HHS Public Access

Author manuscript

Circ Res. Author manuscript; available in PMC 2018 August 04.

Published in final edited form as:

Circ Res. 2017 August 04; 121(4): 376–391. doi:10.1161/CIRCRESAHA.116.310456.

Genome-Wide Temporal Profiling of Transcriptome and Open-Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs

Qing Liu[1], **Chao Jiang**[1], **Jin Xu**[2], **Ming-Tao Zhao**[3], **Kevin Van Bortle**[1], **Xun Cheng**[4], **Guangwen Wang**[4], **Howard Y. Chang**[2], **Joseph C. Wu**[3], and **Michael P. Snyder**[1]

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

[2]Center for Personal Dynamic Regulomes and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

[3]Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, California 94305, USA

[4]Stem Cell Core Facility, Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA

Abstract

Rationale—Recent advances have improved our ability to generate cardiomyocytes from human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs). However, our understanding of the transcriptional regulatory networks underlying early stages (*i.e.* from mesoderm to cardiac mesoderm) of cardiomyocyte differentiation remains limited.

Objective—To characterize transcriptome and chromatin accessibility during early cardiomyocyte differentiation from hiPSCs and hESCs.

Methods and Results—We profiled the temporal changes in transcriptome and chromatin accessibility at genome-wide levels during cardiomyocyte differentiation derived from two hiPSC lines and two hESC lines at four stages: pluripotent stem cells, mesoderm, cardiac mesoderm, and differentiated cardiomyocytes. Overall, RNA-seq analysis revealed that transcriptomes during early cardiomyocyte differentiation were highly concordant between hiPSCs and hESCs, and clustering of four cell lines within each time-point demonstrated that changes in genome-wide chromatin accessibility were similar across hiPSC and hESC cell lines. Weighted gene co-expression network analysis (WGCNA) identified several modules that were strongly correlated with different stages of cardiomyocyte differentiation. Several novel genes were identified with high weighted-connectivity within modules and exhibited co-expression patterns with other genes,

Correspondence to: Michael P. Snyder, Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, mpsnyder@stanford.edu.

In May 2017, the average time from submission to first decision for all original research papers submitted to *Circulation Research* was 12.28 days.

This article was sent to Elizabeth McNally, Consulting Editor, for review by expert referees, editorial decision, and final disposition.

Disclosures: HYC is a founder and consultant of Epinomics.

including non-coding RNA *LINC01124* and uncharacterized RNA *AK127400* in the module related to the mesoderm stage; and *ZEB1* in the module correlated with post-cardiac mesoderm. We further demonstrated that ZEB1 is required for early cardiomyocyte differentiation. In addition, based on integrative analysis of both WCGNA and TF-motif enrichment analysis, we determined numerous TFs likely to play important roles at different stages during cardiomyocyte differentiation, such as *T* and *EOMES* (mesoderm); *LEF1* and *MESP1* (from mesoderm to cardiac mesoderm); *MEIS1* and *GATA4* (post-cardiac mesoderm); *JUN* and *FOS* families, and *MEIS2* (cardiomyocyte).

Conclusions—Both hiPSCs and hESCs share similar transcriptional regulatory mechanisms underlying early cardiac differentiation, and our results have revealed transcriptional regulatory networks and new factors (*e.g.* ZEB1) controlling early stages of cardiomyocyte differentiation.

Keywords

stem cell differentiation; cardiomyocyte; early stage; transcriptome; chromatin accessibility

Introduction

Human pluripotent stem cells, including both human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs), serve as an *in-vitro* cell-based model for investigations of degenerative diseases and predictive developmental toxicology in humans.[1-3] Differentiation of cardiomyocytes from hiPSCs and hESCs is of particular interest for a multitude of reasons. As a “heart-disease model in a dish”, this system provides great opportunities and advantages in the study of cardiac diseases and the evaluation of drug toxicity in cardiac tissue.[4, 5] Recent studies on cardiomyocyte differentiation using human and mouse pluripotent cell lines have primarily focused on transcriptional regulation of the later transition from cardiac progenitors to differentiated cardiomyocytes.[6, 7] However, transcriptomic analysis with genome-wide chromatin accessibility profiling at earlier stages (*i.e.* from mesoderm to cardiac mesoderm) have not been fully documented, despite the fact that cardiac mesoderm formation leads to the differentiation of cardiac progenitors.

In order to fill the knowledge gap regarding the transcriptional regulation underlying early cardiomyocyte differentiation, we profiled dynamic changes in the transcriptome and open chromatin states during different stages of cardiac differentiation. Two parallel genomic assays were used: RNA sequencing (RNA-seq) to evaluate the temporal changes in transcription, and the recently developed assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)[8] to investigate genome-wide chromatin accessibility. We analyzed both hiPSCs and hESCs in this study, and observed that these cell lines exhibited high concordant transcriptomes during cardiac differentiation. Weighted gene co-expression network analysis (WGCNA)[9] revealed stage-specific gene co-expression modules and genes with high connectivity within the modules, identifying novel genes that likely play key roles in the differentiation process. Moreover, integrative analysis of transcription factor (TF) DNA-binding specificities (motifs) analysis with coordinated network analysis delineated stage-specific TFs that are likely to play important roles in different stages of the differentiation.

Methods

Cell culture and cardiomyocyte differentiation

Two hESC lines (H1 and H9) and two hiPSC lines (C15 and C20) were used in this study. They were obtained from the Stanford Cardiovascular Institute (SCVI) Biobank and the Stem Cell Core Facility of Genetics, Stanford University. The C15 and C20 hiPSCs were generated with lentivirus from skin fibroblasts of anonymous healthy persons. All pluripotent cell lines were grown in Matrigel (Corning)-coated 12-well plates in Essential 8™ Medium (Thermo Fisher Scientific) at 37 °C incubators (5% CO_2). Cardiomyocyte differentiation was initiated using a monolayer differentiation method with a PSC Cardiomyocyte Differentiation kit (Thermo Fisher Scientific) according to the manufacturer's instructions. To further increase cardiomyocyte purity, the differentiated cells were subjected to subsequent glucose starvation using non-glucose-supplemented RPMI/B27 medium for three times (two days per time) to decrease non-cardiomyocyte cells, since cardiomyocytes are more tolerant to glucose starvation.[10]

On days 0, 2, 4 and 30 during the differentiation period (before the medium-change at that day), cells were collected using Accutase (Thermo Fisher Scientific). For each cell line and each time-point, cells from two independent differentiation wells were used as two biological replicates, and these collected cells were both used for RNA-seq and ATAC-seq experiments. Thus, a total of 32 samples were collected both for RNA-seq and ATAC-seq. Library preparation and data analyses for RNA-seq and ATAC-seq, *de novo* motif analysis, computational footprinting of TF-binding analysis, immunostaining, flow cytometry and analysis, siRNA transfection, real-time QPCR and western blot are described in the Online Data Supplement available at http://circres.ahajournals.org.

Data availability

Both RNA-seq and ATAC-seq data generated for this work have been deposited in NCBI's Gene Expression Omnibus (GEO), and they are accessible through GEO SuperSeries accession number GSE85332: GSE85331 for the RNA-seq, and GSE85330 for the ATAC-seq. The sequencing depth, quality control (QC), and mapping reads are shown in Online Tables II and III.

Results

Cardiomyocyte differentiation from pluripotent stem cells

In the present study, we used a monolayer-differentiation method (Figure 1A) that reproducibly generates cardiomyocytes from both hiPSCs and hESCs, yielding a population of 85-95% cardiomyocytes (Online Figure II). During our optimization process we found that the initiation of cardiac differentiation depended upon confluency of the pluripotent cells and varied with different lines. For instance, the greatest confluency of initiation of cardiac differentiation in the C15 cell line was between 50%-65%. The best initial confluency of H1, H9 and C20 lines for cardiac differentiation was 60-70%, 70-80%% and 80-85%, respectively.

Two days after the initiation of differentiation, cells expressed the Brachyury protein (encoded by gene *T*) (Figure 1B), which is a typical marker for mesoderm. Four days after initiation of differentiation, cells expressed mesoderm posterior BHLH transcription factor 1 (MESP1), a definitive marker for cardiac mesoderm.[11] The protein of NK2 homeobox 5 (NKX2-5), a marker for cardiomyocyte progenitor, was not observable on day 4 by immunostaining (Figure 1B), its gene expression on day 4 was also low (FPKM<0.5) based on RNA-seq data. ISL LIM homeobox 1 (*ISL1*), another marker for cardiomyocyte progenitor,[12] showed a modest level of gene expression on day 4 (FPKM ranges from1.6-6.7; see Online Data I), suggesting that the day 4 is a time-point of transition from cardiac mesoderm to cardiomyocyte progenitors. Beating cardiomyocytes were readily observed between 7 and 10 days after initiation of differentiation (Online Videos I-IV). At 30 days, both NKX2-5 and Troponin T type 2, cardiac (TNNT2) were strongly expressed (Figure 1B). These results suggest that these four cell lines exhibited the same kinetics in cardiomyocyte differentiation.

Concordant transcriptomic dynamics of cardiomyocytes differentiation derived from hiPSCs and hESCs

To understand changes in global gene expression during cardiac cell differentiation after derivation from four pluripotent stem lines, we performed RNA-Seq experiments on days 0, 2, 4, and 30 followed by clustering analysis, principal component analysis (PCA), and correlation analysis.

More than 15,000 genes were used for WGCNA after filtration based on the FPKM values, leading to 11 modules according to their co-expression patterns (Figure 2A; and Online Data I), and the significances of correlations between “module and factors” were calculated based on their correlation with two factors: “days” and “cell lines”. The modules in WGCNA were assigned based on hierarchical clustering using topological overlap dissimilarity score, which was calculated based on adjacency scores. Adjacency score was calculated as a power function of correlations between individual gene expression profiles across samples. Based on the clustering analysis of WGCNA, the expression of all transcripts showed highly concordant transcriptomic profiling among four cell lines during cardiomyocyte differentiation (Figure 2A). Genes within each module were enriched with a distinct array of biological functions (Online Table IV), and several modules were strongly correlated with differentiation transitions from mesoderm to differentiated cardiomyocytes. For example, genes in module 1 were highly expressed only in differentiated cardiomyocytes (day 30), and their enriched GO terms of biological process (BP) were related to heart functions such as “regulation of heart contraction” and “muscle system process”; module 2 contains genes that were primarily up-regulated during differentiation from cardiac mesoderm (day 4) to differentiated cardiomyocytes, and their functions comprised of “skeletal system development” and “heart morphogenesis”; module 4 contains genes that were highly expressed on days 2 (mesoderm), and module 8 contains genes that were highly expressed on days 2 and 4 (from mesoderm to cardiac mesoderm). Genes in both modules 4and 8 are mostly involved in “anterior/posterior pattern specification” and “pattern specification process”. Interestingly, genes assigned to module 3 were only highly expressed in pluripotent cells (day 0), and their functions were primarily related to “mitotic cell cycle”;

this is expected since stem cell self-renewal involves cell cycle and cell proliferation genes.[13] In addition, genes in module 6 were highly expressed on day 0, and genes in module 5 were highly expressed on day 0 and day 30, and their enriched GO terms (BP) were related to energy production, such as “mitochondrial translation” and “mitochondrial ATP synthesis coupled electron transport”, suggesting that both pluripotent cells and cardiomyocytes consumed large amount of energy for maintaining pluripotency in stem cells or muscle contraction in cardiomyocytes.[14, 15] Conversely, we observed a small number of genes (1.6% of total transcripts) in modules 9 10 and 11 that exhibited cell-line-specific expression patterns; however, no significantly enriched GO terms were observed.

