGF-1, PeproTech), and 5 ng/ml hepatocyte growth factor (HGF, PeproTech). After reaching 60-70% confluent, cells (FBs, ECs, and CPCs) were transduced with lentivirus-based reprogramming vectors encoding four transcription factors (OCT4, SOX2, KLF4, and C-MYC) with pre-determined dilution as previously described.\(^3\) The vectors were removed after 24 hours and replaced with their respective culture media. The iPSC colonies emerged within 5-7 days after viral transduction when the mTeSR-1 medium (STEMCELL Technologies) was introduced. Individual colonies were manually picked under a stereoscope microscope and transferred to Matrigel-coated 12-well plates (Corning) with mTeSR-1 medium. Each colony proliferated and finally formed a stable iPSC line after passage 10. Human iPSCs (2 FB-iPSC lines, 2 EC-iPSC lines, and 2 CPC-iPSC lines) were subjected to RNA-seq and ChIP-seq experiments during passage 23-27.

Cardiac differentiation. Cardiac differentiation was performed using a small molecule-mediated protocol. Human iPSCs were cultured until 80-90% confluent and then treated with 6 \(\mu\)M CHIR-99021 (Selleckchem) for 2 days to induce mesoderm differentiation. Subsequently 5 \(\mu\)M of IWR-1 (Sigma) was added and cells were incubated for another 2 days for further cardiac lineage differentiation. Differentiating cells were maintained in a cardiac differentiation basal medium (RPMI+B27 minus insulin supplement) for 2 days and replaced with a cardiac proliferation medium (RPMI+B27 supplement) for another 4 days. Beating cells were observed at day 9-12 of differentiation. Beating iPSC-CMs were then subjected to a glucose-free RPMI medium (Thermo Fisher Scientific) plus B27 supplement for metabolic purification. After purification, iPSC-CMs were routinely cultured in RPMI medium with B27 supplement.

Immunofluorescence. Cells were first washed using 1X PBS and fixed using 4% paraformaldehyde (PFA) for 15 min at room temperature (RT). Cells were permeabilized by 0.1% Triton X-100 for 20 min at RT followed by washing with 1X PBS twice. Fixed cells were blocked with 10% donkey or goat serum for 1 hr and then incubated with primary antibodies diluted in PBS/0.05% Tween-20 (PBST) at 4°C overnight. Cells were subsequently washed three times using PBST. Next, cells were incubated with the secondary antibodies linked with fluorescence dyes for 1 hr at RT, followed by PBST washing. Nuclei were counterstained with DAPI for 5 min at RT. Cell stains were mounted by the Lab Vision PermaFluor Aqueous Mounting Medium (Thermo Fisher Scientific). Immunofluorescence slides were visualized by Zeiss LSM 510 Meta confocal microscope. The images were processed by ImageJ software (https://imagej.nih.gov/ij/). Antibodies used in this study were Alexa Fluor 488 mouse anti-human OCT3/4 (561628, BD Biosciences), Alexa Fluor 647 mouse anti-human NANOG (561300, BD Biosciences), Alexa Fluor 488 anti-human TRA-1-60-R (330614, BioLegend), Alexa Fluor 647 anti-human SSEA-4 (330408, BioLegend), Alexa Fluor 488 mouse anti-human CD31 (558068, BD Biosciences), Alexa Fluor 647 mouse anti-human CD144 (561567, BD Biosciences), Alexa Fluor 488 mouse IgG-k Isotype Control (557721, BD Biosciences), and Alexa Fluor 647 mouse IgG-k Isotype Control (557714, BD Biosciences).
Flow cytometry. For intracellular proteins, cells were dissociated and filtered through a strainer cap (BD Biosciences). Cells were then resuspended and incubated in BD Cytofix/Cytoperm solution (BD Biosciences) for 20 min at 4°C, followed by washing with 1X BD Perm/Wash buffer (BD Biosciences). Subsequently, they were incubated with primary antibodies for 45 min on ice. Primary antibodies were then removed and washed with 1X BD Perm/Wash buffer twice. Cells were incubated with secondary antibodies for 30 min on ice in dark. The secondary antibodies were eventually washed away using 1X Perm/Wash buffer. Finally, cells were resuspended in FACS buffer (PBS, 1% FBS, and 2mM EDTA) and loaded to BD FACS Aria II for analysis. For membrane proteins, cells were washed with FACS buffer and filtered through a strainer cap. Cells were then incubated with primary antibodies for 30 min on ice and washed with FACS buffer. After secondary antibody incubation, cells were washed, resuspended with FACS buffer, and loaded to BD FACS instrument as described above.

