Deciphering transcriptional networks during human cardiac development

Authors

Robin Canac,[1,3] Bastien Cimarosti,[1,3] Aurore Girardeau,[1] Virginie Forest,[1] Pierre Olchesqui,[1] Jeremie Poschmann,[2] Richard Redon,[1] Patricia Lemarchand,[1] Nathalie Gaborit,[1,4,*] Guillaume Lamirault,[1,4,*]

[1] Nantes Université, CHU Nantes, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France

[2] INSERM, Nantes Université, Center for Research in Transplantation and Translational Immunology, UMR 1064, ITUN, F-44000 Nantes, France

[3] These authors contributed equally

[4] These authors contributed equally

* Correspondance: nathalie.gaborit@univ-nantes.fr (N.G.), guillaume.lamirault@univ-nantes.fr (G.L.)

Abstract

Human heart development is governed by transcription factor (TF) networks controlling dynamic and temporal gene expression alterations. Therefore, to comprehensively characterize these transcriptional regulations, day-to-day transcriptomic profiles were generated throughout the directed cardiac differentiation, starting from three distinct human induced pluripotent stem cell lines from healthy donors (32 days). We applied an expression-based correlation score to the chronological expression profiles of the TF genes, and clustered them into 12 sequential gene expression waves. We then identified a regulatory network of more than 23 000 activation and inhibition links between 216 TFs. Within this network, we observed previously unknown inferred transcriptional activations linking IRX3 and IRX5 TFs to three master cardiac TFs: GATA4, NKX2-5 and TBX5. Luciferase and co-immunoprecipitation assays demonstrated that these 5 TFs could (1) activate each other's expression, (2) interact physically as multiprotein complexes and (3) together, finely regulate the expression of *SCN5A*, encoding the major cardiac sodium channel. Altogether, these results unveiled thousands of interactions between TFs, generating multiple robust hypotheses governing human cardiac development.

Keywords

Stem cell differentiation, Human induced pluripotent stem cells, Heart development, Transcription factor, Gene regulatory networks, Transcriptomics, Transcription factor complexes, Iroquois transcription factors

Abbreviations

CHD : Congenital Heart Diseases
DEG : Differentially Expressed Genes
GO : Gene Ontology
hiPSCs : human induced Pluripotent Stem Cells
IGNiTe : Gene expression-based sub-network involving IRX3, IRX5, GATA4, NKX2-5 and TBX5
IRX : Iroquois homeobox transcription factor family
LEAP : Lag-based Expression Association for Pseudotime-series
MAC : Maximum Absolute Correlation
mESC : Mouse Embryonic Stem Cell
PC : Principal component
PCA : Principal Component Analysis
PPI : Protein-Protein Interaction
TF : Transcription Factor

Introduction

Heart formation is a complex process that requires spatio-temporal interplay between distinct and interdependent cell types through specific signaling and transcriptional pathways, leading to their differentiation and specification (1,2). Defects in this developmental process result in congenital heart disease as well as in a number of inherited cardiac disorders in adults (3). The specific gene expression program, governing the formation of a functional heart, needs precise regulation, in a time-, cell- and space-dependent manner (4). This program is mediated by transcription factors (TFs) regulating the expression of other TF-encoding genes and establish specific TF networks, such as between GATA4, NKX2-5 and TBX5 (5,6). These networks control and permanently remodel over time the transcriptional expression program that govern heart development.

A thorough understanding of these networks is crucial to gain knowledge on the transcriptional regulations and dysregulations that govern normal and pathological cardiac development, respectively. However, full knowledge of the global TF regulatory network of cardiac development is still missing. For instance, while several studies on Iroquois homeobox TF family (IRX) have shown their key roles on the regulation of adult cardiac electrical conduction (7–11), their function during human cardiac development has not been investigated yet. Cellular models derived from human induced Pluripotent Stem Cells (hiPSCs)

offer a unique opportunity to address these challenges, as they reproduce the cellular differentiation processes which lead stem cells to acquire a cardiac cell phenotype, carrying the genome of either healthy subjects or patients with inherited cardiac diseases.

In the present study, we first validated hiPSC cardiac differentiation model as a relevant tool to unravel the global TF regulatory network governing human cardiac development, identifying a network of 216 TFs with time-dependent activations and inactivations. Among these, we identified and biologically validated an undescribed TF regulatory network involving IRX3, IRX5 and three main cardiac TFs, GATA4, NKX2-5 and TBX5. Furthermore, we generated new hypotheses on the potential mechanisms leading to the cooperative effect of these TFs that could form a functional multiprotein complex activating the promoter of *SCN5A,* encoding the main cardiac sodium channel.

Results

Directed cardiac differentiation robustly generates functional cardiac cells

Cardiac differentiation of three hiPSC lines reprogrammed from three healthy donors was used as a cellular model of cardiac development (Fig 1A). After directed cardiac differentiation, all three hiPSC lines expressed cardiac-specific troponin I (Fig 1B), and displayed spontaneous contractions (Fig 1C), demonstrating their capability to form functional cardiomyocytes.

Fig 1. Transcriptomic and functional characterization of cardiac cells derived from hiPSCs.
(A) Diagram illustrating the experimental design involving three distinct cardiac differentiations of three hiPSC lines reprogrammed from healthy donors. (B) Immunocytochemistry staining of troponin I (red) and DAPI (blue) at D30 of cardiac differentiation for all 3 hiPSC lines. (C) Representative contraction patterns captured by MUSCLEMOTION software on movies at D30 of cardiac differentiation for the 3 hiPSC lines. (D) UMAP displaying single-cell RNA-seq data at D30 of cardiac differentiation of hiPSC-A line. The color code indicates the different cell types identified. Cell population fractions are listed on the right. See also Fig S1.

Based on single-cell transcriptomic data from 14 520 cells obtained at the end of directed cardiac differentiation (Fig 1D and Supplementary Fig S1), about 95% of the cells could be successfully annotated to one of the 15 cell types described in the developing human fetal heart (4), including 34% cardiomyocytes, 21% epicardial cells, and 15% fibroblast-like cells. This distribution was similar to previous findings in the adult human heart (12). These data indicate that directed cardiac hiPSC differentiation generated the cellular diversity observed in human fetal heart, known to be not only necessary for cardiac function, but also required for the establishment of cardiomyocytes (13).

To investigate how gene expression variations are orchestrated throughout cardiac differentiation, we then generated daily transcriptomic data, from hiPSC stage (D-1) to day 30 (D30), for three independent cardiac differentiations of each of the three hiPSC lines (Fig 2A). Directed cardiac differentiation was associated to gradual temporal transcriptomic changes, represented on the first principal component (PC1) of the principal component analysis (Fig 2B). PC1 was significantly correlated with time from the onset of cardiac differentiation (spearman correlation coefficient *rho*=0.87, *p-value* < $2.2.10^{-16}$). Cardiac differentiation evolution, represented by PC1 (Fig 2C), showed that 85% of transcriptomic variations were achieved by D14 (Fig 2D). Altogether, these data demonstrate that the first 14 days of hiPSC cardiac differentiation represent the ideal time window to investigate the molecular processes that lead to functional cardiac cells.

Fig 2. Transcriptomic time-course analysis of hiPSC cardiac differentiation.
(A) Methodological workflow. Steps are represented in white rectangle and outputs in red rectangles. (B) Global transcriptomic variations displayed with the first two components of the Principal Component Analysis. Three cardiac differentiations were studied for each of the three hiPSC lines. For each cardiac differentiation, a line connects the time-points in chronological order. (C) Boxplots displaying the distribution of PC1 coordinates of each replicates at each day (median +/- quartile). (D) Histogram comparing distribution of PC1 coordinates at the beginning (D-1), the middle (D14) and the end (D30) of hiPSC cardiac differentiations (Mean +/- SEM; Wilcoxon matched-pairs signed rank test).

Transcriptomic kinetics of hiPSC cardiac differentiation unveiled biological processes involved during cardiac development

Focusing on gene expression changes related to hiPSC cardiac differentiation, the 3 000 genes with the most significant expression variation during directed cardiac differentiation (differentially expressed genes; DEG) were identified and grouped into 12 clusters, chronologically ordered based on the time point when their expression level changes the most, showing distinct temporal gene expression profiles (Fig 3A). The average temporal expression pattern of each cluster was then compared to transcriptomic data obtained for the same genes from an *in vivo* reference model of murine cardiac development ((14); Fig 3B). As cardiac cells derived from hiPSCs are usually described as reaching an equivalent of, at the most, E18.5 stage in murine embryonic development (15), we restricted the comparison of the hiPSC dataset to murine developmental transcriptomic data obtained between murine embryonic stem cells and E18.5 stage. Apart from cluster D, all clusters displayed strikingly similar expression patterns between hiPSC cardiac differentiation and murine cardiac development. Nevertheless, genes of cluster D were associated with gastrulation biological processes (Fig 3B – Cluster D middle panel) which is completed before E7.5. As no data was available between the mouse embryonic stem cell (mESC) and E7.5 stages in the murine experiments, relevant gene expression changes associated to this process were likely to be absent in the murine transcriptomic dataset but remained detectable in the daily hiPSCs cardiac differentiation dataset. For all other 11 clusters, hiPSC cardiac differentiation could be confidently matched to sequential gene expression waves that occur during murine cardiac development. Altogether, these clusters recapitulate key steps of cardiac development, including (1) expression decrease of genes related to pluripotency and stemness maintenance (Fig 3B – Cluster A to C), followed by the transient expression of genes related (2) to gastrulation and mesoderm formation (Fig 3B – Cluster D) and (3) to early cardiac development (Fig 3B – Cluster E). These specific patterns were then followed by the successive implementation and persistence over time of gene expression waves that set up the sequential establishment of the functional cardiac phenotype (Fig 3B – Cluster F to L). To confirm these results, similar analyses were conducted on the top 3 000 DEG during murine cardiac development from mESCs to E18.5 (Supplementary Fig S2). This again revealed consistency of gene expression changes during hiPSC cardiac differentiation and during murine cardiac development. Collectively, these analyses demonstrate that hiPSC cardiac differentiation precisely recapitulates transcriptomic processes related to human and mouse cardiac development.

Fig 3. Expression profile and functional annotation of the top 3 000 differentially expressed genes (DEGs) during hiPSC cardiac differentiation, and comparison with murine cardiac development gene expression dataset.

(A) Heatmap displaying DEG expression levels. The entire data set was used to aggregate the genes into 12 clusters and the mean expression level of 9 replicates is represented. (B) For each cluster, average gene expression level during hiPSC cardiac differentiation (left panel for each cluster) and of their orthologs during murine cardiac development (mESCs to E18.5 stage, right panel for each cluster) are shown. Replicate gene expression levels were averaged for each hiPSC line (n=3 per hiPSC line and per timepoint) and for murine data (n=3 to n=6 per timepoint, depending on the stage). The 15 most significantly related GO terms are displayed for each cluster on the middle panel. See also Fig S2.

Prediction of gene regulatory networks governing hiPSC cardiac differentiation

TFs are known to be key players of developmental processes (6,16). Therefore, to elucidate gene regulatory networks that underlie human cardiac development, gene expression analysis was then focused on all 216 TFs that were found to be differentially expressed during the time-course of cardiac differentiation (Fig 4A; Supplementary Table 1). Overall, 69% of these TFs have been already linked to cardiac (patho)physiological phenotypes (Supplementary Table 1). We chose to adapt an expression correlation score involving time delay (LEAP method, see Methods) to capture gene associations that are hidden by time lags (*i.e.* time delay between the mRNA expression of the source gene and the mRNA expression of its target gene). Using this method on the 216 TFs, we predicted interactions that activated or inhibited the expression of target TFs by source TFs, building a regulatory network. This gene expression-based network included 11 467 activating interactions and 11 539 inhibitory interactions (Fig 4B left panel; Supplementary Table 2). We then evaluated the biological relevance of these TF interactions, using the STRING protein-protein interaction (PPI) database to generate an undirected PPI-based network restricted to the 216 TFs (Fig 4B right panel). Interestingly, 182 TFs (84%) were found to share at least one known PPI interaction. This included interactions between TFs belonging to the same gene cluster but also, interactions between TFs from different gene clusters, suggesting coordination between TFs to regulate the successive gene expression waves.

Fig 4. Inferred TF regulatory network governing hiPSC cardiac differentiation.

(A) Normalized gene expression of the 216 TFs (identified within the top 3 000 differentially expressed genes during hiPSC cardiac differentiation) were quantified and averaged in each gene cluster. UPM: UMI per million. The number in () indicates the TF number per cluster. (B) Graphical representation of gene expression-based network and protein-protein interaction-based network (LEAP- and STRING-based method, respectively) of the same TFs as in A. Interactions between TFs of successive clusters are shown using bold lines. (C-E) Comparative quantitative analysis between both networks. (F-G) Examples of two literature-based sub-networks. Interactions uncovered in gene expression-based network are shown in blue, in PPI-based network, in yellow, and by literature curation, in black. Node colors correspond to the one of their corresponding gene cluster (as in A). Paper PMID associated with literature-based links: [A] 33803261; [B] 23417899; [C] 34901033; [D] 15253934; [E] 17011492; [F] 21632880; [G] 22402664; [H] 22438573.