The segregation of samples by time-points was also observed using PCA analysis (Figure 2B), samples on day 2 were more similar to those on day 4 when compared to other time-points, indicating that global gene expression shares some similarities between mesoderm stage and cardiac-mesoderm stage. Although correlations of expression values (*i.e.* FPKM) of all transcripts (between replicates, hiPSCs, hESCs, and hiPSCs and hESCs) showed slight time-course decreases, the overall correlation coefficients were higher than 0.75, indicating both hiPSCs and hESCs shared similarity in global genes expression (Online Figure III).

GO enrichment analysis of the differential genes for each cell line exhibited similar dynamic changes from the undifferentiated stage to differentiated cardiomyocytes (Figure 2C). Enriched GO terms (BP) included “cell cycle” and “mitosis” of undifferentiated cells (day 0); embryonic development-related GO terms (such as “anterior/posterior pattern formation” and “pattern specification process”) for the mesoderm (day 2) to cardiac mesoderm transition (day 4); and cardiac development- and cardiomyocyte function-related terms from cardiac mesoderm to cardiomyocytes (day 30) (*e.g.*, “heart development”, “heart morphogenesis” and “regulation of heart contraction”). The representative genes assigned to these GO terms included *NANOG* and *SOX2* (known TFs in undifferentiated stem cells);[16] *T*, eomesodermin (*EOMES*), Wnt family member 3A (*WNT3A*), bone morphogenetic protein 2 (*BMP2*), MESP1 and mix paired-like homeobox 1 (*MIXL1*) (which are known key TFs and factors for mesoderm development);[11, 17-21] *NKX2-5*, GATA binding protein 4 (*GATA4*), T-box transcription factor 5 (*TBX5*), heart and neural crest derivatives expressed 2 (*HAND2*) and *ISL1* (which are known TFs regulating cardiac development);[6, 12, 22] and *TNNT2*, troponin I, cardiac muscle (*TNNI3*), myosin light chain 3 (*MYL3*) and calcium/calmodulin-dependent protein kinase II delta (*CAMK2D*) for cardiac structure and heart contraction[23-26] (Figure 2C). The similar dynamic changes in enriched GO terms are correlated with the concordant transcriptome profiles of the four cell lines, and also reflect features of the cardiac differentiation that were observed in previous studies in other model systems.

Modules identified by WGCNA exhibits correlation with stages of cardiomyocyte differentiation

Four modules (*i.e.* networks) identified by WGCNA were strongly related to differentiation transitions from mesoderm to differentiated cardiomyocytes based on their temporal gene-expression patterns (in Figure 2A) and the GO enrichment analysis (see Online Table IV), including: module 4 (mesoderm-related), module 8 (mesoderm to cardiac mesoderm),

module 2 (post-cardiac mesoderm), and module 1 (cardiomyocyte-related) (Figure 3A). In order to better explore the modules, we visualized these modules using Gephi (0.9.1) (Online Figure IV A-D). Since each module contains hundreds to thousands of genes, we presented each module using genes with top weighted connectivity in Figures 3B-3E. The size of each node represents the "weighted connectivity" of this gene within the module, which is the sum of adjacency scores (*i.e.* connection strength) between this gene and all other genes in the network, therefore genes with high connectivity can be inferred as playing important roles in this module.[9]

Module 4 (Figure 3B and Online Figure IV A) is correlated to mesoderm stage based on the enriched GO terms of genes in this module. Many known TFs assigned to this module play important roles in mesoderm development, such as *T*, *EOMES*, *MIXL1*, *WNT3A*, Dickkopf Wnt signaling pathway inhibitor 1 (*DKK1*), Sp5 transcription factor (*SP5*), mesogenin 1 (*MSGN1*), even-skipped homeobox 1 (*EVX1*) and caudal type homeobox 2 (*CDX2*).[18-21, 27-31] We also observed that two transcripts which have not been well characterized before, *AK127400* (coded by *LOC100130256*) and *LOC440925* (known as "long intergenic non-protein coding RNA1124", *LINC01124*), had highly-weighted connectivity within the module 4. Interestingly, *LINC01124* and *AK127400* are juxtaposed on chromosome 2 and surrounded by *SP5*, *MYO3B* (myosin IIIB) and *GAD1* (glutamate decarboxylase 1) as a gene cluster (Online Figure V A), which were also assigned into the module 4. In our study, these genes were all temporally overexpressed at mesoderm stages (day 2) and assigned to the module 4, which was related to mesoderm formation. Based on siRNA-mediated gene knockdown of *AK127400,* we observed decreases in expression of *LINC01124*, *MYO3B, GAD1* and *SP5* (Online Figure V B-D), suggesting a proximity-based co-expression paradigm at this genomics locus. In addition, expression of genes assigned within the module 4 (such as *EOMES*, *T* and *MIXL1*) exhibited significant down-regulation in cells transfected with siRNAs for *AK127400* and *LINC01124* at mesoderm stage (Online Figure V B-D); however, no obvious failure in mesoderm formation was observed after knockdown of expression of *AK127400* and *LINC01124* (Online Figure V E-F); instead the cells continued to be differentiated into beating cardiomyocytes (data not shown).

The genes within module 8 (Figure 3C and Online Figure IV B) were only up-regulated from day 2 to day 4, and their enriched GO terms (BP) included "pattern specification process" and "embryonic organ development", thus, module 8 is more likely related to the transition from mesoderm (day 2) to early cardiac mesoderm (day 4). Module 8 contains two nodes with significantly higher weighted-connectivity, wntless Wnt ligand secretion mediator (*WLS*) and E-box binding homeobox 2 (*ZEB2*). WLS is known as a membrane protein and was found to regulate WNT signaling pathway;[32] and ZEB2 is a TF regulating nervous system development and epithelial-mesenchymal transition (EMT) during the development.[33, 34] Their functions in cardiac development have been rarely documented. In addition, *LEF1*, Wnt family member, 8A (*WNT8A*),[35] apelin receptor (*APLNR*) [36] and DiGeorge syndrome critical region 8 (DGCR8) [37] were also had high weighted connectivity in the module 8.

Module 2 (Figure 3D and Online Figure IV C) was related to post-cardiac-mesoderm stage based on the enriched biological functions of the genes, expression of which was induced at both day4 (cardiac mesoderm) and day 30 (differentiated cardiomyocytes). Two of the genes with highest connectivity in this module included EGF containing fibulin-like extracellular matrix protein 2 (*EFEMP2*) and E-box binding homeobox 1 (*ZEB1*). It has been reported that EFEMP2 (also known as fibulin-4) is to organize elastic fibers of the connective tissue during development.[38-40] Like ZEB2, ZEB1 is also known as a regulator for the EMT during the development.[34, 41] GATA4 and HAND2 are known TFs regulating cardiac development within this module.[6, 22]

We additionally demonstrate that ZEB1 is functionally important for early cardiac differentiation, particularly during the cardiac mesoderm stage, and siRNA-mediated knockdown of *ZEB1* from initiation of differentiation to cardiac mesoderm stage induced failure in cardiomyocyte differentiation (Figures 4A-4D). Knockdown of *ZEB1* was associated with down-regulation of TFs that regulated cardiac differentiation, such as *GATA4*, *MESP1*, *MEIS1* and *LEF1*[6, 11, 42] (Online Figure VI). This result reveals an important role of ZEB1 in the early stages of cardiac differentiation.

Genes assigned in module 1 (Figure 3E and Online Figure IV D) were highly expressed on day 30, and most of them are related to cardiomyocyte structure and regulation of muscle contraction. Module 1 contained multiple genes with similar top levels of weighted connectivity, such as troponin I type 1 (*TNNI1*), *MYL3*, *MYH7* (myosin, heavy chain 7, cardiac muscle, beta), junctophilin 2 (*JPH2*), myosin binding protein C, cardiac (*MYBPC3*), leucine rich repeat containing 10 (*LRRC10*), SET And MYND domain containing 1 (*SMYD1*), myozenin 2 (*MYOZ2*), myomesin 1 (*MYOM1*), and protein phosphatase 1 regulatory subunit 12B (*PPP1R12B*, also known as *MYPT2*).[43-52] Both *NKX2-5* and *TBX5*, known as TFs to regulate cardiac differentiation, were assigned in this module.

Dynamics of chromatin accessibility during cardiomyocyte differentiation

Chromatin accessibility patterns often reveal mechanisms of gene regulation not evident from transcript patterns.[53] We performed ATAC-seq to map the open chromatin patterns on days 0, 2, 4 and 30. The dynamic changes of open-chromatin regions showed a stage-specific clustering pattern among the four cell lines (Online Figure VII), which was concordant with the clustering results observed for the transcriptome (Figure 2A). We observed that the correlation patterns of the ATAC-seq data exhibited similarity to that of the RNA-seq data (Figure 5A), suggesting that overall changes in gene expression and chromatin accessibility closely reflect each stage of cardiomyocyte differentiation. Importantly, clustering of four cell lines within each developmental stage demonstrates that changes in genome-wide chromatin accessibility are similar across the four cell lines, except for a slightly lower correlation between C15 and other cell lines on day 4. The pluripotent stage (day 0) and differentiated cardiomyocyte stage (day 30) were clustered more distinctly than mesoderm stage (day 2) and the cardiac mesoderm stage (day 4), indicating a close transition from mesoderm to cardiac mesoderm. In addition, based on the overlapping peaks among four cell lines, the promoter regions of overlapping open-chromatin across four cell lines showed an increase from 32% to 61% during the 30 days of differentiation (Figure 5B),

suggesting increases of transcriptional events during cardiomyocyte differentiation among four stem cell lines.

Motifs enrichment analysis reveals TF-binding features during cardiomyocyte differentiation

The enriched *de-novo* motifs were searched in the regions +/- 100-bp DNA sequences around the differential peaks from each subset of pairwise comparisons (Online Tables V-VIII). We used both E values (from enrichment analysis by DREME) and coverage (*i.e.*, the percentage of bound sites of the motif pattern) to create the scatter plots to identify transcription factor binding sites enriched for chromatin accessibility in each cell-line at different time points (Online Figure VIII A-D). We observed that the enriched motifs exhibited time-specific profiles during cardiomyocyte differentiation. Some enriched motifs were similarly enriched across all cell lines. For instance, the motif CYCCDCCC was enriched through all time- points for each cell line, and the matched TFs of this motif included Kruppel-like factor 5 (KLF5), SP1/2 transcription factors, early growth response gene 1 and 2 (EGR1/2), zinc finger protein 263 (ZNF263) and zinc finger protein 740 (ZNF740).

A number of motifs were mainly statistically enriched on days 2, 4 and 30, indicating that they probably play important roles during cardiomyocyte differentiation. The matched TFs of these motifs constituted multiple TF families, including V-maf avian musculoaponeurotic fibrosarcoma oncogene homolog K (MAFK), FOS gene family (FOS and FOSL2), lymphoid enhancer-binding factor 1 (LEF1), transcription factor 7-like 1 and 2 (TCF7L1/2), early B-cell factor 1 (EBF1), signal transducer and activator of transcription 3 (STAT3), MAX dimerization protein (MGA), MESP1, EOMES, T, JUN family (JUN, JUNB, and JUND), retinoid X receptor family (RXRA and RXRG), GATA family (GATA2, 3, 4, 5 and 6), TBX family (TBX1, 2, 5 and 20) and MEIS family (MEIS1, 2 and 3). These enriched motifs and expression of the matched TFs were visualized in Figures 6A and 6B. Motifs that were primarily enriched on day 4 and day 30 showed much higher similarity among the four cell lines, and the matched TFs included MAFK, FOS, FOSL2, Jun dimerization protein 2 (JDP2), JUN and GATA families (Figure 6B). The strong correlation between expression patterns for TFs and occurrences of enrichment of their matched motifs were observed in motifs for MESP1, T, EOMES, GATA and MEIS families: 1) motifs for MESP1, T and EOMES were enriched on day 2 (mesoderm); 2) motifs for GATAs were enriched from day 2 to day 30 (cardiomyocyte); and 3) motifs for MEISs were enriched from day 4 (cardiac mesoderm) to day 30. EOMES, T and MESP1 are known to play important roles in mesoderm development,[11, 19, 20, 54] and GATA4, MEIS1/2 were found to regulate cardiac development and morphogenesis.[6, 55] However, not all TF expression patterns were consistent with dynamic motif enrichment. For instance, the motifs for TBXs were enriched on day2, yet its expression was only highly expressed on day 30 (Figure 6A).