Chromatin immunoprecipitation sequencing (ChIP-seq). Chromatin immunoprecipitation were performed using approximately 1x10^7 cells. Cells were first cross-linked with 1% formaldehyde for 10 min at RT, and formaldehyde was quenched by glycine with a final concentration of 0.125 M. Chromatin was broken into small pieces with an average size of 0.5-2 kb using the Bioruptor (Diagenode). The sonicated chromatin was then incubated with 3-5 µg of primary antibodies overnight at 4°C. A small portion (10%) of chromatin without antibody incubation was kept as input DNA for each ChIP reaction. Subsequently, 75 µl of Dynabeads Protein A or Protein G were added and incubated for 4 hours at 4°C with overhead shaking. Magnetic beads were then washed away and chromatin was eluted. Crosslink was reversed and precipitated DNA was purified and resuspended in nuclease-free water. Sequencing libraries of immunoprecipitated DNA and input DNA were constructed according to an Illumina DNA library preparation protocol. Subsequently, ChIP-seq libraries were loaded to an Illumina HiSeq 2500 platform for deep sequencing. All ChIP antibodies used in this study have been previously reported and individually validated for ChIP-seq experiment:4 p300 (sc-585, Santa Cruz Biotechnology), H3K4me1 (ab8895, Abcam), H3K4me3 (35159, Active Motif), H3K27ac (39133, Active Motif), H3K27me3 (39536, Active Motif), and RNA Polymerase II (05-952, EMD Millipore).

RNA sequencing. RNA-seq libraries were generated using the Encore Complete RNA-seq Library Systems (Part No. 0311, NuGEN). Briefly, 150 ng of DNase-treated RNA input was used for each sample. First-strand cDNA synthesis was carried out by adding the first strand primer mix into 5 µL of total RNA sample in a 96-well PCR plate. The PCR plate was placed in a pre-warmed thermal cycler programed for the primer annealing step at 65°C for 5 min. Subsequently, 3 µL of the first-strand master mix was added, mixed by pipetting and incubated at 40°C for 30 min. Second-strand cDNA was synthesized by adding 65 µL of the second-strand master mix, mixed and incubated at 16°C for 60 min. The double-strand cDNA was subjected to fragmentation using Covaris S-Series sonication system and broken into pieces with a median size of 300 bp. Fragmented cDNA was cleaned using the Agencourt RNAClean XP beads. End repair was done by the addition of end repair master mix into the fragmented cDNA, followed by the incubation at 25°C for 30 min and 70°C for 10 min. For adaptor ligation, 3 µL of the appropriate ligation adaptor mix (L2DR-BC1 through L2DR-BC8) was combined with 14 µL of ligation master mix, and incubated at 25°C for 30 min. Adaptor ligated libraries were then subjected to strand selection I and purification using Agencourt RNAClean XP beads, and strand selection II. Final library amplification was completed by adding 55 µL of the amplification master mix to 25 µL of each strand selection library. The PCR plate was placed in a pre-warmed thermal cycler programmed to run the following program: 5 cycles (94°C, 30 s; 55°C, 30 s; 72°C, 1 min); 15 cycles (94°C, 30 s; 63°C, 30 sec; 72°C, 1 min); and 72°C, 5 min. Amplified libraries were then subjected to bead purification. The quantity and quality of purified RNA-seq Libraries were measured by running the Bioanalyzer DNA Chip 1000. RNA-seq libraries were sequenced in an Illumina HiSeq 2500 platform.

RNA-seq data processing. The raw RNA-seq reads were first aligned to the reference human genome (hg38_P2) using the package tool HISAT (http://www.ccb.jhu.edu/software/hisat/index.shtml).5 After the
alignment, HT-seq (http://htseq.readthedocs.io/en/release_0.9.1/) was applied to the reassembled transcriptomes and reads were assigned to individual genes based on GENCODE genome annotation (GRCh38, version 22). Differentially expressed genes (adjusted P-value <0.00001) were sorted out by using the package DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html). Unsupervised hierarchical clustering was implemented on the differentially expressed genes among different cell types.