Comparing both networks, the gene expression-based network (LEAP-based) contained a greater amount of information than the PPI-based network (STRING-based) (Fig 4B). Indeed, although both networks were generated using the same TF query list, the density (i.e. normalized averaged number of neighbors) of the gene expression-based network was 5.5 fold higher, as compared to the PPI-based network. Deeper analysis showed that about 100% of the nodes and 80% of the links found in the PPI-based network were also found in the gene expression-based network (Fig 4C-D). Moreover, focusing on links between successive expression clusters, more than 76% of those found in PPI-based network were also found in the gene expression-based network (Fig 4E). Further confirming the accuracy of gene expression-based strategy, sub-networks that have been well-described in the literature were also present in both networks: (1) the network composed of the main actors of pluripotency (e.g POU5F1) and early phases of cardiac development (e.g EOMES, MESP1; Fig 4F), and (2) the TF network implicated in cardiogenesis (e.g ISL1, MEF2C; Fig 4G). This validated the relevance of such expression correlation score approach taking into account time delay to comprehensively analyze TFs and their interactions throughout cardiac differentiation. Altogether, while the gene expression-based network confirmed already known and validated interactions, it also inferred 21 530 new interactions unveiling numerous new hypotheses on TF networks potentially critical for cardiac development.

IRX3 and IRX5 are involved in triggering expression of GATA4, NKX2-5, TBX5 cardiac transcription factor network

Leveraging this new gene expression-based network to uncover new regulation mechanisms, and based on our previous focus of interest (7,10) we evaluated IRX TF family involvement in the establishment of cardiac developmental processes. Expression levels of the 6 different *IRX* TF genes was analyzed during cardiac differentiation in the three hiPSC lines (Fig 5A). Expression of *IRX6* was undetectable and expression of *IRX1* and *IRX2* did not vary over time. Only IRX3, IRX4 and IRX5 expression increased significantly between D-1 and D30 of cardiac differentiation. Interestingly, based on their expression profiles, *IRX3* and *IRX5* ranged from the earliest cardiac-specific gene cluster with an expression level that was maintained until the end of cardiac hiPSC differentiation (cluster F). This suggested a potential role for IRX3 and IRX5 in the early establishment of gene regulatory networks essential for cardiac fate, and beyond. In contrast, *IRX4* expression was detected in one of the latest clusters (cluster K). Therefore, we then focused on both IRX3 and IRX5 TFs.

Fig 5. Exploration of the inferred IIGNT sub-network.

(A) Table of TF selection criteria for the IRX genes family. Timecourse expression rank is the output of the timecourse package and illustrates the variation in gene expression during directed cardiac differentiation (a lower number indicating a higher variation). The Expression cluster column refers at the expression cluster in which each TF ranged as in Fig 4A. Selection and exclusion criteria are indicated in green and red respectively. (B) Potential target TFs of IRX3 and/or IRX5 identified in the G-to-L clusters based on the gene expression-based network. (C) Gene expression-based network of IRX3, IRX5, GATA4, NKX2-5 and TBX5 TFs. Node colors represent their corresponding clusters as in Fig 4A: IRX3 and IRX5 – cluster F; GATA4 – cluster G; NKX2-5 – cluster I; TBX5 – cluster L. Lag is shown in days. References to literature-based links: [A] Book chapter DOI: 10.1016/B978-0-12-381332-9.00027-X.; [B] PMID: 23457256; [C] PMID: 22449847; [D] PMID: 32450132; [E] PMID: 25280899. (D) Graphs displaying activity levels of luciferase that is under the control of GATA4 (-1800_TSS_+200), NKX2-5 (-2000bp_Start codon) and TBX5 (-1800_TSS_+200) promoter constructs. Mean +/- SD; * and **: $p < 0.05$ and $p < 0.01$, respectively (Mann-Whitney test).

In order to investigate the role of IRX3 and IRX5 in cardiac differentiation progression, all their potential TF targets in the subsequent G to L clusters were extracted from the gene expression-based network (Fig 5B). Interestingly, the master cardiac TF genes *GATA4* (cluster G), *NKX2-5* (cluster I) and *TBX5* (cluster L) were individually found to be potential targets of both IRX3 and IRX5. It is well established that GATA4 acts in a multiprotein complex with NKX2-5 (cluster I) and TBX5 (cluster L) cardiac TFs (6,17,18). To further explore potential new interactions, we then focused on the gene expression-based sub-network involving *IRX3, IRX5, GATA4, NKX2-5* and *TBX5*, referred later as IGNiTe sub-network (Fig 5C). In the IGNiTe sub-network, IRX3 and IRX5 were inferred as activators of *GATA4, NKX2-5* and *TBX5,* and confirming the literature (5,19–21), GATA4 was inferred as activator of *NKX2-5* and both GATA4 and NKX2-5 were activators of *TBX5* expression.

In order to investigate the biological relevance of these inferred interactions, luciferase assays were conducted on *GATA4, NKX2-5* and *TBX5* core promoters (Fig 5D). IRX3 and IRX5 proteins were, separately (fold changes 4.2 and 1.5 respectively) or together (fold change 4.5), able to bind the promoter of *GATA4* and to activate luciferase expression. A slight tendency towards potentiation of both activating effects is observable when IRX3 and IRX5 were present but was not statistically significant. On the *NKX2-5* promoter, IRX5 alone was able to activate luciferase expression (1.3-fold change), but not IRX3, suggesting that the inferred activation of *NKX2-5* by IRX3 found in the IGNiTe sub-network was due to IRX5, and that the high similarity between *IRX3* and *IRX5* expression profiles caused the false-positive link to appear. Together, IRX3 and IRX5 were able to activate *NKX2-5* promoter, with a tendency towards potentiation too (fold change 1.2 between IRX5 alone and IRX3/IRX5 conditions; $p>0.05$). According to the order of appearance of TFs in IGNiTe sub-network, NKX2-5 promoter activation was assessed in the combined presence of IRX3, IRX5 and GATA4, which showed an activator effect (1.8-fold change). Although a potentiation tendency was observed when GATA4 was present in addition with IRX3 and IRX5, this effect was not statistically significant. On *TBX5* promoter, IRX3 and IRX5 were able to bind and activate gene expression either individually (2.7- and 1.2-fold change, respectively) or together (3.6-fold change). Potentiation of both activator effects was clearly observable and statistically significant when IRX3 and IRX5 were together on the *TBX5* promoter. Finally, considering the joint expression of IRX3, IRX5, GATA4 and NKX2-5 from D10, we proved the activator effect of these TFs on *TBX5* promoter

(6.6-fold change), which is statistically increase from the IRX3/IRX5 condition (1.8-fold change). Collectively, these results biologically validated the new interactions inferred with the gene expression-based network and illustrated the progressive temporal activation of the major TFs GATA4, NKX2-5 and TBX5, by IRX3 and IRX5 during cardiac cell lineage establishment.

IRX3 and IRX5 physically interact with GATA4, NKX2-5 and TBX5 to control SCN5A expression

As the expression of the IGNiTe sub-network members was maintained until D30 of hiPSC cardiac differentiation (Fig 6A), the functional role of IRX3, IRX5, GATA4, NKX2-5, and TBX5 as a multiprotein complex was evaluated using co-immunoprecipitation and luciferase assays in heterologous expression systems. Luciferase assays were conducted on the promoter of *SCN5A*, a known target of these TFs (7,22–25). According to the chronological order of expression of these five TFs along cardiac differentiation of hiPSCs (Fig 6A), we first investigated the role of IRX3 and IRX5. As previously described (26), IRX3 and IRX5 physically interacted (Fig 6B top panel and Supplementary Fig S3) and could cooperatively activate *the SCN5A* promoter (Fig 6B bottom panel). While IRX3 alone activated *the SCN5A* promoter (2.3-fold change), IRX5 potentiated its effect with a 1.5-fold change. GATA4 was able to physically interact with IRX5 but not with IRX3 (Fig 6C top panels) and when the three TFs were co-transfected, only GATA4 and IRX5 interacted, suggesting a competitive effect between IRX3 and GATA4 to bind IRX5 (Fig 6C bottom left panel). Also, the addition of GATA4 potentiated (1.5-fold change) the activity of the IRX3/IRX5 couple on *SCN5A* promoter (Fig 6C bottom right panel). NKX2-5 interacted with both IRX3 and IRX5 individually (Fig 6D left panels), but again, when the four TFs were co-transfected we only observed an interaction between IRX5, GATA4, and NKX2-5, suggesting again a competition between IRX3 and IRX5, in favor of IRX5, in these interactions (Fig 6D central panel). NKX2-5 amplified (8.0-fold change) the effect of the IRX3/IRX5/GATA4 trio on the *SCN5A* promoter (Fig 6D right panels). Finally, when IRX3, IRX5, GATA4, NKX2-5 and TBX5 were co-transfected a global protein complex could be formed between IRX5, GATA4, NKX2-5 and TBX5, but not with IRX3, even if IRX3 alone was able to interact with TBX5 (Fig 6E left and central panels). However here, TBX5 slightly reduced (-1.6-fold change) the effect of the IRX3/IRX5/GATA4/NKX2-5 quartet on *SCN5A* promoter (Fig 6E right panel) suggesting a down-regulating role of TBX5 in this global complex. Collectively, we showed that following *IRX3, IRX5, GATA4, NKX2-5* and *TBX5* gene expression increase during

cardiac differentiation, the direct activation of *SCN5A* expression is under the control of a time-changing multi-TFs complex that controls the temporal expression profile of *SCN5A*.

Fig 6. Physical and functional interactions of IRX3/IRX5/GATA4/NKX2-5/TBX5 multiprotein complex. (A) Normalized mean expression level overtime of IRX3, IRX5, GATA4, NKX2-5, TBX5 and of SCN5A ion channel genes. UPM: UMI per million. (B-E) Co-immunoprecipitation and luciferase results associated to the transfection of (B) IRX3 and/or IRX5, (C) IRX3, IRX5 and/or GATA4, (D) IRX3, IRX5, GATA4 and/or NKX2-5, (E) IRX3, IRX5, GATA4, NKX2-5 and/or TBX5. Immunoblots representative of the various co-immunoprecipitations and the schematic illustration of the results. Graphs display activity levels of luciferase that is under the control of the -2109/+1072 region of human SCN5A promoter, in the various transfection conditions. Mean +/- SEM; * and **: $p < 0.05$ and $p < 0.0001$, respectively (Mann-Whitney test). See also Fig S3.

Discussion

In this study, based on a transcriptomic kinetics study on cardiac differentiation of hiPSCs, we identified the global TF regulatory network that is required for heart development. We notably identified novel time-dependent TF-gene regulations that connect *IRX3* and *IRX5* to the core cardiac *GATA4, NKX2-5* and *TBX5* TFs. We also found that these five TFs form protein complexes to regulate target gene expression, such as *SCN5A*. Altogether, this time-course bulk transcriptomic data provided a dynamic model relevant to identify new roles for TFs in developmental processes.

In vitro modeling of time in cardiac development

This study demonstrates that hiPSC cardiac differentiation is a relevant model to study the successive steps leading to the establishment of the gene expression program during human cardiac development. To date, most studies contributing to the knowledge on heart development and TF regulation have been conducted in animal models, mainly in mice (27), as access to human embryonic cardiac tissue, is indeed very limited. If regulatory mechanisms of development are overall highly evolutionary conserved, some are human-specific (28,29). Therefore, investigation of human cardiac development also requires suitable human models. HiPSC cardiac differentiation models have proved to generate functional cardiac cells and

suggested that punctual time points during this differentiation might reflect some key developmental stages (7,30,31). However, fully assessing the relevance of hiPSC cardiac differentiation model to study human cardiac development, requires demonstrating that it thoroughly and accurately reproduces human cardiac development in a temporally coordinated fashion. All phenotypic changes that occur during cardiogenesis are known to be embodied by dynamic alterations in cellular transcriptome. Yet, although the ideal situation would be to compare transcriptomic changes along hiPSC differentiation to the ones occurring during human cardiac development, no public human transcriptomic dataset studying well-distributed stages across the entire cardiac development is available. In the present study, we therefore used murine cardiac transcriptomic data generated from specific stages that appropriately rang from conception to birth (14), to compare with hiPSC cardiac differentiation data. Their high level of consistency confirmed that our hiPSC cardiac differentiation model accurately reproduces cardiogenesis. An important added value of the present data is that it filled a gap of knowledge on the global gene expression changes that occur daily, between these developmental stages in human cells.

A major limitation of hiPSC-derived models is immaturity: cardiac cells produced by current hiPSC differentiation protocols have a fetal-like phenotype far from adult cells (32). Although this limitation does not affect the study of prenatal stages of cardiac development, obtaining mature cardiac cells would broaden the scope of these models to study later stages of development as well as aging processes.