Overall, based on the analyses of the chromatin accessibility, we observed that the four cell lines exhibited similarities in chromatin state and motif enrichment patterns during cardiomyocyte differentiation, suggesting that both hiPSCs and hESCs share highly similar transcriptional regulation mechanisms underlying early cardiac differentiation.

Integration of WGCNA, gene expression, motif enrichment, and TF-footprinting analyses to identify stage-specific TFs during cardiomyocyte differentiation

To identify which TFs are likely to play important roles during cardiomyocyte differentiation, we integrated the results of WGCNA, temporal gene expression, and motif enrichment analysis to determine which TFs exhibited correlations among temporal expression patterns, assignment and connectivity in the modules, and motif enrichment occurrence (Figure 7A)**.** The TFs listed in the table were from the enriched motifs (in Figures 6A and 6B) that were discovered during cardiomyocyte differentiation. TFs with strong correlation between WGCNA (*i.e.* expression patterns) and motif enrichment include: *T* and *EOMES* in module 4; *LEF1* and *MESP1* in module 8; *GATA4* and *MEIS1* in module 2; and JUNs (*JUN*, *JUNB*, and *JUND*), *FOS*, *FOSL2* and *MEIS2* in module 1 (Figure 7B). Our results suggest that LEF1 and MESP1 play important roles in transcriptional regulation from mesoderm to cardiac mesoderm; and GATA4 and MEIS1 play important roles in post-cardiac-mesoderm differentiation. Previous studies showed that T, EOMES, LEF1 and MESP1 play important roles in mesoderm and cardiac development,[11, 19, 20, 42, 54, 56, 57] and MEIS1 was found to interact with GATA4 to regulate heart development.[55, 58] Complex of JUN and FOS proteins are known to form the subunits of Activator protein 1 (AP-1), which is a transcription factor and required for cardiac differentiation.[59, 60] MEIS2 was also found to act as a critical TF for regulation of cardiomyocyte differentiation derived from H7-hESC line.[7] In addition, some TFs showed correlation between WGCNA and motif enrichments, while they were had low weighted-connectivity or showed variability in occurrence of motif-enrichment among cell lines, they included RXRA and MGA (module 8), RXRG, GATA3 and GATA6 (module 2), EBF1, GATA2 and GATA 5 (module 1) (Figure 7B).

To further investigate gene-expression regulated by these TFs, we performed the TF-footprinting analysis to investigate relationships between TF occupancy and expression of genes that were predicted to be regulated by the TFs. We identified strong correlations between TF occupancy and the expression of proximal genes for TFs involved in post-cardiac-mesoderm stage to cardiomyocyte stage, including GATA4, MEIS1, JUNB, JUND, MEIS2, FOS, and FOSL2 (Figure 7C). Target genes that were predicted to be regulated by these TFs (see Online Data II) showed higher expression on day 30 compared to other time points, and the their enriched GO terms were related to cardiac development and relevant signaling pathways (Figure 7C), such as “heart morphogenesis” for GATA4 and “SMAD protein signal transduction” for MEIS1/2. These results are strongly coordinate with the integrative analysis in Figures 7A-7B, demonstrating important roles of these TFs in regulation of cardiac differentiation from post-cardiac-mesoderm to cardiomyocyte stages.

Discussion

In the present study, we used monolayer-differentiation methods to profile temporal genome-wide transcriptome and chromatin accessibility in both hiPSCs and hESCs. This was done to elucidate molecular changes at early stages during cardiomyocyte differentiation, and also to characterize similarities and variability in global gene expression and chromatin state during cardiomyocyte differentiation derived from different human pluripotent stem cells. Based on the PCA, clustering and correlation analyses of the RNA-seq data, both hiPSCs and hiESCs

exhibited high concordance in transcriptomics during cardiomyocyte differentiation (Figure 2). Moreover, the dynamic changes in enriched GO terms (BP) of those over-expressed genes in the four cell lines exhibited similarity. Thus, we can consider the variation in transcriptomic profiles among stem-cell lines as “minor”.

The sample-clustering patterns in the correlation analysis of the data of RNA-seq and ATAC-seq were consistent, indicating highly similar transcriptional regulation mechanisms underlying early cardiac differentiation among the four cell lines. Correlations of chromatin accessibility at the mesoderm stage (day 2) and the cardiac mesoderm stage (day 4) were less distinctly clustered than that in the pluripotent stage (day 0) and the differentiated cardiomyocyte stage (day 30) (Figure 5A). The short time period (2 days) from day 2 to day 4, in which several TFs play important roles at both stages, may explain the higher correlations in genome-wide chromatin accessibility between day 2 and day 4. These TFs probably include the TFs listed in Figures. 6A and 6B, such as MESP1 and LEF1. Regarding the slightly lower correlation in chromatin accessibility of C15 observed on day 4 (Figure 5A), we did not find any cell-line-specific motifs for C15 cell line, and all of four cell lines exhibited the similar motif enrichment patterns (Figures 6A and 6B). Comparison of ATAC-seq data between cell lines further identified motifs for several TFs (such as GATAs and MEISs) were over-representative in C15 on day 4 compared to other cell lines (Online Table IX). It is probable that variation in regulatory sequences (*e.g.,* enhancers) and regulatory factor binding patterns for those TFs (that regulate cardiac mesoderm) leads to the slight differences in chromatin accessibility of between C15 and other cell lines on day 4. It has been reported that differences in genetic background (such as single-nucleotide polymorphisms, SNPs) contribute to variation in chromatin accessibility among cell lineages or cell types,[61-63] and a previous study by Bock *et al.* (2011)[64] showed epigenetic and transcriptional variation between hiPSCs and hESCs. Thus, additional whole genome sequencing for these hiPSCs needs to be performed to identify variations in TF-binding sties among different cell lines in the future. It also suggests a possible mechanism where different cell-lines could potentially undertake plastic regulatory pathways to achieve a highly similar expression profile and eventually the same differentiated cell-type (*e.g.*, differentiated cardiomyocytes in this study), a phenomenon impossible to reveal if only differentiated states are profiled. Although this hypothesis needs to be validated by quantitative trait locus (QTL) analysis using a large number of stem-cell lines, large-scale profiling of other cell-type differentiations (*e.g.*, neural, epithelial and hepatocyte differentiations) derived from human pluripotent cells should to be investigated to determine if this is common among different cell types.

In the present study, we identified a number of genes with high weighted-connectivity in the networks associated with differentiation stages, and our results showed a paradigm of co-expression of these genes within the same module. Examination of both ATAC-seq signals and the topological domain structure of the locus of *AK127400* and *LINC01124*, as defined by high-throughput chromosome conformation capture (Hi-C) profiling across several cell-types, reveals that these genes are clustered within a single self-interacting domain (Online Figure IX).[65-67] Based on siRNA-mediated gene knockdown, we demonstrated a co-expression pattern including of protein-coding genes (*e.g.*, *MYO3B* and *GAD1* and *SP5*) and a long non-coding RNA (*LINC01124*) that was strongly associated with expression of

AK127400. This finding is similar to a previous study by Carcamo-Orive *et al.* (2017),[68] which used a completely different dataset and network analysis method to show that *SP5*, *GAD1*, *LINC01124* and *EOMES* exhibited high correlations within their network. This co-expression pattern probably results from co-expression regulation due to their proximity.[69] On the other hand, the efficiency of siRNA-mediated knockdown of *LINC01124* (around 40%) was not as high as *AK127400* (below 20%). The low efficiency of siRNA-mediated knockdown was also observed in knockdown attempts of long non-coding RNAs, *NEAT1* and *H19*, which were assigned in the module 2 related to post-cardiac-mesoderm stage. We did not find decreases in expression either in *NEAT1* or *H19* (data not shown).

In addition, we identified ZEB1 as an important factor during early cardiac differentiation, particular during cardiac mesoderm stage. Moreover, knockdown of *ZEB1* also induced decreases in expression of *LEF1*, *MESP1*, *GATA4* and *MEIS1*, which were identified as important TFs in the regulation of differentiation from early cardiac-mesoderm to post-cardiac-mesoderm stages (Figure 7B). This suggested an association between ZEB1 and other important regulators responsible for cardiac differentiation, and also supported the results of WGCNA in the present study. ZEB1 belongs to zinc finger transcription factors and plays important roles in the EMT.[34] EMT is essential for gastrulation (*i.e.,* a morphogenetic process of the formation of ectoderm, mesoderm and endoderm) required for heart development. Multiple signaling pathways are involved in the EMT process, including WNT/β-catenin signaling, TGF-β signaling and Notch signaling.[70-73] One important feature of EMT is the down-regulation of Cadherin 1 (CDH1, also known as E-cadherin) by ZEB1 and ZEB2. In the present study, we observed that expression of *CDH1* was down-regulated from day 4 to day 30 (Online Figure X A-B). Although differentiation of cardiomyocytes “in dish” derived from human stem cells is not completely equivalent to heart development *in vivo*, an EMT-like gene-expression pattern can still be observed in the stem cell-derived early cardiac differentiation, based upon the expression patterns and ATAC-seq signals at genomic loci of *ZEB*, *ZEB2* and *CDH1*.

Overall, our results demonstrated high similarities in genome-wide transcriptome and chromatin accessibility of cardiomyocyte differentiation between hiPSCs and hESCs. We also provided new insights into the dynamics of the factors and regulatory pathways controlling cardiomyocyte differentiation, particularly from mesoderm to cardiac mesoderm. These findings will provide a powerful resource for investigating mechanisms underlying early-on-set heart diseases due to disruption of regulatory pathways and factors identified from the study. The detailed roles of these genes and TFs in early cardiac differentiation also need to be addressed in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank both the Stanford Cardiovascular Institute (SCVI) Biobank and Stem Cell Core Facility of Genetics, Stanford University, provided the human pluripotent cells. We thank Rohith Srivas, Xianglong Zhang, Doug Phanstiel, Joshua Gruber and Kun-hsing Yu (Stanford University) for help with critical advice of data analysis. We

thank Sergiu Pasca (Stanford Center for Sleep Sciences and Medicine) and Stanford Neuroscience Microscopy Service (NMS, supported by NIH NS069375) for providing the confocal microscope. The Illumina sequencing services were performed by the Stanford Center for Genomics and Personalized Medicine. We thank Reinhold Hutz (University of Wisconsin Milwaukee) for reading the manuscripts.

Sources of Funding: This work was supported by California Institute for Regenerative Medicine, CIRM GC1R-06673-A (MPS) and National Institutes of Health grants (NIH R24 HL117756 [JCW], NIH R01 HL113006 [JCW], NIH R01 HL128170 [JCW], NIH P50-HG007735 [HYC, MPS] and NIH F32DK107112 [KVB]).