**ChIP-seq bioinformatics.** ChIP-seq reads were first aligned to the human genome (hg38_P2) by the Bowtie (http://bowtie-bio.sourceforge.net/index.shtml). The uniquely aligned reads passing quality control were kept. The peak calling was implemented using the histone marks H3K27ac, H3K4me3, and H3K4me1 by the HOMER package (http://homer.ucsd.edu/homer/index.html). H3K27ac and H3K4me1 were used for enhancer prediction. The H3K27ac and H3K4me1 overlapped peaks were annotated with HOMER, and all peaks within 3 kb of the transcription start sites were excluded, and the rest were defined as putative enhancers. K-means clustering was used to group the putative enhancers into seven clusters with cell type-specific features. Each enhancer was assigned to the nearest gene in the genome.

Histone marks H3K27ac and H3K4me3 were used to define the status of promoters. The H3K27ac and H3K4me3 overlapped peaks were annotated with the HOMER. All peaks outside of 3 kb from transcription start sites were excluded. Peaks with differential H3K27ac enrichment in the given cell types were defined as putative cell type-specific promoters. Similar to putative enhancers, K-means clustering was performed to classify these promoters.

For the ChIP-seq data, integrative genome browser (IGV) was used for the visualization of the multiple peak tracks in designated human genomic loci. Functional enrichment (gene ontology terms) analysis was carried out using the GeneAnswers package (https://www.bioconductor.org/packages/release/bioc/html/GeneAnswers.html). The motif enrichment was analyzed by the HOMER package. Gene regulatory networks in multiple cell types were performed and visualized by the Cytoscape (http://www.cytoscape.org/).

**PCA and t-SNE analysis.** Both principle component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) analysis were performed in this study. For RNA-seq, PCA and t-SNE were performed using the log2-transformed values that were normalized by DESeq2. For ChIP-seq, the raw read counts were extracted by HT-seq and then normalized by DESeq2. For PCA plotting, top 500 genes by variance were selected, and the figures were generated by the R package ggplot2 (https://cran.rproject.org/web/packages/ggplot2/index.html). The t-SNE analysis was performed using R package Rtsne (https://cran.r-project.org/web/packages/Rtsne/index.html), and output figures were generated by ggplot2.

**In vivo validation of human enhancer elements.** Human enhancer elements were cloned into a Lac Z reporter vector and injected into mouse zygotes. The detailed experimental procedure was as previously reported (https://enhancer.lbl.gov/aboutproject.html). The tissue-specific gene expression was examined and recorded in the VISTA Enhancer Browser (https://enhancer.lbl.gov/). In human enhancers that were evolutionally conserved and experimentally validated in transgenic mouse embryos, we screened heart and blood vessel-specific enhancers that were also predicted by ChIP-seq data in this study using the Bedtools package (https://github.com/arq5x/bedtools2).

**Luciferase reporter vector construction.** The luciferase reporter vectors pGL3-Promoter (E1761) and pGL3-Control (E1741) were purchased from Promega. Human enhancer elements were amplified from genomic DNA using the PrimeSTAR GXL DNA Polymerase (Takara, R050B). Linker oligos were inserted upstream and downstream of primers for the following fusion cloning. PCR product was purified and cloned into the linearized pGL3-Promoter vector (cut by Nhel and BglIII) using the Quick-Fusion Cloning Kit (Biotool, B2261). The fusion product (pGL3-enhancer) was subsequently transformed into Mix & Go Competent Cells Strain Zymo 5-α (Zymo Research, T3007). Clones were selected by
Ampicillin and plasmids were prepared using the QIAprep Spin Miniprep Kit (Qiagen, 27106). The plasmids sequenced and aligned with the inserted human enhancer elements. The primers with linker sequences are listed in the Online Table 3.

**Transfection and luciferase assays.** Human iPSCs, iPSC-derived cardiomyocytes (CMs), ECs (fetal aorta), FBs (fetal heart), and HEK293T were transfected in a 96-well plate using the Lipofectamine LTX with PLUS Reagent (Invitrogen, 15338030) and Opti-MEM Reduced Serum medium (Invitrogen, 31985-070). On the day of transfection, cell density was 60-80% confluent. For each well, 100 ng of pGL3-enhancer, pGL3-control, or pGL3-promoter was co-transfected with 2 ng of pRL-CMV (Promega, E2261) as an internal control for the normalization of luciferase activity. Cells were incubated with DNA-lipid complex overnight and media was changed to their respective culture medium for another 2 days. The firefly and renilla luciferase activity were measured respectively using a Dual-Glo Luciferase Assay System (Promega, E2920). The ratio of firefly versus renilla luminescence was calculated and normalized to the control samples in each cell type.