In vitro modeling of cardiac development-associated cellular diversity

Cardiomyocytes require substantial cell diversity to support both the proper execution of their biological functions and their differentiation, since many signaling pathways regulating their formation are sourced from other cell types (12,13). In this study, we confirmed that hiPSC cardiac differentiation generates the cellular diversity typically reported in the human fetal heart and thus provides the opportunity to investigate regulatory mechanisms occurring between these different cardiac cell types. However, hiPSC cardiac differentiation in two dimensions does not reproduce the spatial organization of the cell types as observed in the context of a heart. The emergence of more integrated hiPSC-derived models such as cardioids (33), will therefore undoubtedly enhance our insights into transcriptional regulation between cardiac cell types.

Uncovering new regulatory networks using a gene expression kinetics-based strategy

An original aspect of this study was the identification of expression regulations occurring between TFs in a temporal manner. For that we chose to adapt the LEAP bioinformatic tool designed for single-cell data to kinetic transcriptomic bulk data (34). Importantly, with this tool, these gene regulations are oriented, indicating not only the interaction but also which partner is expected to be the target/source. This higher level of information is important to design more efficiently confirmation experiment, and cannot be obtain in traditionally-used protein-protein interaction databases, such as STRING (35). Moreover, our strategy allowed to biologically link genes in a time-dependent manner during cardiac differentiation, and thus provided important new insights on cardiac gene regulatory networks (36). Of note, one cannot exclude that some of the inferred links may not reflect biological interactions (e.g. TF does not directly bind to an inferred target gene). Other studies embarked in different strategies to study cardiac gene regulation. For instance, Gonzalez-Teran et al. combined, PPI data associated with GATA4 and TBX5 TFs, and genetic data generated on patients presenting congenital heart diseases (CHD) to identify CHD candidate genes (37). This integrated strategy of PPI data and CHD-associated genetic data could be a relevant complementary approach of our chronological gene expression-based strategy in order to identify new CHD-associated TF regulatory networks and to offer a better understanding of cardiac disease underlying mechanisms.

Activation cascade of GATA4, NKX2-5, TBX5 genes triggered by IRX3 and IRX5

It is well established that cardiac transcription factors regulate the expression of other TF-coding genes. For instance, *GATA4* activates *NKX2-5* expression and both *GATA4* and *NKX2-5* activates *TBX5* expression (5,19–21). However, the precise molecular bases of these regulations were still to be uncovered. Using daily-generated transcriptomic data, we characterized the course of expression of these major cardiac TFs showing that, in accordance with the functional data, they are successively launched, starting with *GATA4* around day 5, followed by *NKX2-5* two days later and finally by *TBX5* two days later too. Obviously, this raised the question of how *GATA4* expression is, in the first place, launched. Using gene expression-based network we identified IRX3 and IRX5 TFs as potential activators of *GATA4* expression. Accordingly, the expression of these TFs was launched simultaneously about 1 day prior to *GATA4* expression. These TFs are of growing interest as, while most studies were performed

406 in knockout mice showing that they play redundant roles in cardiac development leading to
407 embryonic lethality, and in postnatal electrophysiological function, their role in human cardiac
408 function now also emerges (7,38). In this context, the present study therefore further explored
409 and specified the role of IRX TFs in the course of human cardiac development.

410 ***Exploring the functional interplay between IRX3/IRX5 and GATA4, NKX2-5, TBX5***

411 It is has been shown that GATA4, NKX2-5 and TBX5 act as multiprotein complex to
412 regulate cardiac gene expression (19). Here, we completed this knowledge by showing that
413 IRX3 and IRX5 can also physically bind to this TF regulatory complex. Furthermore, all five TFs
414 could physically and functionally interact on the promoter of *SCN5A* that encodes the major
415 cardiac sodium ion channel. Accordingly, *SCN5A* expression gradually increases during hiPSC
416 cardiac differentiation, paralleling the progressive expression establishment of the five TFs.
417 Some of the interactions between IRX TFs and GATA4, NKX2-5 or TBX5 have previously been
418 published. For instance, physical and functional interactions between Irx3, Nkx2-5 and Tbx5
419 have been shown in mice to regulate genes implicated in ventricular conduction system
420 establishment and maturation (39). Furthermore, our group has previously demonstrated
421 physical and functional interactions between IRX5 and GATA4 on *SCN5A* promoter (7). In this
422 study we further detailed the complexity of the interactions between IRX3, IRX5, GATA4,
423 NKX2-5 and TBX5, and how these TF complex compositions impact the expression of a target
424 gene.

425 ***Perspectives***

426 Altogether, this study provides a comprehensive dynamic blueprint of transcription
427 factors that control transcriptional regulation during human cardiac development as well as a
428 new methodological approach that may be applied to other research fields. These insights
429 may help to further understand both pathological cardiac development leading to CHD, as
430 well as physiological cardiac development, which is a prerequisite to emerging cardiac
431 regenerative therapy strategies (40). Moreover, in recent years, transcription factor
432 regulation of cardiac functions was widely supported by Genome Wide Association Studies,
433 linking numerous common genetic variations at loci harboring TF genes to cardiac diseases
434 ((41,42); Supplementary Table 1). Confronting the present knowledge to the one obtained
435 from cardiac differentiation of hiPSCs reprogrammed from patients carrying such genetic

variants may provide important information regarding their impact on cardiac development and therefore may lead to new targets for treatment and clinical management improvement.

Limitations of the study

We have identified several limitations in our study. First, the kinetic transcriptomic analysis has been performed using bulk-based strategy, however, the use of single-cell analysis instead would have provided us with a better overview of the cellular transcriptomic diversity. Second, the limited sample size that has been used prevented us from identifying the impact of the gender and ethnicity on transcriptomic regulation. Therefore, further studies will have to investigate if the identified TF networks (1) are activated in a cell-specific manner, and (2) whether they are specific to gender and/or ethnicity.

Materials and Methods

Reprogramming and maintenance of hiPSCs

All cell lines, from 3 healthy donors, were previously characterized. The hiPSC-A (C2a in (43)) line was generated using lentivirus method while hiPSC-B (IRX5-Wt in (44); RRID:CVCL_B5QD) and hiPSC-C (WT8288 in (45); RRID:CVCL_B5Q5) lines were generated using Sendai virus method. hiPSC lines were maintained at 37°C, 5% CO2, 21% O2 in StemMACS™ iPS Brew XF Medium (Miltenyi Biotec) on culture plates coated with Matrigel® hESC-Qualified Matrix (0.05 mg/mL, Corning). At 75% confluency, cells were passaged using Gentle Cell Dissociation Reagent (STEMCELL™ Technologies).

Cardiac differentiation of hiPSCs

Directed cardiac differentiations of hiPSCs were performed using the established matrix sandwich method (Fig 1A; (46)). Briefly, when hiPSCs reached 90% confluency, an overlay of Growth Factor Reduced Matrigel (0.033 mg/ml, BD Corning) was added. Differentiation was initiated 24h later by culturing the cells in RPMI1640 medium (Life Technologies) supplemented with B27 (without insulin, Life Technologies), 2 mM L-glutamine (Life Technologies), 1% NEAA (Life Technologies), 100 ng/mL Activin A (Miltenyi Biotec), 1X Pen/Strep (Life Technologies) and 10 ng/mL FGF2 for 24 hours. On the next day, the medium was replaced by RPMI1640 medium supplemented with B27 without insulin, 2 mM L-glutamine, 1% NEAA, 10 ng/mL BMP4 (Miltenyi Biotec), 1X Pen/Strep and 5 ng/mL FGF2 for 4

days. By day 5, cells were cultured in RPMI1640 medium supplemented with B27 complete (Life Technologies), 2 mM L-glutamine, 1X Pen/Strep and 1% NEAA and changed every two days until day 30. Specifically, for video analysis and immunofluorescence staining, glucose starvation was performed to obtain purified cardiomyocyte population: at day 10 the medium was replaced by Depletion medium (RPMI 1640 medium without glucose (Life Technologies) supplemented with B27 complete, and 1X Pen/Strep) for 3 days. Cells were dissociated at day 13 with 10X TrypLE solution (Life Technologies) and replated in CMs medium (RPMI1640 medium supplemented with B27 complete, 2 mM L-glutamine, 1X Pen/Strep, 1% NEAA) supplemented with Y-27632 Rho-kinase inhibitor (STEMCELL™ Technologies). On day 14, the medium was replaced by Depletion medium for 3 days. From day 17, cells were maintained in CMs medium.

Bulk transcriptomics

RNA extraction and sequencing

For each hiPSC line, samples were harvested daily from D-1 to D30 of the cardiac differentiation protocol from three independent cardiac differentiations. Total RNA were extracted using the NucleoSpin RNA kit (MACHEREY-NAGEL) and their quality assessed by NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific). From D-1 to D14 samples, all cells were collected while, from D15 to D30, to obtain samples enriched with cardiomyocytes, only spontaneously beating cell clusters were collected following mechanical isolation using a needle. 3'RNA libraries were prepared by GenoBiRD core facility according to their published method (47) and sequenced on 8 individual runs on a NovaSeq 6000 or HiSeq 2500 Sequencing System (Illumina).

Primary analysis of bulk transcriptomic data

Demultiplexing, alignment on GRCh38 reference genome and counting steps were conducted on each sequencing run with the Snakemake pipeline developed by the GenoBiRD core facility (47). Normalized and log-transformed expression matrices were generated using the *multiplates* function correcting potential batch effects by treating cardiac differentiation time points as replicates.

PCA

494 Principal Component Analysis (PCA) was performed with the R package *FactoMineR*
495 ((48); RRID:SCR_014602) on the entire mean-centered and log-transformed matrix.

496 *Time-course gene expression analysis*

497 Genes with significant expression variation between the different cardiac
498 differentiation time-points (indicated as Differentially Expressed Genes; DEG) were identified
499 by multivariate empirical Bayes statistics using the R package *timecourse* ((49);
500 RRID:SCR_000077) applied to the entire log-transformed matrix. We selected the top 3 000
501 DEG based on their highest Hotelling $\tilde{T}^2$ statistics. The same method was used to select genes
502 with significant expression variation during murine cardiac development from a published
503 transcriptomic dataset (14). When necessary, human and murine orthologous gene names
504 were identified using the R package *biomaRt* ((50); RRID:SCR_019214) and Ensembl databases.

505 *Clustering and heatmap*

506 DEG were grouped into clusters, based on their expression level variation across the
507 288 samples, using the R function *k-means* set on 2 000 iterations, and visualized with the R
508 package *ComplexHeatmap* ((51); RRID:SCR_017270).

509 *Gene Ontology analyses*

510 Gene Ontology (GO) analysis was performed using the R package *ClusterProfiler* ((52);
511 RRID:SCR_016884), based on GO Biological Process terms from org.Hs.eg.db_3.14.0 and
512 org.Mm.eg.db_3.14.0 databases for human and mouse annotations, as appropriate.
513 Significantly enriched (bonferroni-corrected p-value < 0.05) biological processes, as compared
514 to reference transcriptome, and with a Gene Set Size (GSSize) between 10 and 500, were
515 considered for further analysis. The 15 GO terms with the lowest corrected p-value were
516 visualized with treeplot.

517 *Network construction and analysis*

518 For each hiPSC line, the gene regulatory network was inferred using the R package
519 *LEAP* (Lag-based Expression Association for Pseudotime-series*;* Specht and Li, 2017), based on
520 the average from the log-transform data of triplicate cardiac differentiations. Cardiac
521 differentiation time points were used to rank samples as required by the *LEAP* tool. The
522 *max_lag_prop* parameter was set to 1/10, meaning that, at most, 3-day windows were used

to calculate the maximum absolute correlation (MAC) score. Only links with a significant MAC score (determined by a permutation test; *p-value* < 0.05) and related to a non-null time delay were considered. Links with a positive correlation score were interpreted as activation relationships and those with a negative correlation score as repression relationships. STRING software (35) was used to obtain information on physical and functional interactions between proteins of interest, with a *minimum required interaction score* of 0.4. Nodes without any interaction were excluded. STRING-based or LEAP-based interactions were processed using Cytoscape 3.9.1 for network reconstruction ((53); RRID:SCR_003032). Networks parameters were obtained using the *Analyze network* function.

Single-cell transcriptomic

Single-cell RNA-seq data generation

Cells at D30 of hiPSC-A cardiac differentiation were harvested from three distinct beating wells, dissociated, using the Multi Tissue Dissociation Kit 3 (Miltenyi Biotec), and pooled. This experiment was performed in duplicates. Cell suspensions were filtered on a 40 µm Flowmi® Cell Strainer, counted and cell viability was assessed (viability was 92% for the first experiment and 94% for the second). For each replicate, single-cell droplet libraries were generated from 16 000 cells with the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3 (10X Genomics). After qPCR quantification, libraries were pooled and sequenced on a single run, on a NovaSeq 6000 Sequencing System (Illumina), providing a read depth of >20,000 read pairs per cell, according to manufacturer's instructions.