References

1. Anson B, Kolaja K, Kamp T. Opportunities for use of human ips cells in predictive toxicology. Clin Pharmacol Ther. 2011; 89:754–758. [PubMed: 21430658]
2. Davila J, Cezar G, Thiede M, Strom S, Miki T, Trosko J. Use and application of stem cells in toxicology. Toxicol Sci. 2004; 79:214–223. [PubMed: 15014205]
3. Zhu Z, Huangfu D. Human pluripotent stem cells: An emerging model in developmental biology. Development. 2013; 140:705–717. [PubMed: 23362344]
4. Navarrete E, Liang P, Lan F, Sanchez-Freire V, Simmons C, Gong T, Sharma A, Burridge P, Patlolla B, Lee A, Wu H, Beygui R, Wu S, Robbins R, Bers D, Wu J. Screening drug-induced arrhythmia [corrected] using human induced pluripotent stem cell-derived cardiomyocytes and low-impedance microelectrode arrays. Circulation. 2013; 129:S3–13.
5. Wu H, Lee J, Vincent L, Wang Q, Gu M, Lan F, Churko J, Sallam K, Matsa E, Sharma A, Gold J, Engler A, Xiang Y, Bers D, Wu J. Epigenetic regulation of phosphodiesterases 2a and 3a underlies compromised β-adrenergic signaling in an ipsc model of dilated cardiomyopathy. Cell Stem Cell. 2015; 17:89–100. [PubMed: 26095046]
6. Luna-Zurita L, Stirnimann C, Glatt S, Kaynak B, Thomas S, Baudin F, Samee M, He D, Small E, Mileikovsky M, Nagy A, Holloway A, Pollard K, Müller C, Bruneau B. Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. Cell. 2016; 164:999–1014. [PubMed: 26875865]
7. Paige S, Thomas S, Stoick-Cooper C, Wang H, Maves L, Sandstrom R, Pabon L, Reinecke H, Pratt G, Keller G, Moon R, Stamatoyannopoulos J, Murry C. A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. Cell. 2012; 151:221–232. [PubMed: 22981225]
8. Buenrostro J, Wu B, Chang H, Greenleaf W. Atac-seq: A method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol. 2015(21):29, 1–9.
9. Langfelder P, Horvath S. Wgcna: An r package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9doi: 10.1186/1471-2105-1189-1559
10. Sharma A, Li G, Rajarajan K, Hamaguchi R, Burridge P, Wu S. Derivation of highly purified cardiomyocytes from human induced pluripotent stem cells using small molecule-modulated differentiation and subsequent glucose starvation. J Vis Exp. 2015; 97doi: 10.3791/52628
11. Bondue A, Blanpain C. Mesp1: A key regulator of cardiovascular lineage commitment. Circ Res. 2010; 107:1414–1427. [PubMed: 21148448]
12. Cai C, Liang X, Shi Y, Chu P, Pfaff S, Chen J, Evans S. Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. Dev Cell. 2003; 5:877–889. [PubMed: 14667410]
13. Orford K, Scadden D. Deconstructing stem cell self-renewal: Genetic insights into cell-cycle regulation. Nat Rev Genet. 2008; 9:115–128. [PubMed: 18202695]
14. Zhang J, Nuebel E, Daley G, Koehler C, Teitell M. Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. Cell Stem Cell. 2012; 11:589–595. [PubMed: 23122286]
15. Kolwicz SJ, Purohit S, Tian R. Cardiac metabolism and its interactions with contraction, growth, and survival of cardiomyocytes. Circ Res. 2013; 113:603–616. [PubMed: 23948585]
16. Yu J, Vodyanik M, Smuga-Otto K, Antosiewicz-Bourget J, Frane J, Tian S, Nie J, Jonsdottir G, Ruotti V, Stewart R, Slukvin I, Thomson J. Induced pluripotent stem cell lines derived from human somatic cells. Science. 2007; 318:1917–1920. [PubMed: 18029452]

17. Doss M, Chen S, Winkler J, Hippler-Altenburg R, Odenthal M, Wickenhauser C, Balaraman S, Schulz H, Hummel O, Hübner N, Ghosh-Choudhury N, Sotiriadou I, Hescheler J, Sachinidis A. Transcriptomic and phenotypic analysis of murine embryonic stem cell derived bmp2+ lineage cells: An insight into mesodermal patterning. Genome Biol. 2007; 8:R184. [PubMed: 17784959]
18. Lim S, Pereira L, Wong M, Hirst C, Van Vranken B, Pick M, Trounson A, Elefanty A, Stanley E. Enforced expression of mixl1 during mouse es cell differentiation suppresses hematopoietic mesoderm and promotes endoderm formation. Stem Cells. 2009; 27:363–374. [PubMed: 19038793]
19. Russ A, Wattler S, Colledge W, Aparicio S, Carlton M, Pearce J, Barton S, Surani M, Ryan K, Nehls M, Wilson V, Evans M. Eomesodermin is required for mouse trophoblast development and mesoderm formation. Nature. 2000; 404:95–99. [PubMed: 10716450]
20. Showell C, Binder O, Conlon F. T-box genes in early embryogenesis. Dev Dyn. 2004; 229:201–218. [PubMed: 14699590]
21. Tran T, Wang X, Browne C, Zhang Y, Schinke M, Izumo S, Burcin M. Wnt3a-induced mesoderm formation and cardiomyogenesis in human embryonic stem cells. Stem cell. 2009; 27:1869–1878.
22. McFadden D, Barbosa A, Richardson J, Schneider M, Srivastava D, Olson E. The hand1 and hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. Development. 2005; 132:189–201. [PubMed: 15576406]
23. Bell J, Bernasochi G, Wollermann A, Raaijmakers A, Boon W, Simpson E, Curl C, Mellor K, Delbridge L. Myocardial and cardiomyocyte stress resilience is enhanced in aromatase-deficient female mouse hearts through camkiiδ activation. Endocrinology. 2015; 156:1429–1440. [PubMed: 25625588]
24. Hershberger R, Pinto J, Parks S, Kushner J, Li D, Ludwigsen S, Cowan J, Morales A, Parvatiyar M, Potter J. Clinical and functional characterization of tnnt2 mutations identified in patients with dilated cardiomyopathy. Circ Cardiovasc Genet. 2009; 2:306–313. [PubMed: 20031601]
25. Li M, Wang X, Sykes B. Structural based insights into the role of troponin in cardiac muscle pathophysiology. J Muscle Res Cell Motil. 2004; 25:559–579. [PubMed: 15711886]
26. Olson T, Karst M, Whitby F, Driscoll D. Myosin light chain mutation causes autosomal recessive cardiomyopathy with mid-cavitary hypertrophy and restrictive physiology. Circulation. 2002; 105:2337–2340. [PubMed: 12021217]
27. Foley A, Mercola M. Heart induction by wnt antagonists depends on the homeodomain transcription factor hex. Genes Dev. 2006; 19:387–396.
28. Nowotschin S, Ferrer-Vaquer A, Concepcion D, Papaioannou V, Hadjantonakis A. Interaction of wnt3a, msgn1 and tbx6 in neural versus paraxial mesoderm lineage commitment and paraxial mesoderm differentiation in the mouse embryo. Dev Biol. 2012; 367:1–14. [PubMed: 22546692]
29. Bernardo A, Faial T, Gardner L, Niakan K, Ortmann D, Senner C, Callery E, Trotter M, Hemberger M, Smith J, Bardwell L, Moffett A, Pedersen R. Brachyury and cdx2 mediate bmp-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. Cell Stem Cell. 2011; 9:144–155. [PubMed: 21816365]
30. Mendjan S, Mascetti V, Ortmann D, Ortiz M, Karjosukarso D, Ng Y, Moreau T, Pedersen R. Nanog and cdx2 pattern distinct subtypes of human mesoderm during exit from pluripotency. Cell Stem Cell. 2014; 15:310–325. [PubMed: 25042702]
31. Kalisz M, W M, Bisgaard H, Serup P. Even-skipped homeobox 1 controls human es cell differentiation by directly repressing goosecoid expression. Dev Biol. 2012; 362:94–103. [PubMed: 22178155]
32. Bänziger C, Soldini D, Schütt C, Zipperlen P, Hausmann G, B K. Wntless, a conserved membrane protein dedicated to the secretion of wnt proteins from signaling cells. Cell. 2006; 125:509–522. [PubMed: 16678095]
33. Hegarty S, Sullivan A, O'Keeffe G. Zeb2: A multifunctional regulator of nervous system development. Prog Neurobiol. 2015; 132:81–95. [PubMed: 26193487]
34. Vandewalle C, Van Roy F, Berx G. The role of the zeb family of transcription factors in development and disease. Cell Mol Life Sci. 2009; 66:773–787. [PubMed: 19011757]

35. Kwon C, Arnold J, Hsiao E, Taketo M, Conklin B, Srivastava D. Canonical wnt signaling is a positive regulator of mammalian cardiac progenitors. Proc Natl Acad Sci U S A. 2007; 104:10894–10899. [PubMed: 17576928]
36. Deshwar A, Chng S, Ho L, Reversade B, Scott I. The apelin receptor enhances nodal/tgfβ signaling to ensure proper cardiac development. Elife. 2016:ii–e13758.
37. Chapnik E, Sasson V, Blelloch R, Hornstein E. Dgcr8 controls neural crest cells survival in cardiovascular development. Dev Biol. 2012; 362:50–56. [PubMed: 22138056]
38. Kobayashi N, Kostka G, Garbe J, Keene D, Bächinger H, Hanisch F, Markova D, Tsuda T, Timpl R, Chu M, Sasaki T. A comparative analysis of the fibulin protein family. Biochemical characterization, binding interactions, and tissue localization. J Biol Chem. 2007; 282:11805–11816. [PubMed: 17324935]
39. Dasouki M, Markova D, Garola R, Sasaki T, Charbonneau N, Sakai L, ML C. Compound heterozygous mutations in fibulin-4 causing neonatal lethal pulmonary artery occlusion, aortic aneurysm, arachnodactyly, and mild cutis laxa. Am J Med Genet A. 2007; 143A:2635–2641. [PubMed: 17937443]
40. Hoyer J, Kraus C, Hammersen G, Geppert J, Rauch A. Lethal cutis laxa with contractural arachnodactyly, overgrowth and soft tissue bleeding due to a novel homozygous fibulin-4 gene mutation. Clin Genet. 2009; 76:276–281. [PubMed: 19664000]
41. Liu Y, El-Naggar S, Darling D, Higashi Y, Dean D. Zeb1 links epithelial-mesenchymal transition and cellular senescence. Development. 2008; 135:579–588. [PubMed: 18192284]
42. Klaus A, Müller M, Schulz H, Saga Y, Martin J, Birchmeier W. Wnt/β-catenin and bmp signals control distinct sets of transcription factors in cardiac progenitor cells. Proc Natl Acad Sci U S A. 2012; 109:10921–10926. [PubMed: 22711842]
43. Beavers D, Landstrom A, Chiang D, Wehrens X. Emerging roles of junctophilin-2 in the heart and implications for cardiac diseases. Cardiovasc Res. 2014; 103:198–205. [PubMed: 24935431]
44. Brody M, Cho E, Mysliwiec M, Kim T, Carlson C, Lee K, Lee Y. Lrrc10 is a novel cardiac-specific target gene of nkx2-5 and gata4. J Mol Cell Cardiol. 2013; 62:237–246. [PubMed: 23751912]
45. Brody M, Feng L, Grimes A, Hacker T, Olson T, Kamp T, Balijepalli R, Lee Y. Lrrc10 is required to maintain cardiac function in response to pressure overload. Am J Physiol Heart Circ Physiol. 2016; 310:H269–278. [PubMed: 26608339]
46. Meredith C, Herrmann R, Parry C, Liyanage K, Dye D, Durling H, Duff R, Beckman K, de Visser M, van der Graaff M, Hedera P, Fink J, Petty E, Lamont P, Fabian V, Bridges L, Voit T, Mastaglia F, Laing N. Mutations in the slow skeletal muscle fiber myosin heavy chain gene (myh7) cause laing early-onset distal myopathy (mpd1). Am J Hum Genet. 2004; 75:703–708. [PubMed: 15322983]
47. Okamoto R, Kato T, Mizoguchi A, Takahashi N, Nakakuki T, Mizutani H, Isaka N, Imanaka-Yoshida K, Kaibuchi K, Lu Z, Mabuchi K, Tao T, Hartshorne D, Nakano T, Ito M. Characterization and function of mypt2, a target subunit of myosin phosphatase in heart. Cell Signal. 2006; 18:1408–1416. [PubMed: 16431080]
48. Osio A, Tan L, Chen S, Lombardi R, Nagueh S, Shete S, Roberts R, Willerson J, Marian A. Myozenin 2 is a novel gene for human hypertrophic cardiomyopathy. Circ Res. 2007; 100:766–768. [PubMed: 17347475]
49. Park C, Pierce S, von Drehle M, Ivey K, Morgan J, Blau H, Srivastava D. Sknac, a smyd1-interacting transcription factor, is involved in cardiac development and skeletal muscle growth and regeneration. Proc Natl Acad Sci U S A. 2010; 107:20750–20755. [PubMed: 21071677]
50. Schaub M, Hefti M, Zuellig R, Morano I. Modulation of contractility in human cardiac hypertrophy by myosin essential light chain isoforms. Cardiovasc Res. 1998; 37:381–404. [PubMed: 9614495]
51. Winegrad S. Cardiac myosin binding protein c. Circ Res. 1999; 84:1117–1126. [PubMed: 10347086]
52. Wohlgemuth S, Crawford B, Pilgrim D. The myosin co-chaperone unc-45 is required for skeletal and cardiac muscle function in zebrafish. Dev Biol 2007 Mar 15. 2007; 303:483–492.