**Statistical analysis.** Statistical analyses for each experiment are described in the figure legends or in the appropriate text. Multiple group comparisons were calculated using one-way ANOVA. Pairwise comparisons were carried out using the two-tailed unpaired Student’s t-test. All error bars are defined as standard error of the mean (S.E.M.) unless otherwise indicated.
Online Figure I

A  FiPSCs  CiPSCs  EiPSCs

B  FiPSCs  C  CiPSCs  D  EiPSCs

E  DAPI/NANOG/CD144

F  CD144  CD31

G  Sca1

H  DAPI/DD2/VIMENTIN

I  DDR2  VIMENTIN
Online Figure I. Molecular characterization of human iPSCs and their parental somatic cells (FBs, ECs, and CPCs). (A) Immunofluorescence staining of human iPSCs using monoclonal antibodies against OCT4 and NANOG. (B-D) High percentages of OCT4⁺ NANOG⁺ pluripotent cells in human iPSCs derived from multiple cell types. (E) ECs expressed endothelial surface markers CD31 and CD144. (F) CD31⁺CD144⁺ cells were prevalent in EC culture. (G) Sca1⁺ cells (>98%) in CPCs identified by FACS analysis. (H) Expression of DDR2 and VIMENTIN in human fibroblasts (blue: DAPI, green: DDR2, red: VIMENTIN). (I) FACS analysis showed most human fibroblasts were DDR2⁺ and VIMENTIN⁺. Scale bars: 100 µm.
Online Figure II. Differentially expressed genes (DEGs) identified between human somatic cells and their respective iPSCs. (A-D) Representative gene expression patterns in the clusters of DEGs. Cluster B: somatic cell-specific; Cluster C: EC-specific; Cluster D: FB-specific; Cluster E: FB/CPC-specific. (E) PCA analysis of DEGs in FBs, ECs, CPCs, and iPSCs (q<0.1). Transcriptional variations across different cell types were dominant.
Online Figure III. t-SNE analysis of histone marks, transcription co-factors, and RNA Pol II in human iPSCs and parental somatic cells. Human iPSCs, FBs, ECs, and CPCs were clearly separated based on general enhancer mark H3K4me1 (A), general promoter mark H3K4me3 (B), active chromatin mark H3K27ac (C), repressive chromatin mark H3K27me3 (D), histone acetyltransferase p300 (E), and RNA Polymerase II (F).
Online Figure IV. Chromatin features of Class I enhancers and gene expression. (A-B) Enrichment of active transcription mark H3K4me3 and repressive transcription mark H3K27me3 in Class I enhancers across human iPSCs and somatic cells. (C) The average mRNA abundance of genes regulated by Class I enhancers. Y-axis indicates the log2 values of normalized gene expression. Box plot values show 25% and 75% quartiles of transcriptional levels. The cluster number is shown in the far right column. (D-E) Genomic distribution of Class I and Class II enhancers. Most of Class I enhancers (>70%) were located in gene bodies whereas a substantial portion of Class II enhancers were situated in the gene deserts.
Online Figure V

(A-D) Active histone mark H3K27ac was positively correlated with H3K4me1, p300, and Pol II, but was negatively correlated with the repression mark H3K27me3 in human CPCs (A), ECs (B), FBs (C), and iPSCs (D).