Primary analysis of single-cell transcriptomic data

Data were processed using *cellranger* 4.0.0 (10X Genomics). First, demultiplexing of *raw base call* files into FASTQ files was accomplished using *cellranger mkfastq* function. Second, alignment on GRCh38 reference genome, filtering and counting steps were performed separately on each replicate with *cellranger count* function. Lastly, aggregation with normalization of duplicates was performed using *cellranger aggr* function.

Secondary analysis of single-cell transcriptomic data

The gene expression matrix was analyzed using the R package *Seurat* ((54); RRID:SCR_016341). Doublets were identified and removed using the R package *DoubletFinder*

((55); RRID:SCR_018771), assuming a 7.5% doublet formation rate. Also, only cells with 200 to 5 000 detected features and with <25% reads aligned to mitochondrial genes were selected for further analysis. After normalization, unwanted sources of intercellular variations such as number of detected genes or differences between cell cycle phases were regressed using the *ScaleData* function. A principal component analysis was then performed using the 2 000 most variable genes according to the *FindVariableFeatures* function and the first 10 components were used to calculate the UMAP. Cell-type labelling was performed using published single-cell RNA-seq data from a human fetal heart as a reference (4). Cell-type labels from reference were automatically transferred after cell-to-cell matching at the individual cell level using the R package *CellID* (56).

Musclemotion

hiPSC-CMs were filmed after glucose starvation at D30 in routine culture condition (37°C, 5% CO2), without electrical stimulation, using Nikon A1 RSI confocal microscope with X20 Dry N.A 0.75 objective. MUSCLEMOTION software (*Gaussian Blur*: No; *Speed Window*: 5; *Noise Reduction*: Yes; *Automatic Reference Frame Detection*: Yes; *Transient analysis*: Yes; (57)) was used to obtain contraction traces from 120fps videos. Contraction profiles were analyzed using homemade R pipeline.

HEK293 cell culture and transfection

HEK293 cells were maintained at 37°C, 5% CO2, in DMEM media with 10% FBS, 5% L-Glutamine and 5% Pen/Strep. Cells were plated in 24-well plate or 6-well plate and transfected next day using FuGENE® 6 (Promega, E2691). For luciferase assay, cells were transfected with a total of 2µg of plasmid including: (1) pGL2-Renilla plasmid, (2) plasmid containing Firefly luciferase gene upstream promoter of interest and (3) expression plasmids coding for proteins of interest (Table 1). DNA quantities were equalized in each condition using empty pcDNA3.1 plasmid. Media was changed 24h post transfection, and cell lyses performed 48h post transfection. For co-immunoprecipitation, cells were transfected only with expression plasmids prior lysis 24h post transfection.

Co-immunoprecipitation

Protein sample extraction and quantification

Previously transfected HEK293 cells were lysed (4°C, 15min, with rotation) in lysis buffer: 1% TritonX-100, 100mM NaCl, 50mM Tris-HCl, 1mM EGTA, 1mM Na3VO4, 50mM NaF, 1mM phenylmethylsulfonyl fluoride, protease inhibitors cocktail (Sigma-Aldrich, P8340), and centrifuged at 15 000g (4°C, 15 min). Protein quantification was carried out using Pierce™ BCA Protein Assay Kit (Thermo Fisher, 23225).

Bead-antibody complexes preparation

Co-Immunoprecipitation was performed using Dynabeads® Protein G (Invitrogen, 10004D) and DynaMag™-2 Magnet (Invitrogen, 12321D). First, 12.5µL of beads were conjugated (Room temperature (RT), 40min, with rotation) with 2µg of antibody (Table 1). The bead-antibody complexes were cross-linked (RT, 30min, with rotation) using 5,4mg/ml dimethyl pimelimidate (ThermoScientific, 21667). The cross-linking was quenched with 50mM Tris pH7.5 (RT, 15min, with rotation). Beads were washed using (1) PBS 1X, (2) 0.1M citrate pH3.1, (3) Na-phosphate solutions, then incubated in PBS 0.5% NaDOC (RT, 15min, with rotation) and were finally washed with lysis buffer.

Immunoprecipitation and western blotting analysis

Bead-antibody complexes were incubated with 1mg protein samples (4°C, 2h, with rotation). Supernatant was then discarded and beads were washed 3 times with lysis buffer. Beads-protein complexes were then heated (50°C, 10min) in NuPAGE™ LDS Sample Buffer (4X) (Invitrogen, NP0008). Samples were magnetized prior supernatants collection and incubated (70°C, 10min) in NuPAGE® Sample Reducing Agent 10X (Invitrogen, NP0009). Finally, samples were loaded onto a 4–15% precast polyacrylamide gel (Biorad, 4568083) together with 10µg of total protein used as control. Revelation was performed using corresponding antibody (Table 1) with ECL Clarity Max (Biorad, 1705062). Images were acquired with ChemiDoc camera (Biorad) and analysed using Image Lab Software (Biorad).

Table 1. Plasmids and antibodies references

Plasmid Name	Sequence/Reference	Supplier
Nkx2.5 promoter - FireflyLuc	-2000bp_Start codon	Vectorbuilder
GATA4 promoter - FireflyLuc	-1800_TSS_+200	Vectorbuilder
Tbx5 promoter - FireflyLuc	-1800_TSS_+200	Vectorbuilder
SCN5A promoter - FireflyLuc	-2109_TSS_ +1072	Adapted from (7)
pGL2 Renilla luciferase		Promega
IRX5	RG234228	Origene

IRX3	RG205722	Origene	
GATA4	RC210945	Origene	
Tbx5	SC120046	Origene	
Nkx2.5	SC122678	Origene	
pcDNA3.1		Invitrogen	
Antibody	**Reference**	**RRID**	**Supplier**
anti-GFP	TA150041	AB_2622256	Origene
anti-Myc Tag	05-724	AB_309938	Merck Millipore
anti-IRX5	sc-81102	AB_1124818	Santa Cruz
anti-IRX3	sc-166877	AB_10609525	Santa Cruz
anti-GATA4	sc-25310	AB_627667	Santa Cruz
anti-Tbx5	sc-515536		Santa Cruz
anti-Nkx2.5	sc-8697	AB_650280	Santa Cruz
anti-Troponin I	sc-15368	AB_793465	Santa Cruz
Mouse IgG Isotype Control	02-6502	AB_2532951	Thermo Fisher Scientific

Luciferase assay

Cells were lysed according to the manufacturer recommendations and luciferase activity was measured using Dual Luciferase reporter assay system (Promega, E1910) with Varioskan™ LUX microplate reader (Thermofisher). Mann-Whitney statistical tests were performed with Prism software (v8.0.1).

Immunofluorescence

Cells were fixed with 4% paraformaldehyde for 15min at room temperature (RT) in Matrigel®-coated µ-Slide 8 Well (IBIDI) prior permeabilization with 0.1% PBS-BSA 1% Saponin (RT, 15min) and blocking with 3% PBS-BSA (RT, 30min). Cells were then incubated with primary antibodies (dilution 1/250) in PBS 0.1% BSA 0.1% Saponin solution (4°C, overnight). Finally, cells were washed and incubated with secondary antibodies and DAPI (RT, 1h) and stored in 0.5% paraformaldehyde (4°C). Images were acquired using an inverted epifluorescence microscope (Zeiss Axiovert 200 M).

TF and cardiac phenotypes association

The association between cardiac phenotypes and transcription factors was performed using the DisGeNET (v7.0; (58)) and NHGRI-EBI GWAS Catalog (59) databases, filtering on cardiovascular traits, which were then manually validated.

Acknowledgments

We are most grateful to the Genomics Core Facility GenoA, member of Biogenouest and France Genomique and to the Bioinformatics Core Facility BiRD, member of Biogenouest and Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013) for the use of their resources and their technical support. We acknowledge the IBISA MicroPICell facility (Biogenouest), member of the national infrastructure France-Bioimaging supported by the French national research agency (ANR-10-INBS-04). We thank Jérôme Montnach for helping with muscle contraction analyses, Elise Douillard and Stéphane Minvielle for technical help with single-cell experiments, and Béatrice Guyomarch for helpful discussions on statistical analyses.

Fundings

B.C. is laureate of fellowships from the Fondation pour la Recherche Médicale (FRM, PBR201810007614) and the Fondation Genavie. R.C. is laureate of fellowship from the Fonds Marion Elizabeth Brancher. This work was funded by grants from The National Reasearch Agency (HEART iPS ANR-15-CE14-0019-01), from the *Fédération Française de Cardiologie,* and from International Incoming Fellowship FP7-PEOPLE-2012-IIF [PIIF-GA-2012-331436] to N.G, and by a grant from The French Regional Council of Pays de la Loire (RFI project VaCaRMe).

Author contributions

Conceptualization, R.C., B.C., R.R., P.L., N.G. and G.L., Data curation, R.C., B.C. and A.G., Formal Analysis, R.C. and B.C., Funding acquisition, R.R., P.L., N.G. and G.L., Investigation, R.C., B.C., A.G. and V.F., Methodology, R.C., B.C., A.G., N.G. and G.L., Project administration, R.R., P.L., N.G. and G.L., Software, R.C., B.C. and P.O., Supervision, J.P., R.R., P.L., N.G. and G.L., Validation, R.C., B.C., A.G., V.F., J.P., R.R., P.L., N.G. and G.L., Visualization, R.C., B.C., Writing – original draft, R.C., B.C., A.G., V.F., J.P., R.R., P.L., N.G. and G.L.

Declaration of interests

The authors declare no competing interests.

Data and code availability

650 The authors declare that all data supporting the findings of this study are available
651 within the article and its supplementary information files. Bulk and single-cell transcriptomic
652 data are available in ArrayExpress database at EMBL-EBI
653 (https://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-11822 and E-MTAB-
654 11817, respectively. All data will be publicly available as of the date of publication. All original
655 code have been deposited at Gitlab (https://gitlab.univ-nantes.fr/E132534J/cardiff.git) and
656 will be publicly available as of the date of publication. Any additional information required to
657 reanalyze the data reported in this work is available from the lead contact upon reasonable
658 request.

References

1. Meilhac SM, Buckingham ME. The deployment of cell lineages that form the mammalian heart. Nat Rev Cardiol. 2018 Nov;15(11):705–24.

2. Vincent SD, Buckingham ME. How to make a heart: the origin and regulation of cardiac progenitor cells. Curr Top Dev Biol. 2010;90:1–41.

3. Postma AV, Christoffels VM, Bezzina CR. Developmental aspects of cardiac arrhythmogenesis. Cardiovasc Res. 2011 Jul 15;91(2):243–51.

4. Asp M, Giacomello S, Larsson L, Wu C, Fürth D, Qian X, et al. A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. Cell. 2019 Dec 12;179(7):1647-1660.e19.

5. Bruneau BG. Signaling and transcriptional networks in heart development and regeneration. Cold Spring Harb Perspect Biol. 2013 Mar 1;5(3):a008292.

6. Luna-Zurita L, Stirnimann CU, Glatt S, Kaynak BL, Thomas S, Baudin F, et al. Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. Cell. 2016 Feb 25;164(5):999–1014.

7. Al Sayed ZR, Canac R, Cimarosti B, Bonnard C, Gourraud JB, Hamamy H, et al. Human model of IRX5 mutations reveals key role for this transcription factor in ventricular conduction. Cardiovasc Res. 2021 Jul 27;117(9):2092–107.

8. Bruneau BG, Bao ZZ, Fatkin D, Xavier-Neto J, Georgakopoulos D, Maguire CT, et al. Cardiomyopathy in Irx4-deficient mice is preceded by abnormal ventricular gene expression. Mol Cell Biol. 2001 Mar;21(5):1730–6.

9. Costantini DL, Arruda EP, Agarwal P, Kim KH, Zhu Y, Zhu W, et al. The Homeodomain Transcription Factor Irx5 Establishes the Mouse Cardiac Ventricular Repolarization Gradient. Cell. 2005 Oct 21;123(2):347–58.

10. Gaborit N, Sakuma R, Wylie JN, Kim KH, Zhang SS, Hui CC, et al. Cooperative and antagonistic roles for Irx3 and Irx5 in cardiac morphogenesis and postnatal physiology. Dev Camb Engl. 2012 Nov;139(21):4007–19.

11. Zhang SS, Kim KH, Rosen A, Smyth JW, Sakuma R, Delgado-Olguín P, et al. Iroquois homeobox gene 3 establishes fast conduction in the cardiac His–Purkinje network. Proc Natl Acad Sci U S A. 2011 Aug 16;108(33):13576–81.

12. Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL, et al. Cells of the adult human heart. Nature. 2020 Dec;588(7838):466–72.

13. Meilhac SM. Cardiac growth I: Cardiomyocyte proliferation. ESC Textb Cardiovasc Dev. 2018;75.

14. Li X, Martinez-Fernandez A, Hartjes KA, Kocher JPA, Olson TM, Terzic A, et al. Transcriptional atlas of cardiogenesis maps congenital heart disease interactome. Physiol Genomics. 2014 Jul 1;46(13):482–95.