53. Buenrostro J, Giresi P, Zaba L, Chang H, Greenleaf W. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10:1213–1218. [PubMed: 24097267]

54. Chan S, Shi X, Toyama A, Arpke R, Dandapat A, Iacovino M, Kang J, Le G, Hagen H, Garry D, Kyba M. Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. Cell Stem Cell. 2013; 12:587–601. [PubMed: 23642367]

55. Wamstad J, Alexander J, Truty R, Shrikumar A, Li F, Eilertson K, Ding H, Wylie J, Pico A, Capra J, Erwin G, Kattman S, Keller G, Srivastava D, Levine S, Pollard K, Holloway A, Boyer L, Bruneau B. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell. 2012; 151:206–220. [PubMed: 22981692]

56. Kitajima S, Takagi A, Inoue T, Saga Y. Mesp1 and mesp2 are essential for the development of cardiac mesoderm. Development. 2000; 127:3215–3226. [PubMed: 10887078]

57. Lu H, Li Y, Wang Y, Liu Y, Wang W, Jia Z, Chen P, Ma K, Zhou C. Wnt-promoted isl1 expression through a novel tcf/lef1 binding site and h3k9 acetylation in early stages of cardiomyocyte differentiation of p19cl6 cells. Mol Cell Biochem. 2014; 39:183–192.

58. Dupays L, Shang C, Wilson R, Kotecha S, Wood S, Towers N, Mohun T. Sequential binding of meis1 and nkx2-5 on the popdc2 gene: A mechanism for spatiotemporal regulation of enhancers during cardiogenesis. Cell Rep. 2015; 13:183–195. [PubMed: 26411676]

59. Eriksson M, Leppä S. Mitogen-activated protein kinases and activator protein 1 are required for proliferation and cardiomyocyte differentiation of p19 embryonal carcinoma cells. J Biol Chem. 2002; 277:15992–16001. [PubMed: 11884386]

60. Jahangiri L, Sharpe M, Novikov N, González-Rosa J, Borikova A, Nevis K, Paffett-Lugassy N, Zhao L, Adams M, Guner-Ataman B, Burns C, Burns C. The ap-1 transcription factor component fosl2 potentiates the rate of myocardial differentiation from the zebrafish second heart field. Development. 2016; 143:113–122. [PubMed: 26732840]

61. Arvey A, van der Veeken J, Plitas G, Rich S, Concannon P, Rudensky A. Genetic and epigenetic variation in the lineage specification of regulatory t cells. Elife. 2015; 4

62. Kim K, Ban H, Seo J, Lee K, Yavartanoo M, Kim S, Park K, Cho S, Choi J. Genetic factors underlying discordance in chromatin accessibility between monozygotic twins. Genome Biol. 2014; 15:R72. [PubMed: 24887574]

63. Thurman R, Rynes E, Humbert R, Vierstra J, Maurano M, Haugen E, Sheffield N, Stergachis A, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield T, Diegel M, Dunn D, Ebersol A, Frum T, Giste E, Johnson A, Johnson E, Kutyavin T, Lajoie B, Lee B, Lee K, London D, Lotakis D, Neph S, Neri F, Nguyen E, Qu H, Reynolds A, Roach V, Safi A, Sanchez M, Sanyal A, Shafer A, Simon J, Song L, Vong S, Weaver M, Yan Y, Zhang Z, Zhang Z, Lenhard B, Tewari M, Dorschner M, Hansen R, Navas P, Stamatoyannopoulos G, Iyer V, Lieb J, Sunyaev S, Akey J, Sabo P, Kaul R, Furey T, Dekker J, Crawford G, Stamatoyannopoulos J. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75–82. [PubMed: 22955617]

64. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith Z, Ziller M, Croft G, Amoroso M, Oakley D, Gnirke A, Eggan K, Meissner A. Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines. Cell. 2011; 144:439–452. [PubMed: 21295703]

65. Symmons O, Uslu V, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 2014; 24:390–400. [PubMed: 24398455]

66. Le Dily F, Baù D, Pohl A, Vicent G, Serra F, Soronellas D, Castellano G, Wright R, Ballare C, Filion G, Marti-Renom M, Beato M. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev. 2014; 28:2151–2162. [PubMed: 25274727]

67. Rajapakse I, Perlman M, Scalzo D, Kooperberg C, Groudine M, Kosak S. The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. Proc Natl Acad Sci U S A. 2009; 106:6679–6684. [PubMed: 19276122]

68. Carcamo-Orive I, Hoffman G, Cundiff P, Beckmann N, D'Souza S, Knowles J, Patel A, Papatsenko D, Abbasi F, Reaven G, Whalen S, Lee P, Shahbazi M, Henrion M, Zhu K, Wang S, Roussos P, Schadt E, Pandey G, Chang R, Quertermous T, Lemischka I. Analysis of transcriptional variability

in a large human ipsc library reveals genetic and non-genetic determinants of heterogeneity. Cell Stem Cell. 2017; 20:518–532.e519. [PubMed: 28017796]

69. Engreitz J, Haines J, Perez E, Munson G, Chen J, Kane M, McDonel P, Guttman M, Lander E. Local regulation of gene expression by lncrna promoters, transcription and splicing. Nature. 2016; 539:452–455. [PubMed: 27783602]
70. Kennedy M, Chalamalasetty R, Thomas S, Garriock R, Jailwala P, Yamaguchi T. Sp5 and sp8 recruit β-catenin and tcf1-lef1 to select enhancers to activate wnt target gene transcription. Proc Natl Acad Sci U S A. 2016; 113:3545–3550. [PubMed: 26969725]
71. Kovacic J, Mercader N, Torres M, Boehm M, Fuster V. Epithelial-to-mesenchymal and endothelial-to-mesenchymal transition: From cardiovascular development to disease. Circulation. 2012; 125:1795–1808. [PubMed: 22492947]
72. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol. 2014; 15:178–196. [PubMed: 24556840]
73. von Gise A, Pu W. Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. Circ Res. 2012; 110:1628–1645. [PubMed: 22679138]

Non-standard Abbreviations and Acronyms

hiPSCs	human induced pluripotent stem cells
hESCs	human embryonic stem cells
RNA-seq	RNA sequencing
ATAC-seq	assay for transposase-accessible chromatin with high-throughput sequencing
WGCNA	weighted gene co-expression network analysis
GO	gene ontology
TF	transcription factor
TSS	transcription start sites
EMT	epithelial-mesenchymal transition
EOMES	eomesodermin
MEIS1/2/3	meis homeobox 1/2/3
MESP1/2	mesoderm posterior BHLH transcription factor 1/2
MGA	MAX dimerization protein
RXRA and RXRG	retinoid X receptor alpha and gamma
ZEB1/2	E-box binding homeobox 1/2
GATA2/3/4/5/6	GATA binding protein 2/3/4/5/6
TBX1/2/5/20	T-box transcription factor 1/2/5/20
TNNT2	cardiac Troponin T
NKX2-5	NK2 homeobox 5

ISL1	ISL LIM homeobox 1
LEF1	lymphoid enhancer-binding factor 1
HAND2	heart and neural crest derivatives expressed 2
WNT3A	Wnt family member 3A
WNT8A	Wnt family member 8A
BMP2	bone morphogenetic protein 2
MIXL1	mix paired-like homeobox 1
TNNI1	troponin I type 1
TNNI3	troponin I type 3, cardiac muscle
MYL3	myosin light chain 3
CAMK2D	calcium/calmodulin-dependent protein kinase II delta
DKK1	Dickkopf Wnt signaling pathway inhibitor 1
SP5	Sp5 transcription factor
MSGN1	mesogenin 1
EVX1	even-skipped homeobox 1
CDX2	caudal type homeobox 2
MYO3B	myosin IIIB
GAD1	glutamate decarboxylase 1
WLS	wntless Wnt ligand secretion mediator
MYH7	myosin, heavy chain 7, cardiac muscle, beta
JPH2	junctophilin 2
MYBPC3	myosin binding protein C, cardiac
LRRC10	leucine rich repeat containing 10
SMYD1	SET And MYND domain containing 1
MYOZ2	myozenin 2
MYOM1	myomesin 1
DGCR8	DiGeorge syndrome critical region 8
TCF7L1/2	transcription factor 7-like 1/2
EFEMP2	EGF containing fibulin-like extracellular matrix protein 2

EBF1	early B-cell factor 1
STAT3	signal transducer and activator of transcription 3
JDP2	Jun dimerization protein 2
CDH1	cadherin 1

Novelty and Significance

What is known?

- Cardiomyocyte differentiation derived from both hESCs and hiPSCs has been widely used employed in biomedical research.
- Weighted gene co-expression network analysis (WGCNA) of transcriptomes revealed correlation patterns among genes across different conditions.
- Evaluation of chromatin accessibility revealed important information regarding the DNA immediately surrounding transcription factor motifs and underlying mechanisms of gene regulation.

What new information does this article contribute?

- We discovered that the genome-wide transcriptome and chromatin accessibility exhibited high concordance during early cardiomyocyte differentiation among four stem cell lines (two hiPSC and two hESC lines).
- We investigated transcriptional regulation networks in early cardiomyocyte differentiation, especially the transition from mesoderm to cardiac mesoderm; and we validated ZEB1 as one of the key regulators during the early stages of cardiomyocyte differentiation.

Although *in-vitro* differentiated cardiomyocytes derived from both hESCs and hiPSCs have been widely used to investigate cardiac diseases and evaluate drug toxicity, gene regulation at earlier stages (*i.e.,* the period from mesoderm to cardiac mesoderm) among different stem cell lines has not been fully investigated. We characterized the genome-wide transcriptome and chromatin accessibility of cardiomyocyte differentiation between hiPSCs and hESCs, and our results demonstrated that cardiomyocyte differentiation exhibited high concordance between hESC and hiPSCs on a genome-wide scale. Moreover, we investigated the gene expression networks by integrating WGCNA and transcription factor (TF)-motif enrichment analysis. We determined a number of TFs and genes that are likely to play important roles during cardiomyocyte differentiation; and ZEB1 was specifically validated as one of the key regulators in early differentiation stages. Our study provided new insights into the dynamics of the regulatory pathways and factors controlling early cardiomyocyte differentiation. The methodology and findings from this study will provide a powerful resource for investigating mechanisms underlying early-onset heart diseases that are due to disruption of regulatory pathways and factors during early stages of cardiac development.