**Online Figure V. Correlation analysis of histone marks, transcription co-factors, and RNA Pol II in the Class I enhancers.**
Online Figure VI. Gene ontology (GO) terms and motif enrichment by Class I enhancers. (A) Gene functions enriched by cluster 1 (iPSC-specific) of Class I enhancers. (B) GO terms enriched by cluster 4 (EC-specific) of Class I enhancers. (C-D) Predicted transcription factor binding motifs in clusters 1 and 4 of Class I enhancers. Class I enhancers were mostly associated with cell lineage determination.
Online Figure VII. Enrichment of histone marks and cell type-specific gene expression affected by Class II enhancers. Cell type-specific enrichment was observed for active transcription mark H3K4me3 (A), but not significant for repressive transcription mark H3K27me3 (B). (C) Class II enhancers directed the cell type-specific gene expression across different cell types. Y-axis indicates the log2 values of normalized gene expression. Box plot values show 25% and 75% quartiles of transcriptional levels. The cluster number is shown in the far right column. (D-E) The GO terms of cluster 1 and cluster 4 genes that were top enriched and potentially regulated by the Class II enhancers.
Online Figure VIII. Distinct biological functions regulated by Class I and Class II enhancers. (A) GO terms (biological process) of genes regulated by Class I enhancers. (B) GO terms of genes modulated by Class II enhancers. Each cluster was enriched by specific GO functional terms that were associated with each cell type, and each class of enhancers apparently regulated distinct biological functions. Values show the enrichment scores (-log (P-value)).
Online Figure IX. Enrichment of histone mark H3K4me1 and transcription co-factor p300 in Class I promoters. (A-B) H3K4me1 and p300 were not differentially enriched in Class I promoters. (C) Cell type-specific gene expression driven by Class I promoters. Median gene expression was shown in each cell type in each cluster. Box plot values show 25% and 75% quartiles of transcriptional levels. Y-axis indicates the log2 values of normalized gene expression. The cluster number is shown in the far right column.
Online Figure X. Enrichment of histone mark H3K4me1 and histone acetyltransferase p300, and gene expression regulated by Class II promoters. (A-B) The distribution of H3K4me1 and p300 exhibited cell type-specific pattern in Class II promoters and was positively correlated with H3K27ac enrichment. (C) Gene transcription was significantly affected by Class II promoters, with low levels of transcription being observed when Class II promoters were repressed in a given cell type. Median gene expression was shown in each cell type in each cluster. Box plot values show the 25% and 75% quartiles of transcriptional levels. Y-axis indicates the log2 values of normalized gene expression.
Online Figure XI. Gene ontology (GO) terms enriched by the relevant genes that were potentially regulated by Class I and Class II promoters. (A-B) Class I and Class II promoters regulated genes associated with distinct cellular and biological functions. Values show the enrichment scores (-log (P-value)) and the degree of darkness in each dot indicates the number of relevant genes.
Online Figure XII. Common genes regulated by both Class I and II regulatory DNA elements. (A) Venn diagram of the number of common genes regulated by the combination of promoters and enhancers. (B) Box plot of gene expression levels potentially regulated by Class I enhancers together with Class I promoters (upper) or Class II promoters (bottom). (C) Transcriptional levels potentially modulated by Class II enhancers combined with Class I promoters (upper) or Class II promoters (bottom). Class I promoters drove a higher gene expression than Class II promoters, regardless of the combination with enhancers. Median gene expression was shown in each cell type. Box plot values show the 25% and 75% quartiles of transcriptional levels. Y-axis indicates the log2 values of normalized gene expression. (D) Common genes regulated by Class I promoters exhibited higher enrichment scores than those regulated by Class II promoters, regardless of distal enhancer activity. Values show the enrichment scores (-log (P-value)) and the degree of darkness in each dot indicates the number of relevant genes.
Online Figure XIII. Prospective gene regulation by the cell type-specific enhancers. (A) The flanking gene *TEAD3* that was preferentially expressed in cardiomyocytes and potentially regulated by the enhancer element HS1887. (B) *TANC1* (adjacent to enhancer HS2027) was highly expressed in both CMs and ECs. (C-D) Representative mouse embryos injected with human enhancer elements displayed tissue-specific gene expression in the heart (Vista Enhancer Brower). (E) Histone marks (H3K27ac and H3K4me1) enrichment in human enhancer elements HS2027. H3K27ac was highly enriched in CPCs and ECs but absent in iPSCs and FBs.
Online Table 1. Classification of Class I and Class II enhancers and their cell type-specific activities.