15. DeLaughter DM, Bick AG, Wakimoto H, McKean D, Gorham JM, Kathiriya IS, et al. Single-Cell Resolution of Temporal Gene Expression during Heart Development. Dev Cell. 2016 Nov 21;39(4):480–90.

16. Ng AHM, Khoshakhlagh P, Rojo Arias JE, Pasquini G, Wang K, Swiersy A, et al. A comprehensive library of human transcription factors for cell fate engineering. Nat Biotechnol. 2021 Apr;39(4):510–9.

17. Ang YS, Rivas RN, Ribeiro AJS, Srivas R, Rivera J, Stone NR, et al. Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. Cell. 2016 Dec 15;167(7):1734-1749.e22.

18. Durocher D, Charron F, Warren R, Schwartz RJ, Nemer M. The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. EMBO J. 1997 Sep 15;16(18):5687–96.

19. McCulley DJ, Black BL. Transcription factor pathways and congenital heart disease. Curr Top Dev Biol. 2012;100:253–77.

20. Nemer G, Nemer M. GATA4 in Heart Development and Disease. Heart Dev Regen. 2010 Dec 31;599–616.

21. Waardenberg AJ, Ramialison M, Bouveret R, Harvey RP. Genetic Networks Governing Heart Development. Cold Spring Harb Perspect Med. 2014 Nov 1;4(11):a013839.

22. Arnolds DE, Liu F, Fahrenbach JP, Kim GH, Schillinger KJ, Smemo S, et al. TBX5 drives Scn5a expression to regulate cardiac conduction system function. J Clin Invest. 2012 Jul;122(7):2509–18.

23. Briggs LE, Takeda M, Cuadra AE, Wakimoto H, Marks MH, Walker AJ, et al. Perinatal Loss of Nkx2-5 Results in Rapid Conduction and Contraction Defects. Circ Res. 2008 Sep 12;103(6):580–90.

24. Koizumi A, Sasano T, Kimura W, Miyamoto Y, Aiba T, Ishikawa T, et al. Genetic defects in a His-Purkinje system transcription factor, IRX3, cause lethal cardiac arrhythmias. Eur Heart J. 2016 07;37(18):1469–75.

25. Tarradas A, Pinsach-Abuin ML, Mackintosh C, Llorà-Batlle O, Pérez-Serra A, Batlle M, et al. Transcriptional regulation of the sodium channel gene (SCN5A) by GATA4 in human heart. J Mol Cell Cardiol. 2017 Jan;102:74–82.

26. He W, Jia Y, Takimoto K. Interaction between transcription factors Iroquois proteins 4 and 5 controls cardiac potassium channel Kv4.2 gene transcription. Cardiovasc Res. 2009 Jan 1;81(1):64–71.

27. Krishnan A, Samtani R, Dhanantwari P, Lee E, Yamada S, Shiota K, et al. A detailed comparison of mouse and human cardiac development. Pediatr Res. 2014 Dec;76(6):500–7.

28. Anzai T, Yamagata T, Uosaki H. Comparative Transcriptome Landscape of Mouse and Human Hearts. Front Cell Dev Biol. 2020;8:268.

29. Olson EN, Srivastava D. Molecular pathways controlling heart development. Science. 1996 May 3;272(5262):671–6.

30. Kathiriya IS, Rao KS, Iacono G, Devine WP, Blair AP, Hota SK, et al. Modeling Human TBX5 Haploinsufficiency Predicts Regulatory Networks for Congenital Heart Disease. Dev Cell. 2021 Feb 8;56(3):292-309.e9.

31. Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, et al. Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. Cell. 2012 Sep 28;151(1):206–20.

32. Yoshida Y, Yamanaka S. Induced Pluripotent Stem Cells 10 Years Later: For Cardiac Applications. Circ Res. 2017 Jun 9;120(12):1958–68.

33. Hofbauer P, Jahnel SM, Papai N, Giesshammer M, Deyett A, Schmidt C, et al. Cardioids reveal self-organizing principles of human cardiogenesis. Cell. 2021 Jun 10;184(12):3299-3317.e22.

34. Specht AT, Li J. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. Bioinforma Oxf Engl. 2017 Mar 1;33(5):764–6.

35. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1):D607–13.

36. Wang Y, Yi N, Hu Y, Zhou X, Jiang H, Lin Q, et al. Molecular Signatures and Networks of Cardiomyocyte Differentiation in Humans and Mice. Mol Ther Nucleic Acids. 2020 Sep 4;21:696–711.

37. Gonzalez-Teran B, Pittman M, Felix F, Thomas R, Richmond-Buccola D, Hüttenhain R, et al. Transcription factor protein interactomes reveal genetic determinants in heart disease. Cell. 2022 Mar 3;185(5):794-814.e30.

38. Bonnard C, Strobl AC, Shboul M, Lee H, Merriman B, Nelson SF, et al. Mutations in IRX5 impair craniofacial development and germ cell migration via SDF1. Nat Genet. 2012 May 13;44(6):709–13.

39. Kimura Y, Aiba T, Sasano T, Furukawa T, Kusano K, Shimizu W. IRX3 variant as a modifier of Brugada syndrome with frequent ventricular fibrillation. Hear Case Rep. 2016 Aug 2;2(6):465–8.

765 40. Rowton M, Guzzetta A, Rydeen AB, Moskowitz IP. Control of cardiomyocyte
766 differentiation timing by intercellular signaling pathways. Semin Cell Dev Biol. 2021
767 Oct;118:94–106.

768 41. Barc J, Tadros R, Glinge C, Chiang DY, Jouni M, Simonet F, et al. Genome-wide association
769 analyses identify new Brugada syndrome risk loci and highlight a new mechanism of
770 sodium channel regulation in disease susceptibility. Nat Genet. 2022 Mar;54(3):232–9.

771 42. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, et al. Multi-ethnic
772 genome-wide association study for atrial fibrillation. Nat Genet. 2018 Sep;50(9):1225–33.

773 43. Si-Tayeb K, Noto FK, Nagaoka M, Li J, Battle MA, Duris C, et al. Highly efficient generation
774 of human hepatocyte-like cells from induced pluripotent stem cells. Hepatol Baltim Md.
775 2010 Jan;51(1):297–305.

776 44. Canac R, Caillaud A, Cimarosti B, Girardeau A, Hamamy H, Reversade B, et al. Generation
777 of three human induced pluripotent stem cell lines with IRX5 knockout and knockin
778 genetic editions using CRISPR-Cas9 system. Stem Cell Res. 2022 Jan;58:102627.

779 45. Girardeau A, Atticus D, Canac R, Cimarosti B, Caillaud A, Chariau C, et al. Generation of
780 human induced pluripotent stem cell lines from four unrelated healthy control donors
781 carrying European genetic background. Stem Cell Res. 2021 Dec 28;59:102647.

782 46. Es-Salah-Lamoureux Z, Jouni M, Malak OA, Belbachir N, Al Sayed ZR, Gandon-Renard M,
783 et al. HIV-Tat induces a decrease in IKr and IKsvia reduction in phosphatidylinositol-(4,5)-
784 bisphosphate availability. J Mol Cell Cardiol. 2016 Oct;99:1–13.

785 47. Charpentier E, Cornec M, Dumont S, Meistermann D, Bordron P, David L, et al. 3'RNA
786 sequencing for robust and low-cost gene expression profiling. Protoc Exch. 2021 Jan 28;

787 48. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. J Stat Softw.
788 2008;25:1–18.

789 49. Tai YC. timecourse: Statistical Analysis for Developmental Microarray Time Course Data
790 [Internet]. Bioconductor version: Release (3.14); 2022 [cited 2022 Apr 26]. Available from:
791 https://bioconductor.org/packages/timecourse/

792 50. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of
793 genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–
794 91.

795 51. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
796 multidimensional genomic data. Bioinforma Oxf Engl. 2016 Sep 15;32(18):2847–9.

797 52. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment
798 tool for interpreting omics data. Innov N Y N. 2021 Aug 28;2(3):100141.

799 53. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software
800 environment for integrated models of biomolecular interaction networks. Genome Res.
801 2003 Nov;13(11):2498–504.

802 54. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive
803 Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21.

804 55. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA
805 Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019 Apr 24;8(4):329-
806 337.e4.

807 56. Cortal A, Martignetti L, Six E, Rausell A. Gene signature extraction and cell identity
808 recognition at the single-cell level with Cell-ID. Nat Biotechnol. 2021 Sep;39(9):1095–102.

809 57. Sala L, van Meer BJ, Tertoolen LGJ, Bakkers J, Bellin M, Davis RP, et al. MUSCLEMOTION: A
810 Versatile Open Software Tool to Quantify Cardiomyocyte and Cardiac Muscle Contraction
811 In Vitro and In Vivo. Circ Res. 2018 02;122(3):e5–16.

812 58. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The
813 DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020
814 Jan 8;48(D1):D845–55.

815 59. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The
816 NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays
817 and summary statistics 2019. Nucleic Acids Res. 2019 Jan 8;47(D1):D1005–12.

818

Supplemental information titles and legends

Fig S1 – Transcriptomic signatures of cell types generated by hiPSC cardiac differentiation.

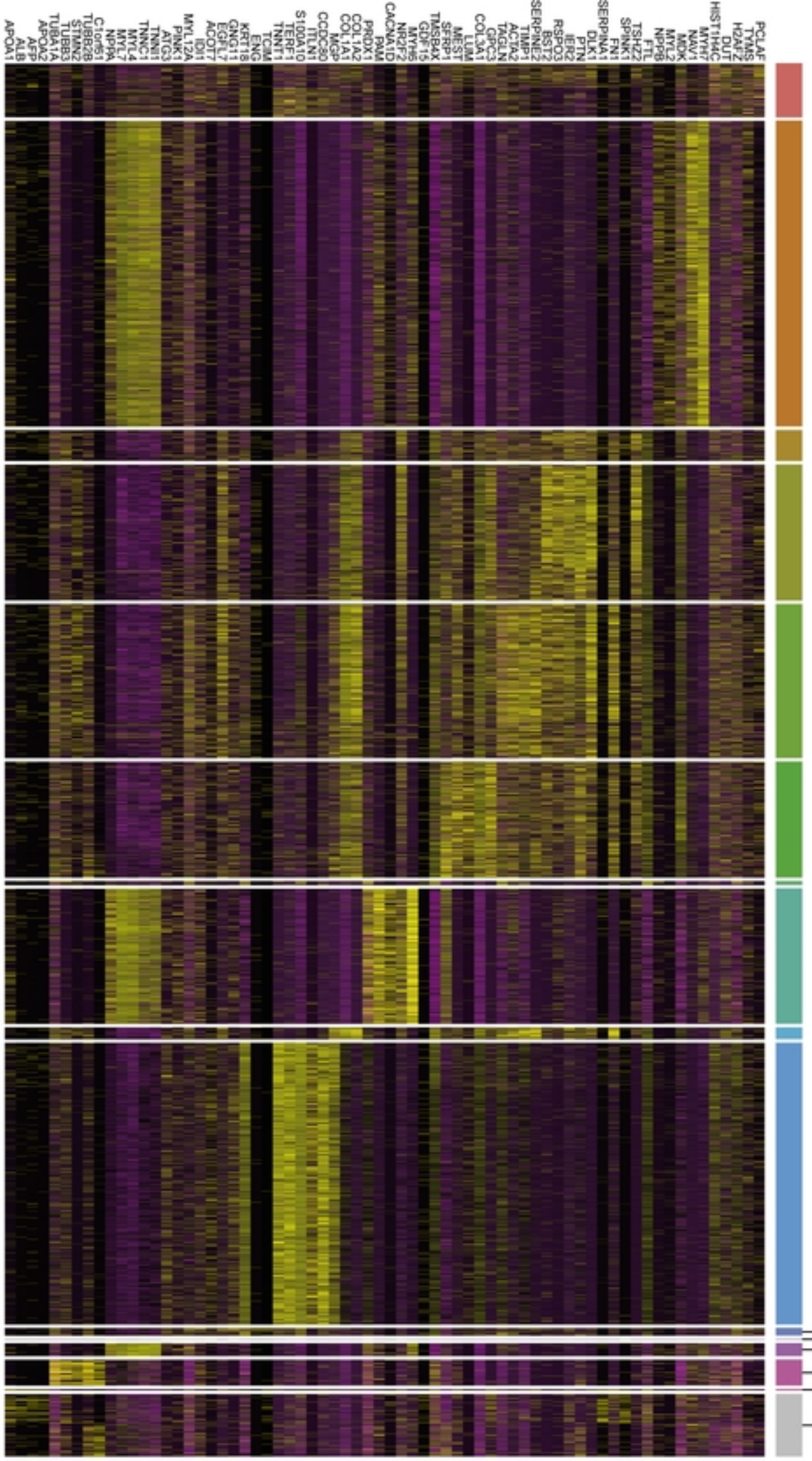
Heatmap displaying expression levels of the top 5 markers of each cell population found in Fig 1D. Markers were identified using the FindAllMarkers function from the R package Seurat.