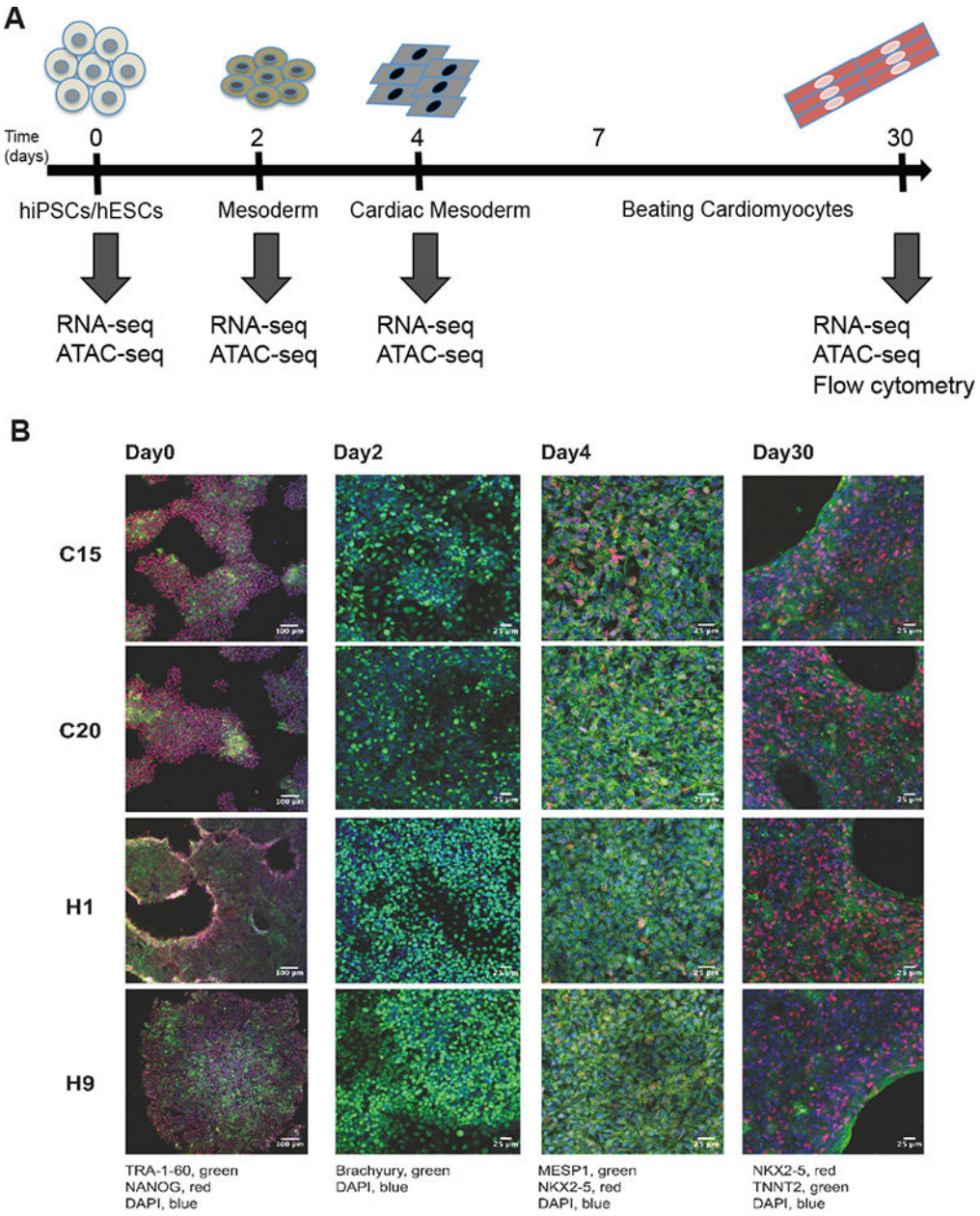


Figure 1. Cardiomyocyte differentiation from hiPSCs (C15 and C20) and hESCs (H1 and H9

(A) Schematic experimental design of cardiomyocyte differentiation and sample collection. Cells were collected on days 0, 2, 4 and 30 for RNA-seq and ATAC-seq. **(B)** Immunostaining of cells with special markers for each time points, including TRA-1-160 (green) and NANOG (red) for undifferentiated cells (day 0), Brachyury (green) for mesoderm (day 2), MESP1 (green) and NKX2-5 (red) for cardiac mesoderm (day 4), and NKX2-5 (green) and TNNT2 (red) for cardiomyocytes (day 30). The nucleus was stained with DAPI (blue).

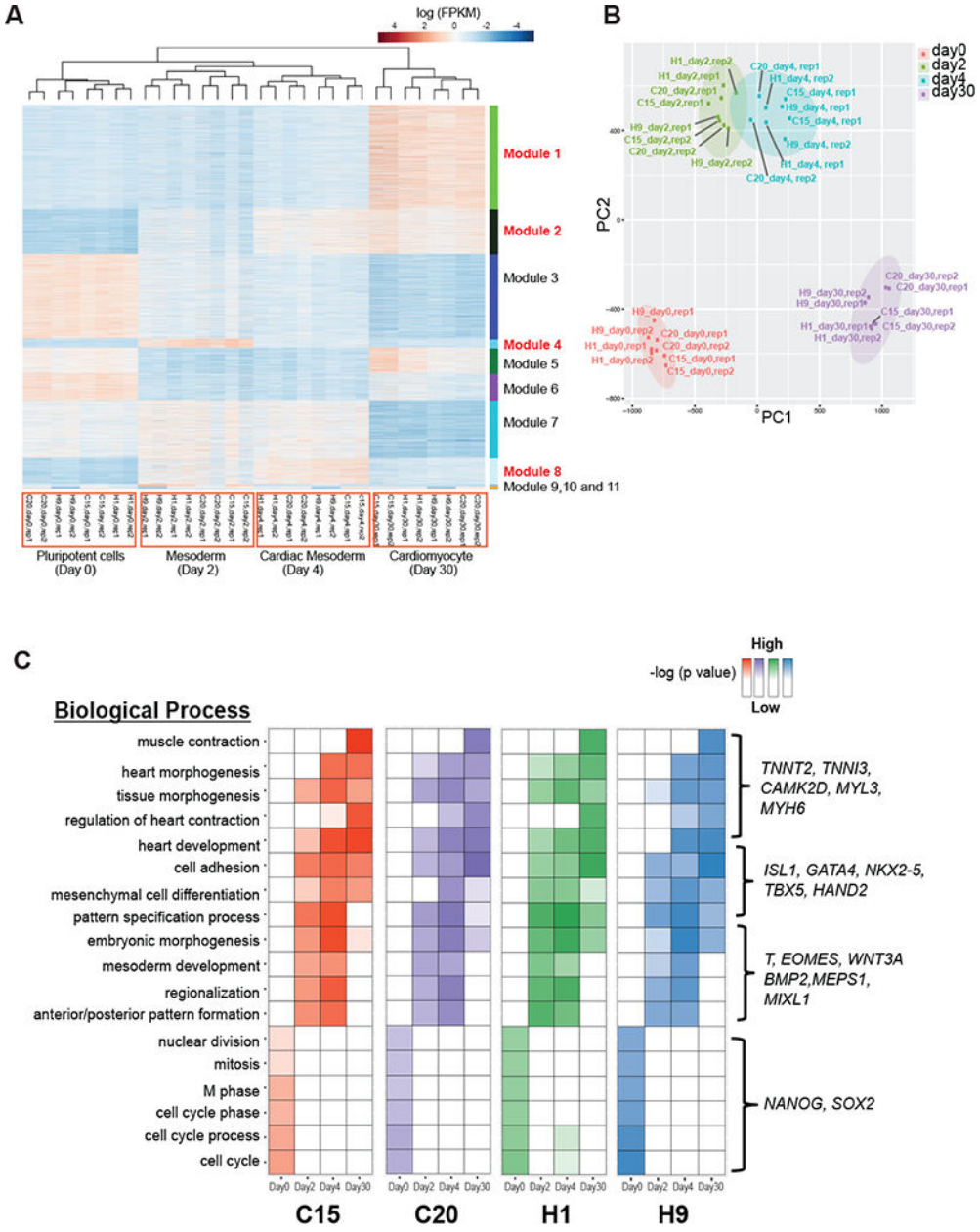


Figure 2. Temporal profiling of transcriptome at four stages of the cardiomyocyte differentiation

(A) Heatmap of gene expression profiles clustered by weighted gene co-expression network analysis (WGCNA). 11 modules were identified in total and labeled accordingly. **(B)** PCA analysis of genome-wide transcriptome during cardiomyocyte differentiation derived from four human-pluripotent stem cell lines. **(C)** Dynamic changes in enriched GO terms (BP) of top 1000 genes differentially expressed among four cell lines during cardiomyocyte differentiation, and representative genes assigned to the GO terms are shown in the figure.

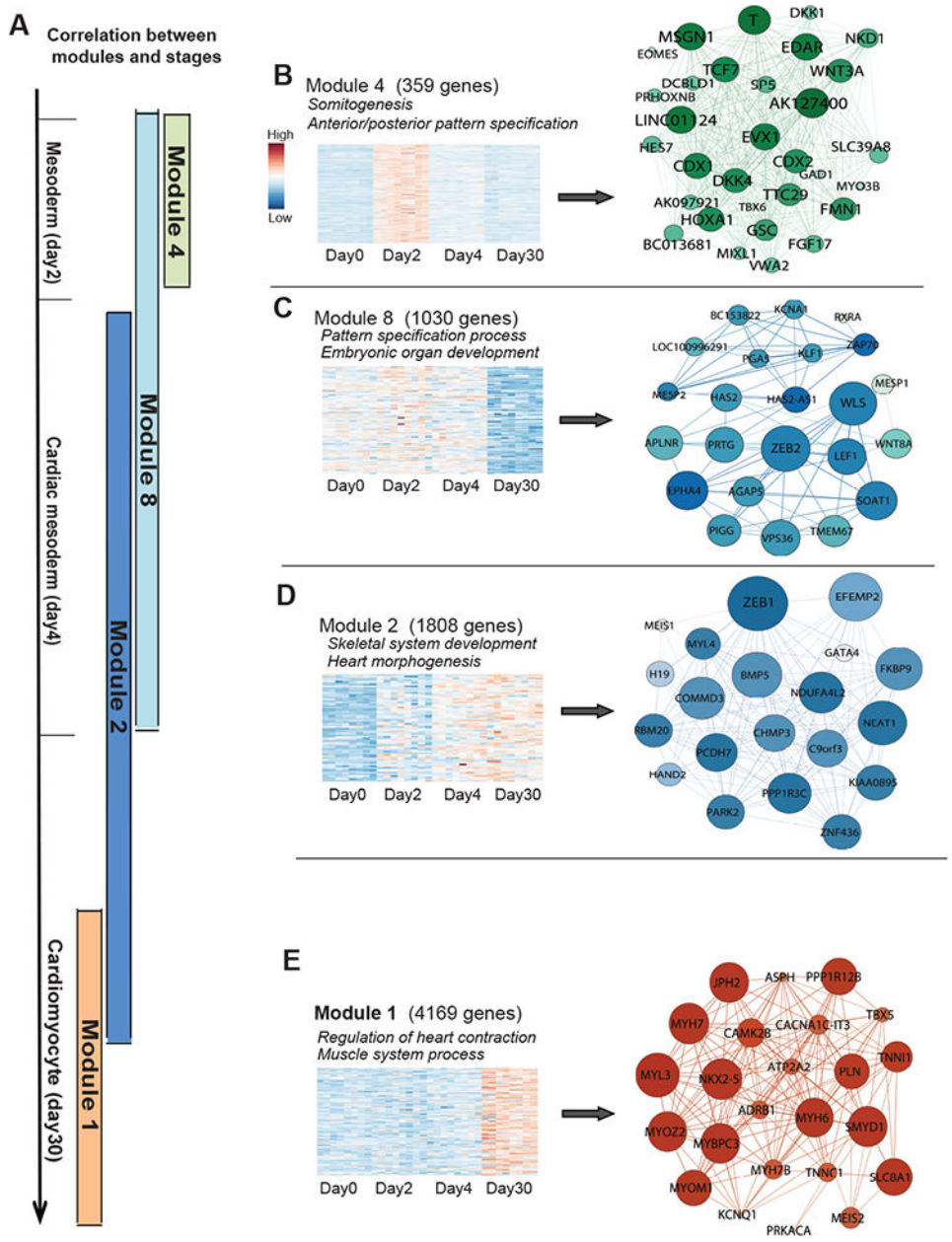


Figure 3. Differentiation-stage-related modules identified by WGCNA

(A) Four modules were strongly related to differentiation transitions from mesoderm to differentiated cardiomyocytes based on their temporal gene-expression patterns and the enriched GO terms (see Online Table IV). Module 4 is mesoderm-related; module 8 is related to mesoderm to cardiac mesoderm; module 2 is correlated to post-cardiac mesoderm; and module 1 is related to differentiated cardiomyocytes on day 30. **(B-E)** Left, gene expression patterns of each module with their representative GO terms; Right, visualization of networks with the genes with top weighted connectivity and TFs previously known to control cardiac differentiation. Size of each node represents the weighted connectivity of this gene within the module, which is the sum of adjacency scores (*i.e.* connection strength) between this gene and all other genes in the network.