<table>
<thead>
<tr>
<th>Class</th>
<th>Cluster</th>
<th>Number of Enhancers</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>Cluster 1</td>
<td>1,686 (62.4%)</td>
<td>iPSC</td>
</tr>
<tr>
<td></td>
<td>Cluster 2</td>
<td>469 (17.4%)</td>
<td>iPSC and FB</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>104 (3.85%)</td>
<td>FB and CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>115 (4.26%)</td>
<td>CPC and EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 5</td>
<td>80 (2.96%)</td>
<td>EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 6</td>
<td>209 (7.74%)</td>
<td>FB, EC, and CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 7</td>
<td>37 (1.37%)</td>
<td>CPC</td>
</tr>
<tr>
<td>Class I</td>
<td>Total: 2,700</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class II</td>
<td>Cluster 1</td>
<td>12,304 (28.2%)</td>
<td>iPSC</td>
</tr>
<tr>
<td></td>
<td>Cluster 2</td>
<td>8,182 (18.8%)</td>
<td>FB</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>5,573 (12.8%)</td>
<td>FB and CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>4,374 (10.0%)</td>
<td>CPC and EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 5</td>
<td>4,888 (11.2%)</td>
<td>EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 6</td>
<td>6,378 (14.6%)</td>
<td>FB, EC, and CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 7</td>
<td>1,862 (4.27%)</td>
<td>CPC</td>
</tr>
<tr>
<td>Class II</td>
<td>Total: 43,561</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Online Table 2. Categorization of Class I and Class II promoters and their cell type-specific activities.

<table>
<thead>
<tr>
<th>Class</th>
<th>Cluster</th>
<th>Number of Promoters</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>Cluster 1</td>
<td>1,728 (44.0%)</td>
<td>iPSC</td>
</tr>
<tr>
<td>Total: 3,925</td>
<td>Cluster 2</td>
<td>78 (2.0%)</td>
<td>CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>1,414 (36.0%)</td>
<td>FB, EC, and CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>239 (6.1%)</td>
<td>CPC and EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 5</td>
<td>129 (3.3%)</td>
<td>EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 6</td>
<td>167 (4.3%)</td>
<td>CPC and FB</td>
</tr>
<tr>
<td></td>
<td>Cluster 7</td>
<td>170 (4.3%)</td>
<td>FB</td>
</tr>
<tr>
<td>Class II</td>
<td>Cluster 1</td>
<td>373 (28.6%)</td>
<td>iPSC</td>
</tr>
<tr>
<td>Total: 1,305</td>
<td>Cluster 2</td>
<td>43 (3.3%)</td>
<td>CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>418 (32.0%)</td>
<td>FB, EC, and CPC</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>220 (16.9%)</td>
<td>CPC and EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 5</td>
<td>48 (3.7%)</td>
<td>EC</td>
</tr>
<tr>
<td></td>
<td>Cluster 6</td>
<td>147 (11.3%)</td>
<td>CPC and FB</td>
</tr>
<tr>
<td></td>
<td>Cluster 7</td>
<td>56 (4.3%)</td>
<td>FB</td>
</tr>
</tbody>
</table>
Online Table 3. Primer sequence information for the construction of enhancer-reporter vectors.

<table>
<thead>
<tr>
<th>Enhancer elements</th>
<th>Sequence (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS1665_Foward</td>
<td>cgagctttaaegegtctagcccccctcttgaaattgtgc</td>
</tr>
<tr>
<td>HS1665_Reverse</td>
<td>tgagatgeagatgeagatcttttttacccctgctgcctt</td>
</tr>
<tr>
<td>HS1967_Foward</td>
<td>cgagctttaaegegtctagccacctgctcagtggggatcagag</td>
</tr>
<tr>
<td>HS1967_Reverse</td>
<td>tgagatgeagatgeagatctgaaccacacacacagagatgag</td>
</tr>
<tr>
<td>HS1887_Foward</td>
<td>cgagctttaaegegtctagcagtgagtgcttcgcttgctt</td>
</tr>
<tr>
<td>HS1887_Reverse</td>
<td>tgagatgeagatgeagatctcttgagggctgagtctccac</td>
</tr>
<tr>
<td>HS1951_Foward</td>
<td>cgagctttaaegegtctagctgtccctgtctcaagaagcg</td>
</tr>
<tr>
<td>HS1951_Reverse</td>
<td>tgagatgeagatgeagatctggaggctgagtctccacctact</td>
</tr>
<tr>
<td>HS2181_Foward</td>
<td>cgagctttaaegegtctagctgtcaagctgaaggaaaaacag</td>
</tr>
<tr>
<td>HS2181_Reverse</td>
<td>tgagatgeagatgeagatctgtagtgactgcaccact</td>
</tr>
<tr>
<td>HS2027_Foward</td>
<td>cgagctttaaegegtctagcagcttccctttgtaaaacccttgagatgeagatgtctcttgcttgcttgctt</td>
</tr>
<tr>
<td>HS2027_Reverse</td>
<td>tggagatgeagatgcagatctcagcacaattcttggctctg</td>
</tr>
</tbody>
</table>
References