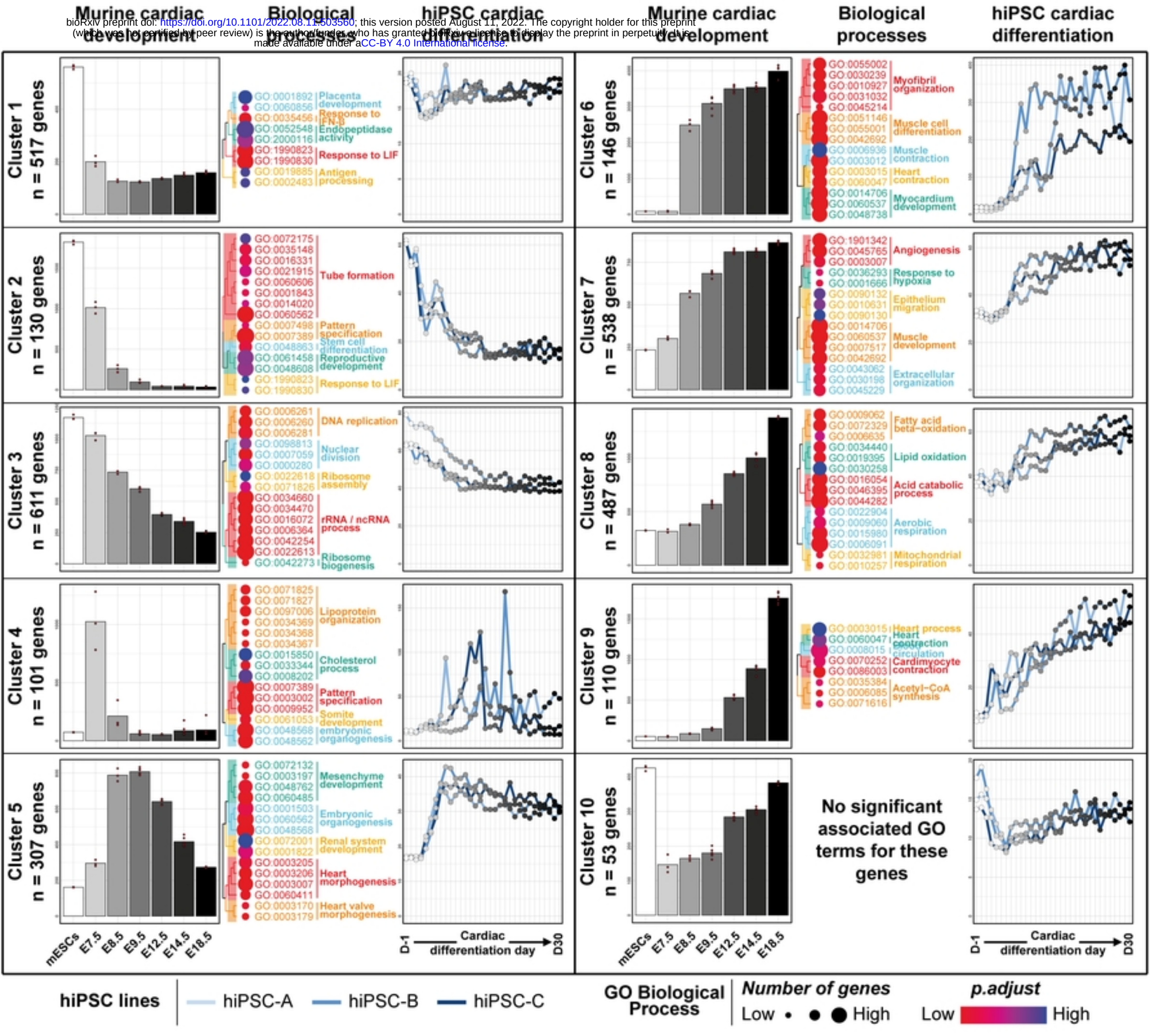
Fig S2 – Expression profile and functional annotation of the top 3 000 differentially expressed genes during murine cardiac development, and comparison with hiPSC cardiac differentiation gene expression dataset.

For each cluster, average gene expression level during murine cardiac development (left panel for each cluster) and their human orthologs during hiPSC cardiac differentiation (right panel for each cluster) are shown. Replicates of gene expression levels were averaged for murine data (n=3 to n=6 per timepoint, depending on the stage) and hiPSC cardiac differentiation (n=3 per hiPSC line and per timepoint). The 15 most significantly related GO terms are displayed on the middle panel for each cluster.

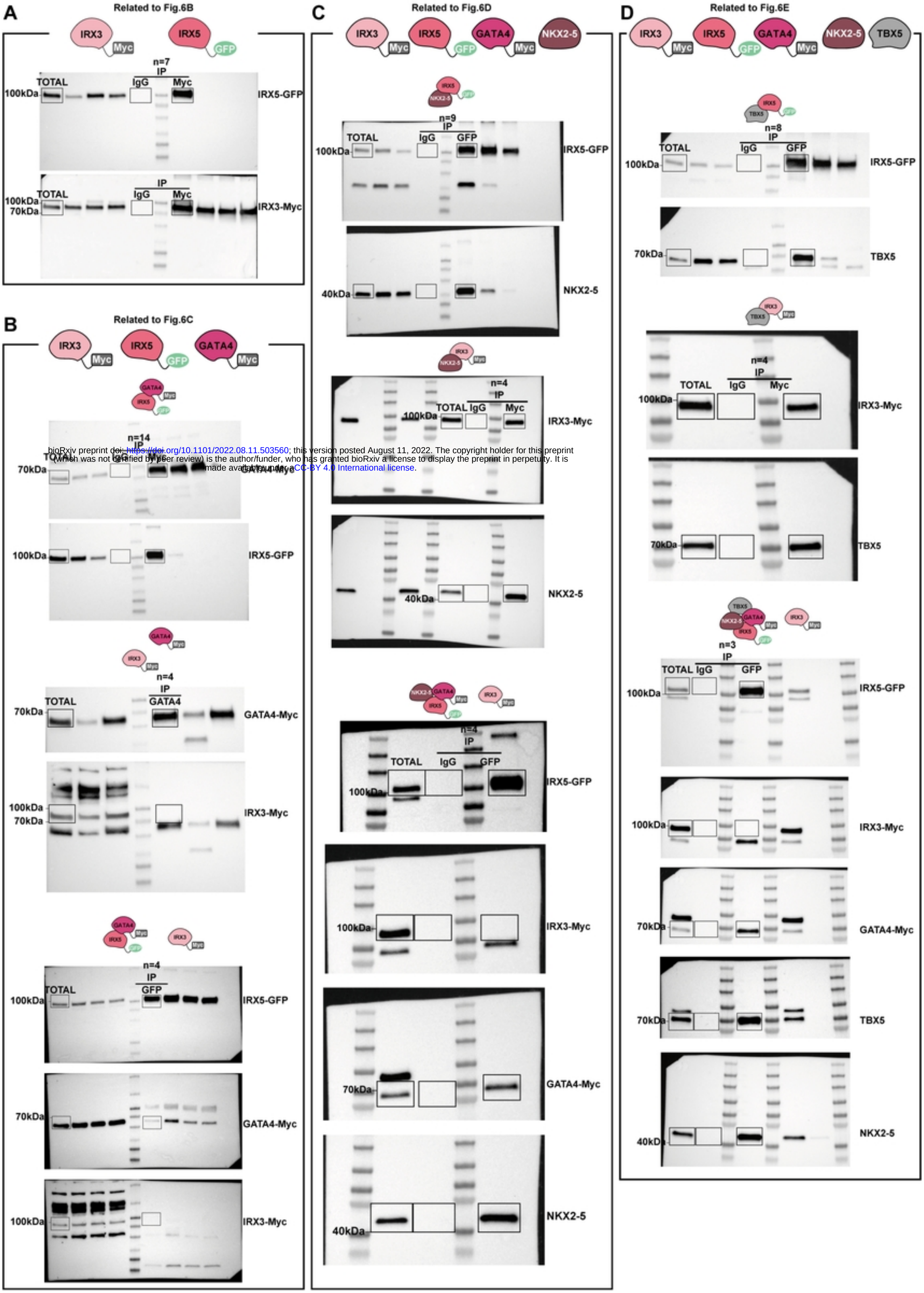
Fig S3 – Original western blots related to Fig 6.

S1 Fig


APOA1
ALB
AFP
APOA2
TUBA1A
TUBB3
STMN2
TUBB2B
C1orf61
NPPA
TNNI1
TNNC1
MYL4
MYL7
ATG3
PINK1
MYL12A
ID1
ACOT7
EGFL7
GNG11
KRT18
ENG
TCIM
TNNT1
TERF1
S100A10
ITLN1
CCDC80
MGP
COL1A1
COL1A2
PRDX1
PAM
CACNA1D
NR2F2
MYH6
GDF15
TMSB4X
SFRP1
MEST
LUM
COL3A1
GPC3
TAGLN
ACTA2
TIMP1
SERPINE2
BST2
RSPO3
IER2
PTN
DLK1
SERPINA1
FN1
SPINK1
TSHZ2
FTL
NPPB
MYL2
MDK
NAV1
MYH7
HIST1H4C
DUT
H2AFZ
TYMS
PCLAF
Capillary endothelium
Ventricular cardiomyocytes
Fibroblast−like
(cardiac skeleton connective tissue)
Epicardium−derived cells
Fibroblast−like
(smaller vascular development)
Smooth muscle cells/fibroblast−like
Erythrocytes
Atrial cardiomyocytes
Fibroblast−like
(larger vascular development)
Epicardial cells
Endothelium/pericytes/adventitia
Erythrocytes
Myoz2−enriched cardiomyocytes
Immune cells
Cardiac neural crest cells &
Schwann progenitor cells
Unassigned
Relative expression
Low
High