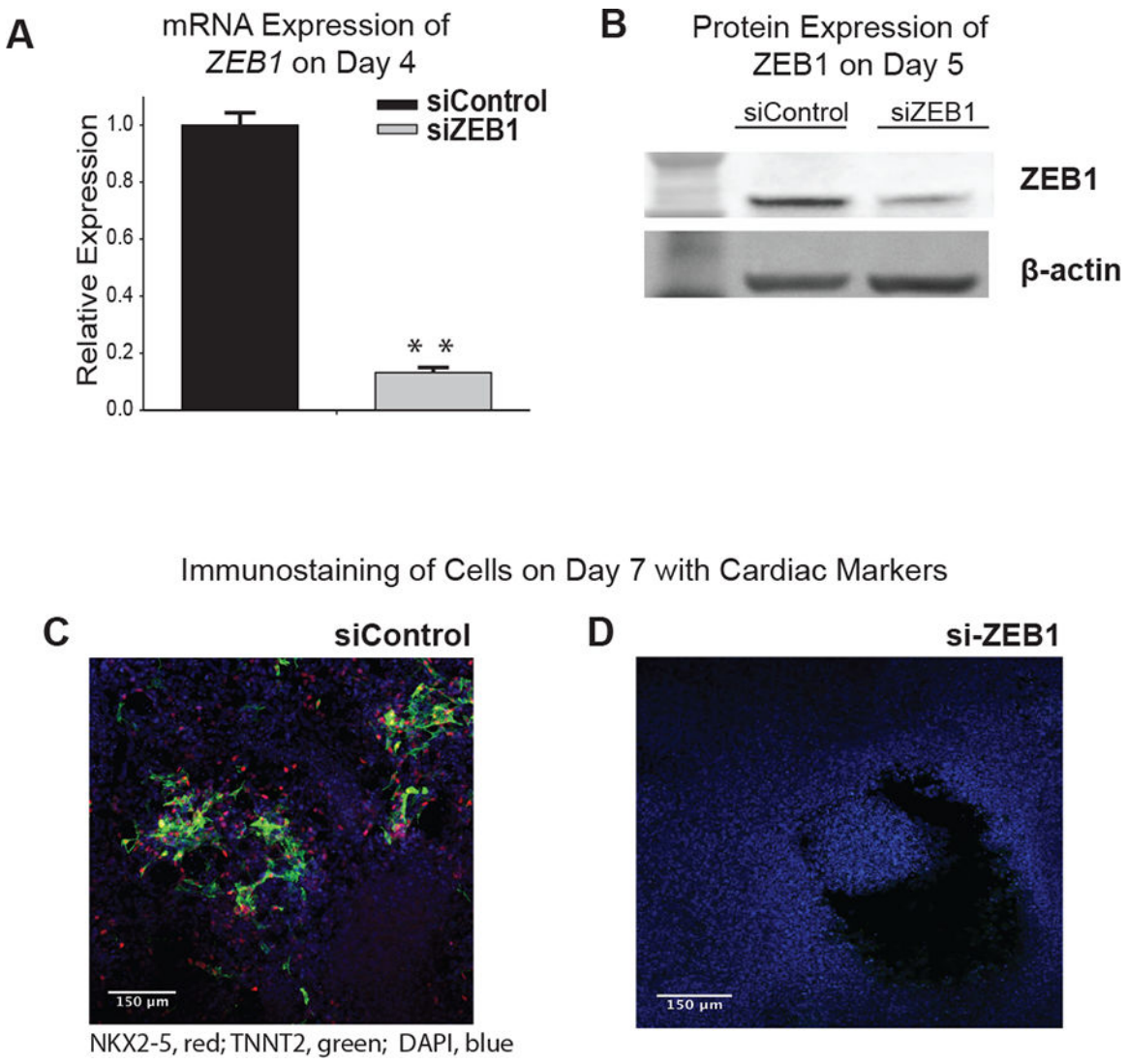


Figure 4. siRNA-mediated gene knockdown of *ZEB1* induced failure in cardiac differentiation from H1-hESCs

(A) Barchart represents relative expression of *ZEB1* at cardiac mesoderm stage (day 4) in H1-hESC cells which were constantly transfected with siControl and siZEB1 from initiation of differentiation. mRNA levels in cells transfected with siControl were set as 1, and relative expression levels were compared to that in siControl after normalization against 18S rRNA. T-test were perform for statistical analysis, * $p<0.05$ and ** $p<0.001$. **(B)** Expression of ZEB1 protein on day 5 in H1-hESC cells transfected with siControl and siZEB1 using Western-blot analysis. ZEB1 protein showed a down-regulation in siZEB1-transfected cells compared to that in control. **(C-D)** Immunostaining of cells transfected with siZEB1 **(C)** and siControl **(D)** at early cardiomyocyte stage (day 7). Cells transfected with siControl showed early cardiomyocyte feature, and expression of NKX2-5 (red) and TNNT2 (green) were detected. By contrast, continuous siRNA-mediated gene knockdown of *ZEB1* gradually failed cardiomyocyte differentiation.

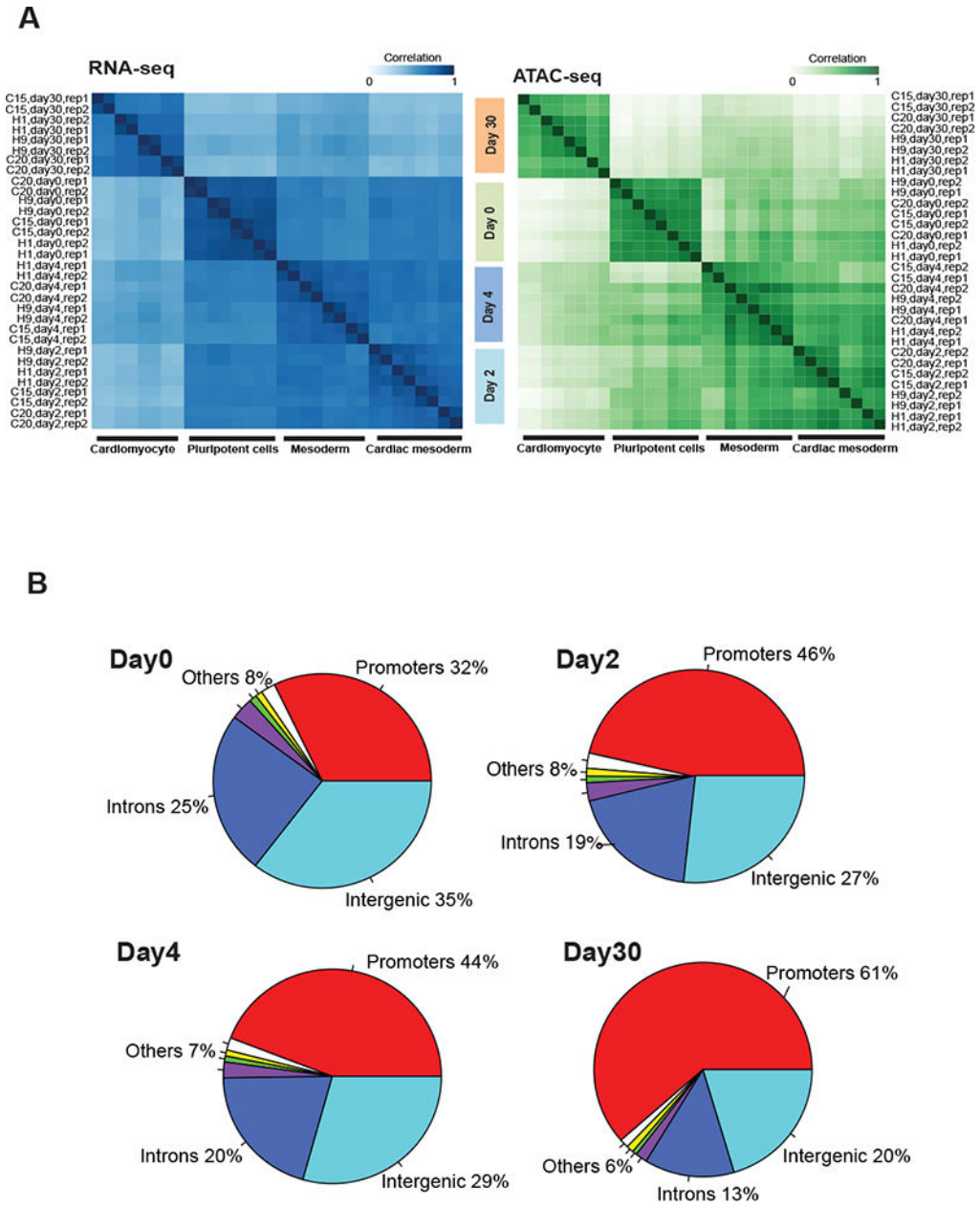


Figure 5. Dynamics of chromatin accessibility during cardiomyocyte differentiation derived from four human-pluripotent cell lines

(A) Correlation analysis of RNA-seq data (left) and ATAC-seq (right) data. The color represents the kendall correlation coefficients between samples, which were calculated using FPKM values of gene expression (for RNA-seq) and genome-wide chromatin accessibility (for ATAC-seq), respectively. **(B)** Changes of overlapping open-chromatin regions between cell lines. “Others” includes immediate downstream, 5′ UTRs, 3′ UTRs and exons.

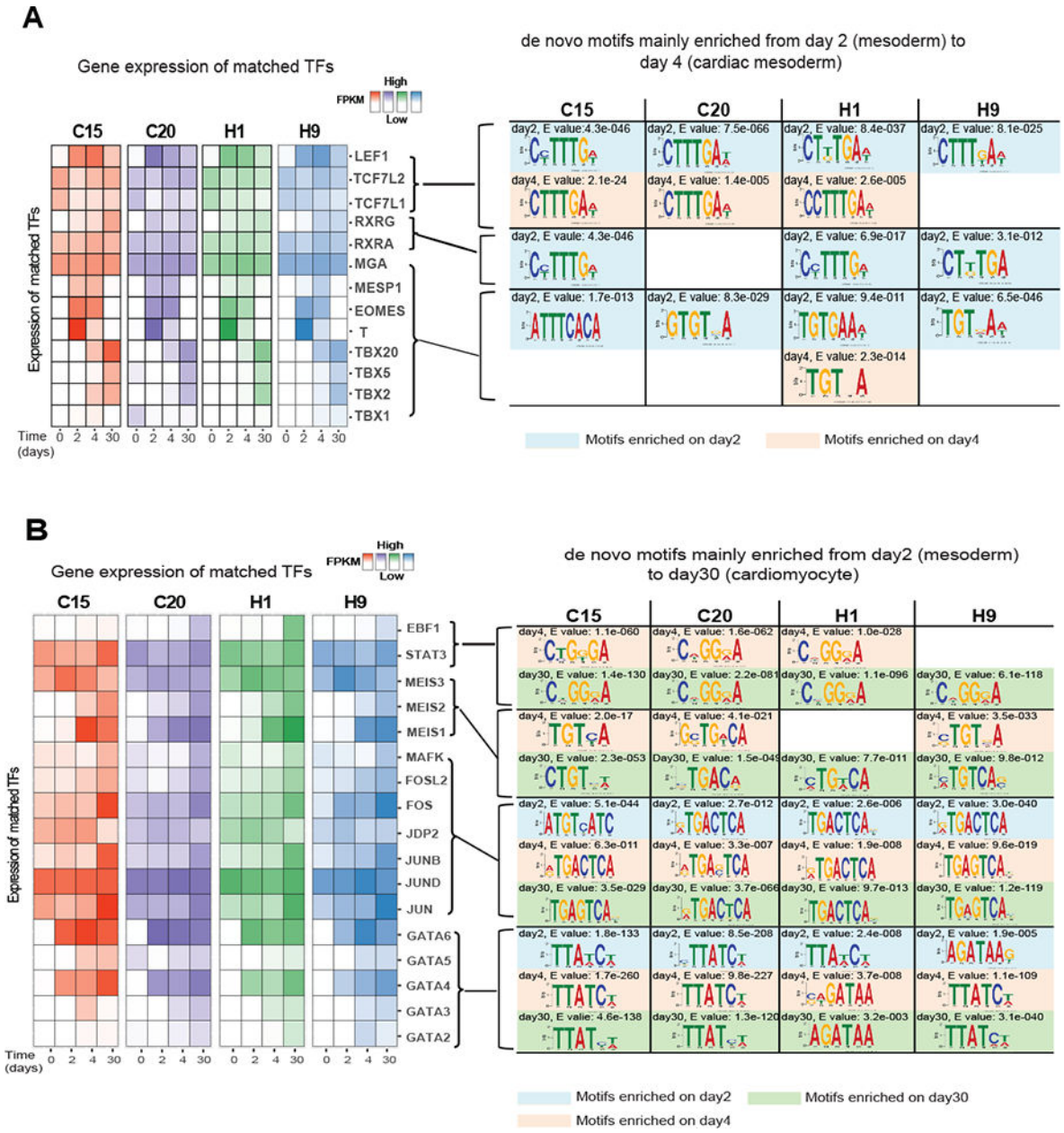
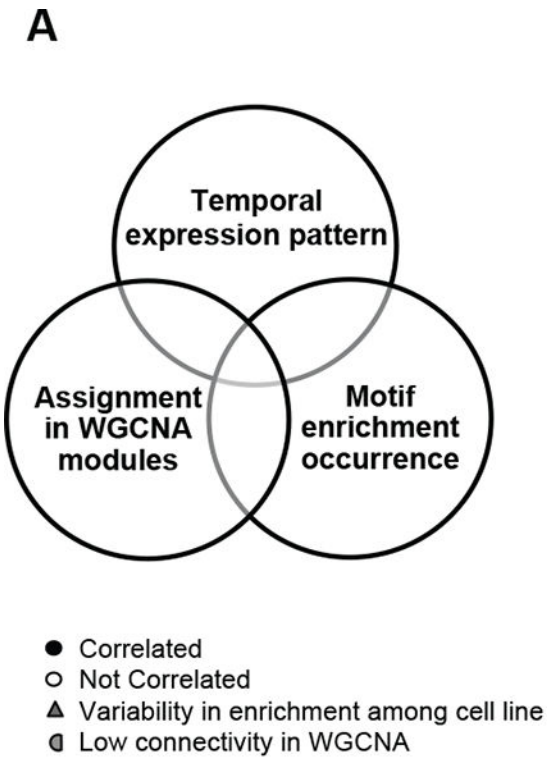
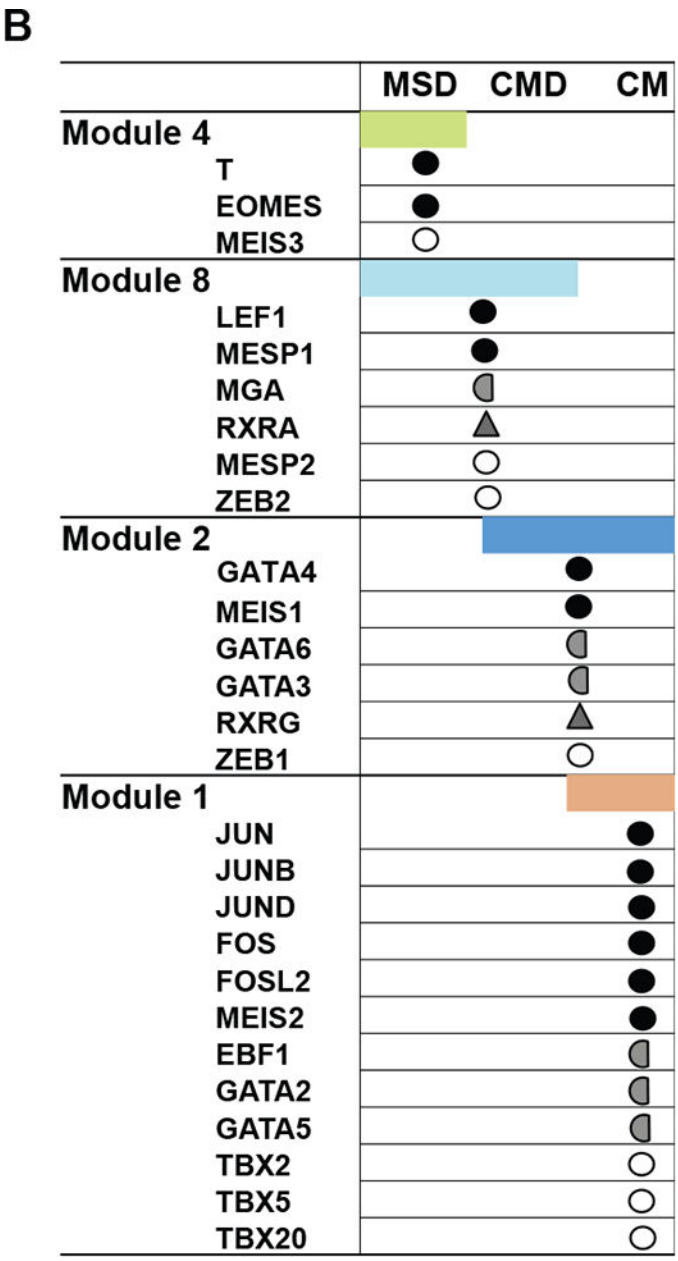


Figure 6. *De novo* motif analysis

De novo motifs were discovered using MEME-chip based on differential peaks of ATAC-seq data. Selected enriched de-novo motifs and expression of matched TFs during cardiomyocyte differentiation, including motifs mainly enriched **(A)** from mesoderm (day 2) to cardiac mesoderm (day 4), and **(B)** from mesoderm to cardiomyocyte (day 30). The heatmaps represent the expression values (FPKM) of the TFs.





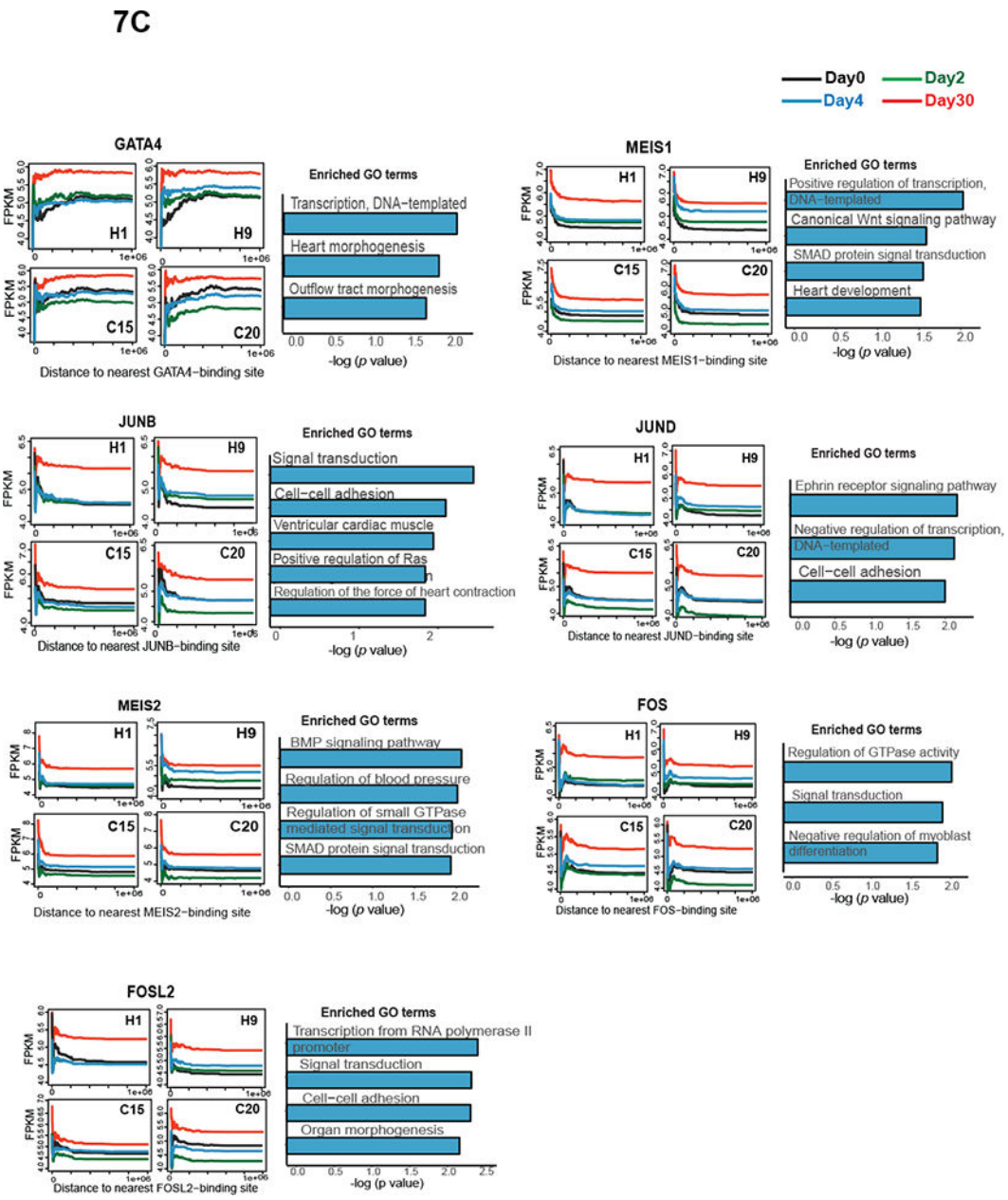


Figure 7. Integrative analysis of WGCNA, gene expression and motif enrichment analysis to identify stage-specific transcription factors

(A) Methodology of the integrative analysis. TFs were classified as four categories based on correlations among temporal expression patterns, assignment in the modules, and motif enrichment occurrence. **(B)** TFs identified based on the integrative analysis. The TFs listed in the table were achieved from the enriched motifs from day 2 to day 30 of cardiomyocyte differentiation. The TFs were assigned to the modules that were correlated to different stages of cardiomyocyte differentiation, including module 4 (related to mesoderm [MSD]), module 8 (differentiation from mesoderm to cardiac mesoderm [CMD]), module 2 (post-cardiac-mesoderm differentiation), and module 1 (related to cardiomyocyte [CM] stage). The color

bars represent correlations between each module (1, 2, 4, and 8) and differentiation stages, suggesting stage-specific roles of the TFs within these modules.

(C) Right, plotting of expression values (*i.e.* FPKMs) of differential genes in close spatial proximity to the genomic binding sites of these TFs in four cell lines (specifically, within a moving window of 1500-bp distance to the transcription start site [TSS]); left, statistically enriched ($p<0.05$) GO terms of the genes that were likely regulated by the TFs.