S2 Fig



S3 Fig

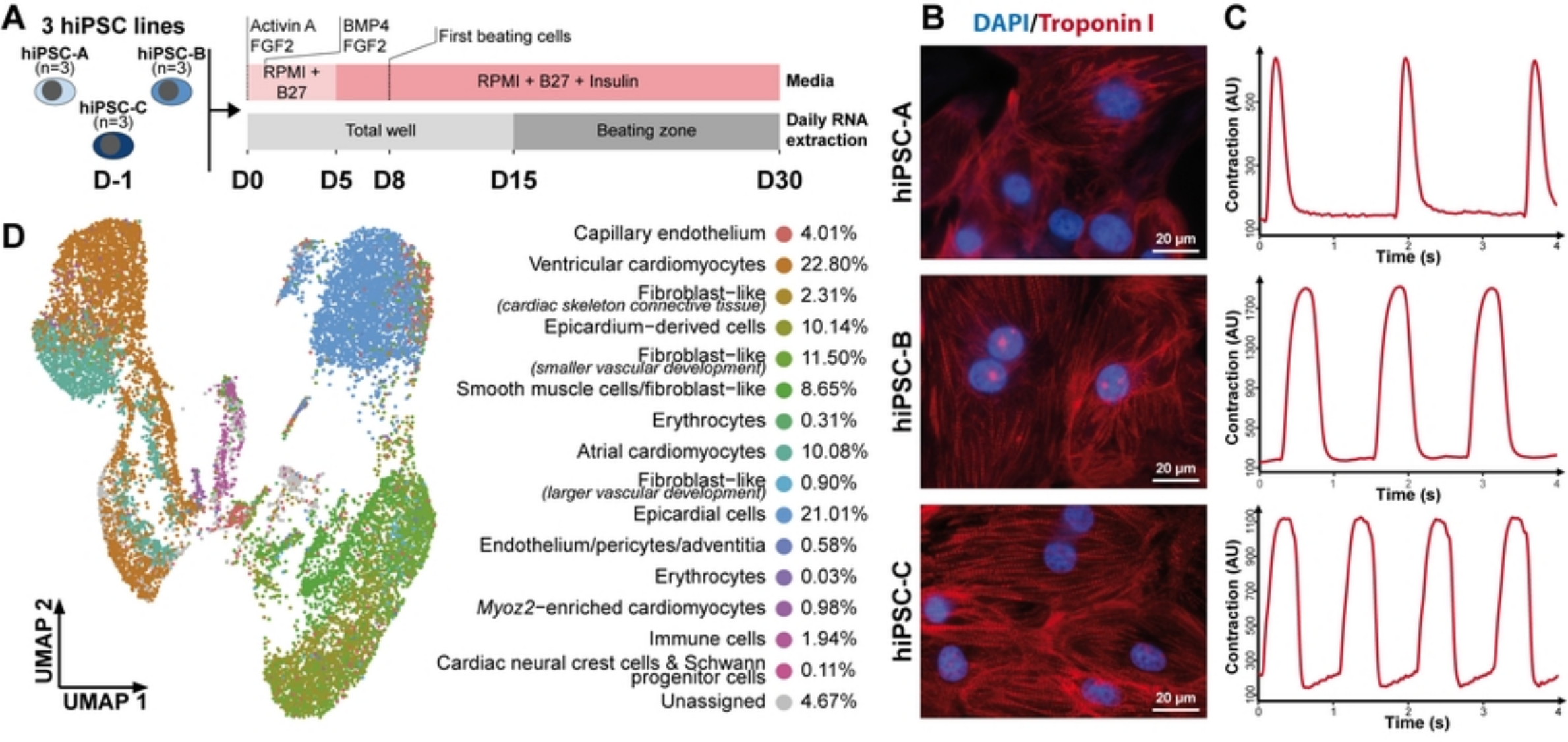

A
3 hiPSC lines
hiPSC-A (n=3)
hiPSC-B (n=3)
hiPSC-C (n=3)
Activin A FGF2
BMP4 FGF2
First beating cells
RPMI + B27
RPMI + B27 + Insulin
Media
Total well
Beating zone
Daily RNA extraction
D-1
D0
D5
D8
D15
D30
B
DAPI/Troponin I
hiPSC-A
hiPSC-B
hiPSC-C
20 µm
C
Contraction (AU)
Time (s)
D
UMAP 2
UMAP 1
Capillary endothelium 4.01%
Ventricular cardiomyocytes 22.80%
Fibroblast-like (cardiac skeleton connective tissue) 2.31%
Epicardium-derived cells 10.14%
Fibroblast-like (smaller vascular development) 11.50%
Smooth muscle cells/fibroblast-like 8.65%
Erythrocytes 0.31%
Atrial cardiomyocytes 10.08%
Fibroblast-like (larger vascular development) 0.90%
Epicardial cells 21.01%
Endothelium/pericytes/adventitia 0.58%
Erythrocytes 0.03%
Myoz2-enriched cardiomyocytes 0.98%
Immune cells 1.94%
Cardiac neural crest cells & Schwann progenitor cells 0.11%
Unassigned 4.67%

Fig 1

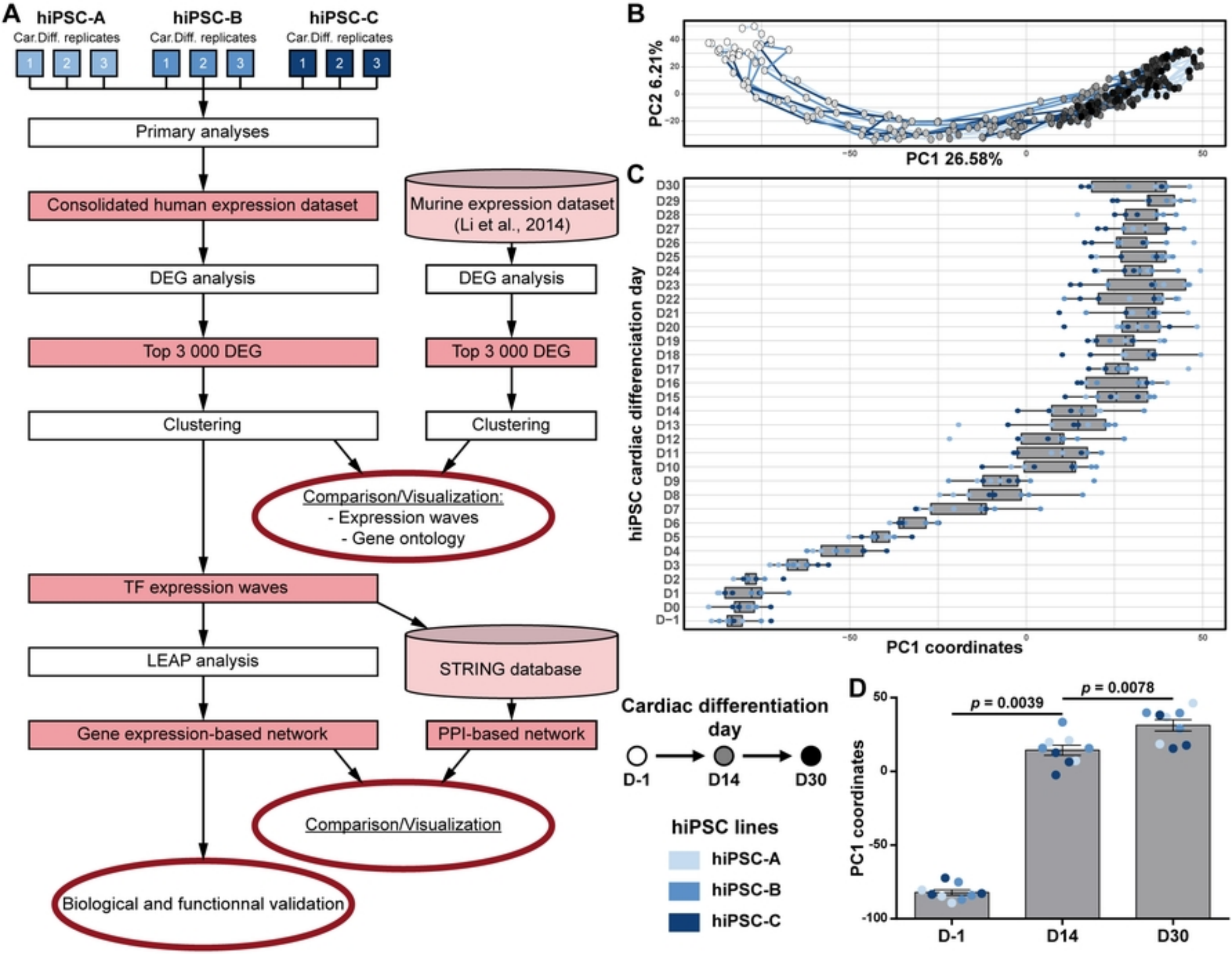

A
hiPSC-A
Car.Diff. replicates
1
2
3
hiPSC-B
Car.Diff. replicates
1
2
3
hiPSC-C
Car.Diff. replicates
1
2
3
Primary analyses
Consolidated human expression dataset
DEG analysis
Top 3 000 DEG
Clustering
Murine expression dataset
(Li et al., 2014)
DEG analysis
Top 3 000 DEG
Clustering
Comparison/Visualization:
- Expression waves
- Gene ontology
TF expression waves
LEAP analysis
STRING database
Gene expression-based network
PPI-based network
Comparison/Visualization
Biological and functionnal validation
B
PC2 6.21%
PC1 26.58%
C
hiPSC cardiac differenciation day
D30
D29
D28
D27
D26
D25
D24
D23
D22
D21
D20
D19
D18
D17
D16
D15
D14
D13
D12
D11
D10
D9
D8
D7
D6
D5
D4
D3
D2
D1
D0
D−1
PC1 coordinates
Cardiac differentiation day
D-1
D14
D30
hiPSC lines
hiPSC-A
hiPSC-B
hiPSC-C
D
p = 0.0039
p = 0.0078
PC1 coordinates
D-1
D14
D30

Fig 2

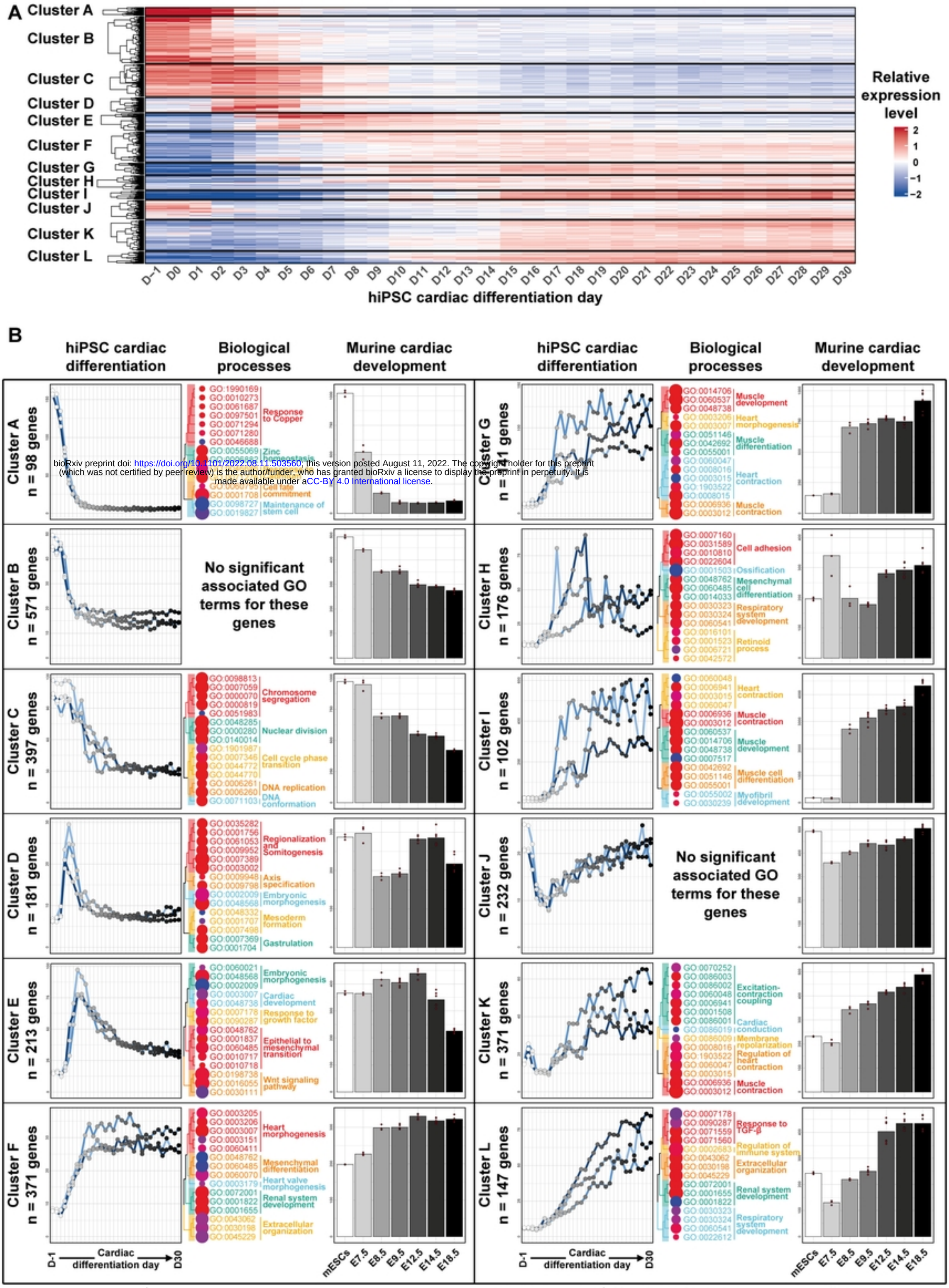


Fig 3

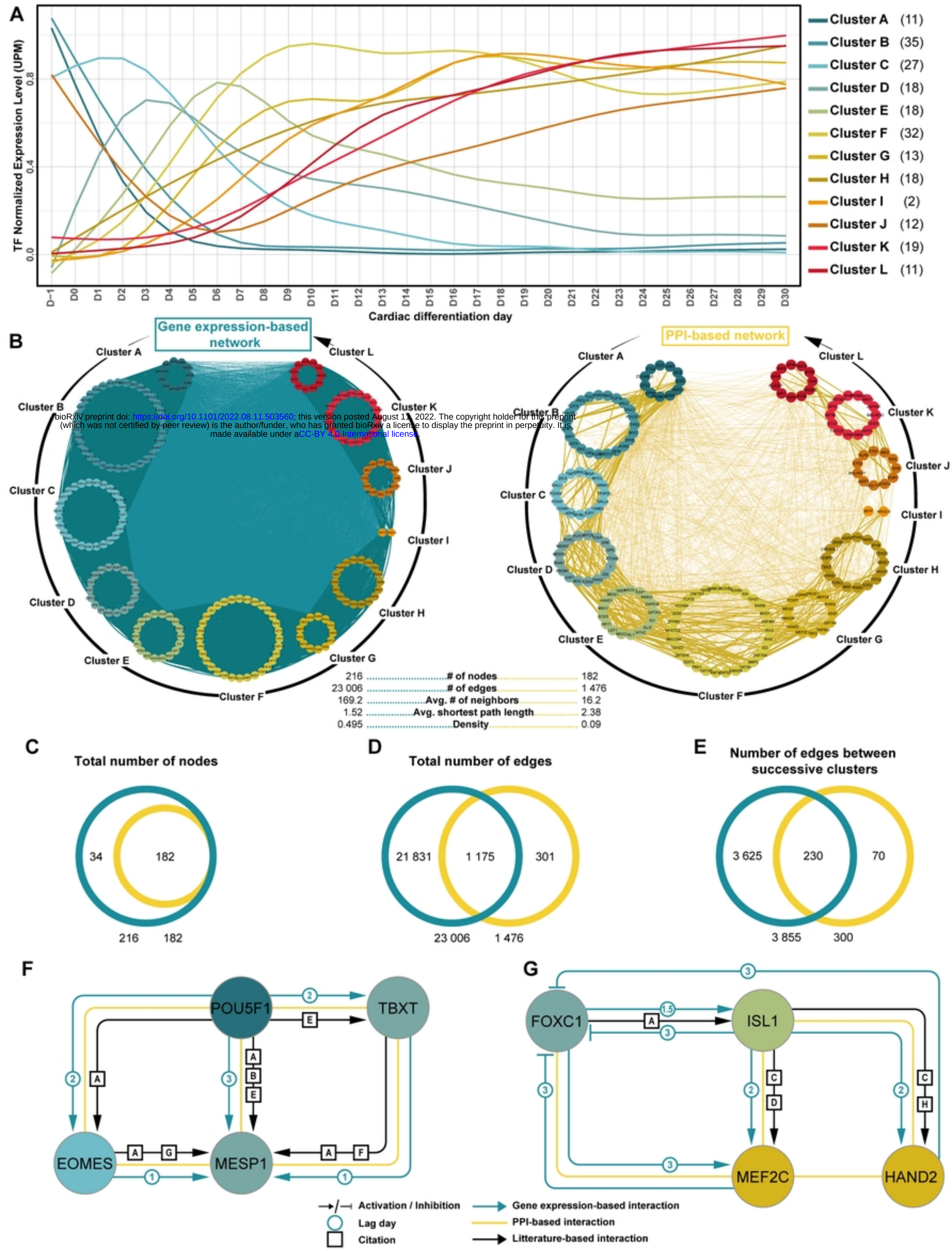
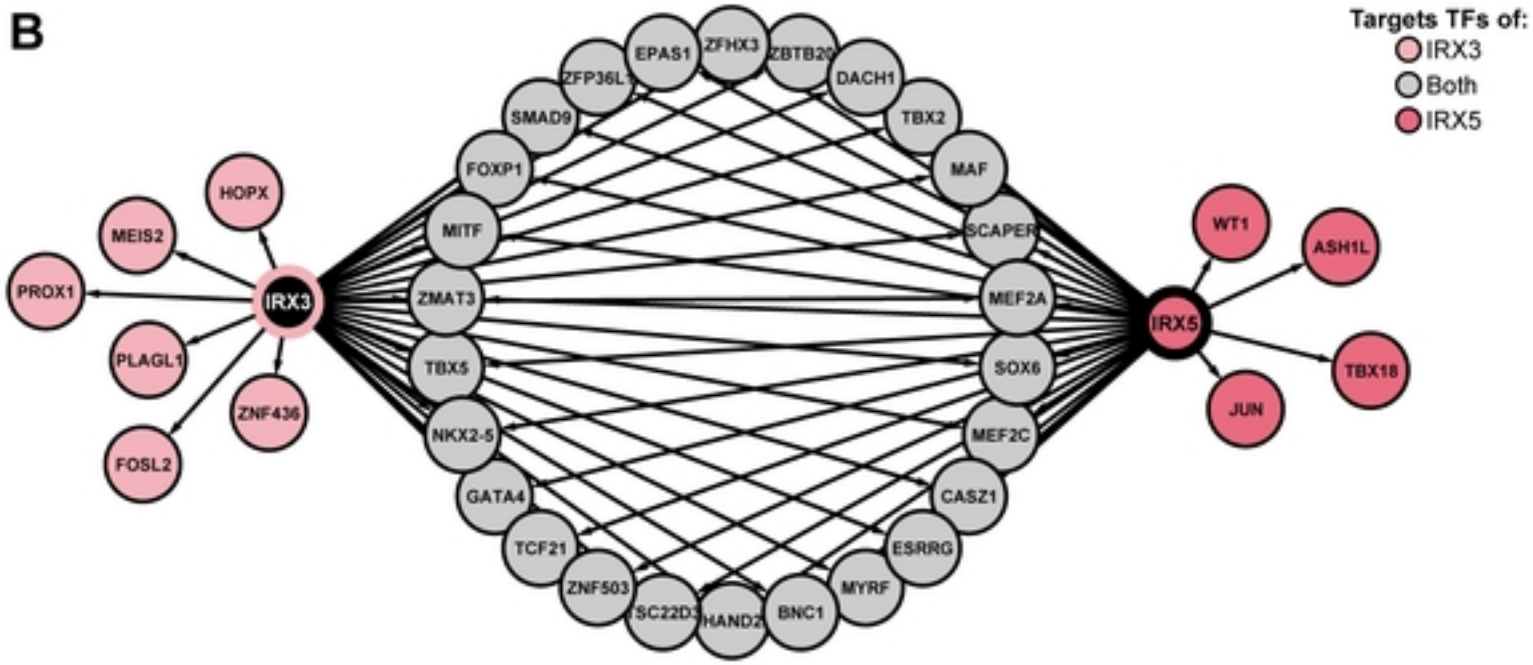
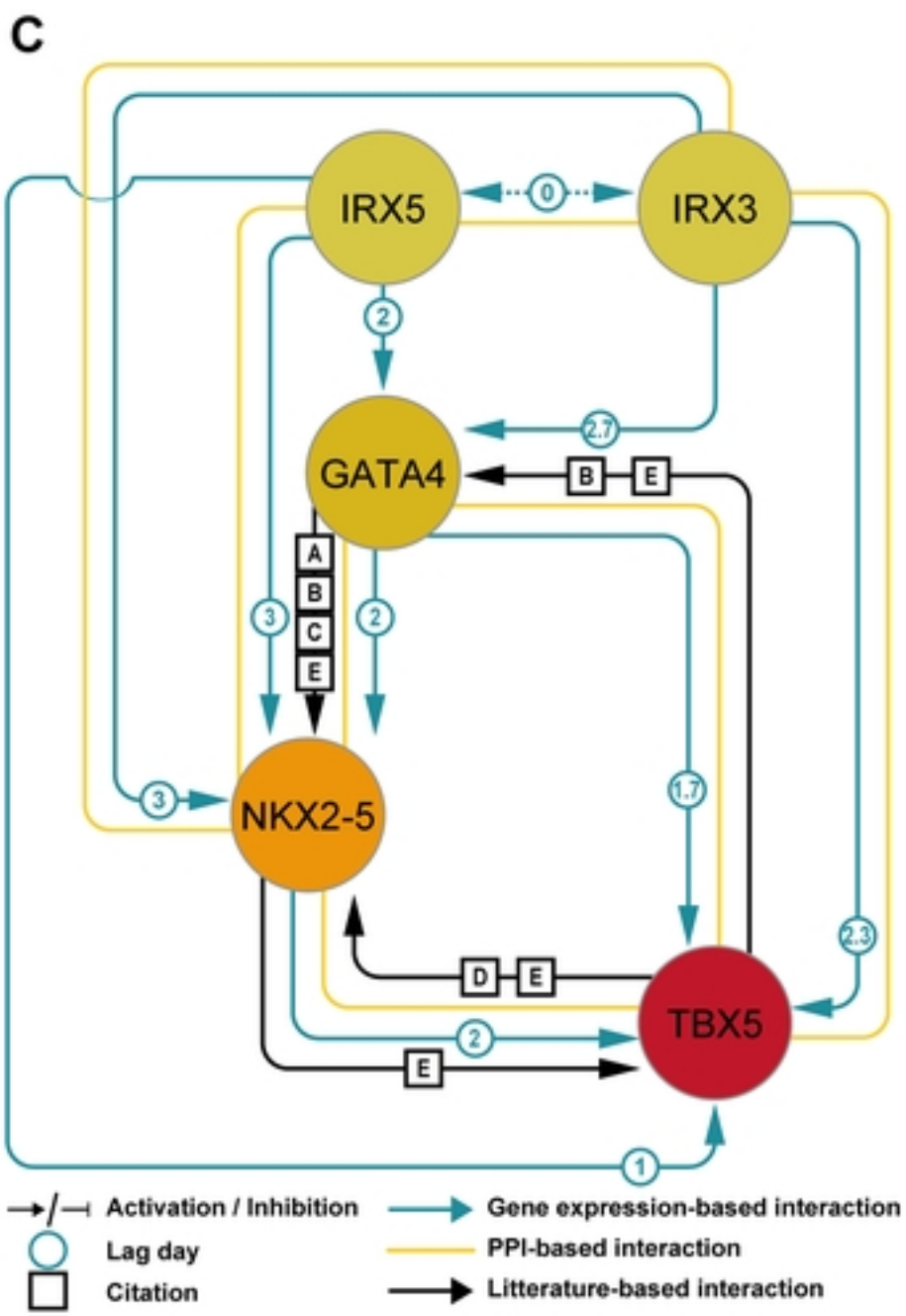


Fig 4

A

	Detected expression	Timecourse expression rank	Expression cluster
IRX1	✓	20 177	
IRX2	✓	20 165	
IRX3	✓	187	F
IRX4	✓	1 592	K
IRX5	✓	831	F
IRX6	✗	19 604	





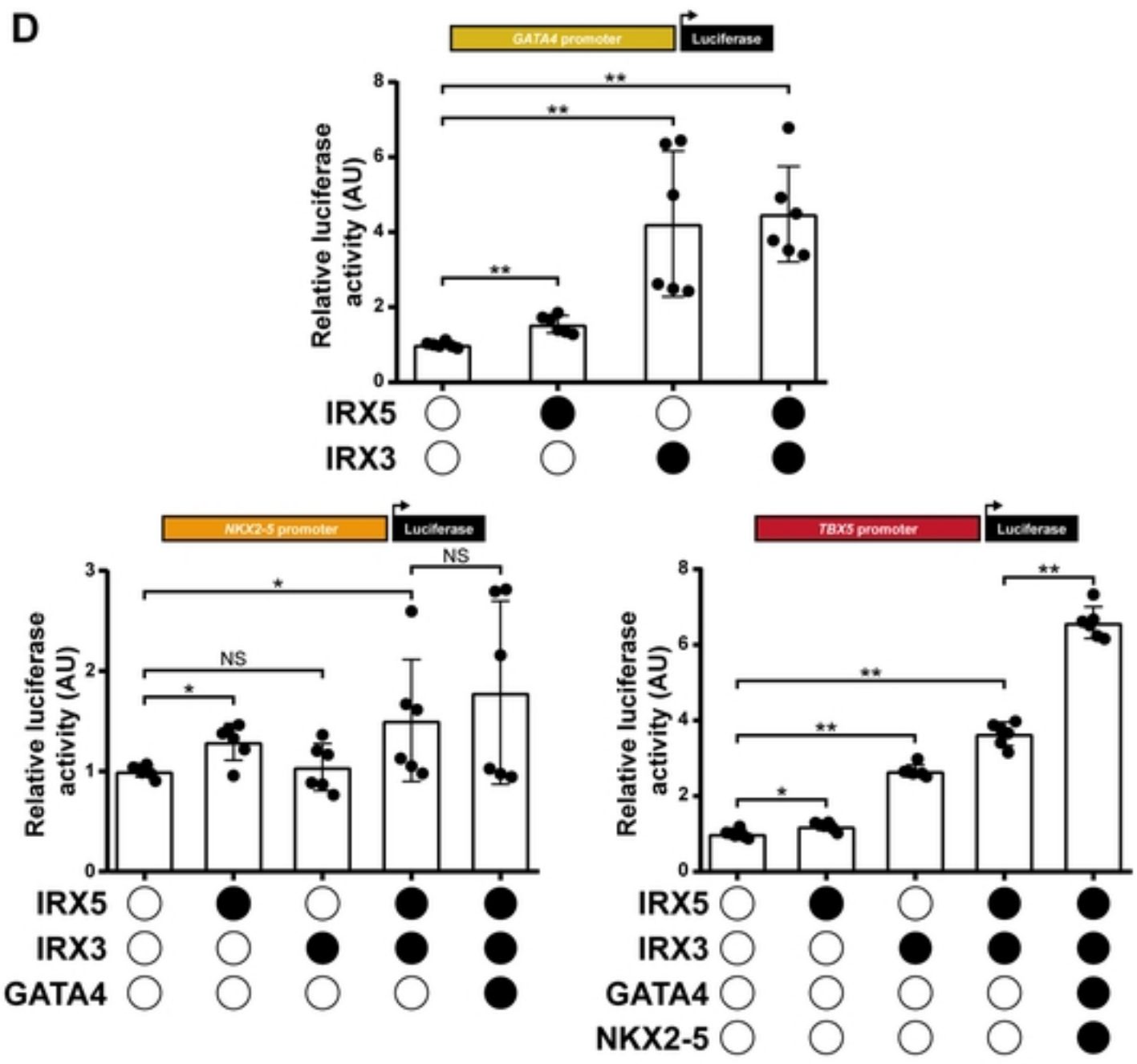


Fig 5

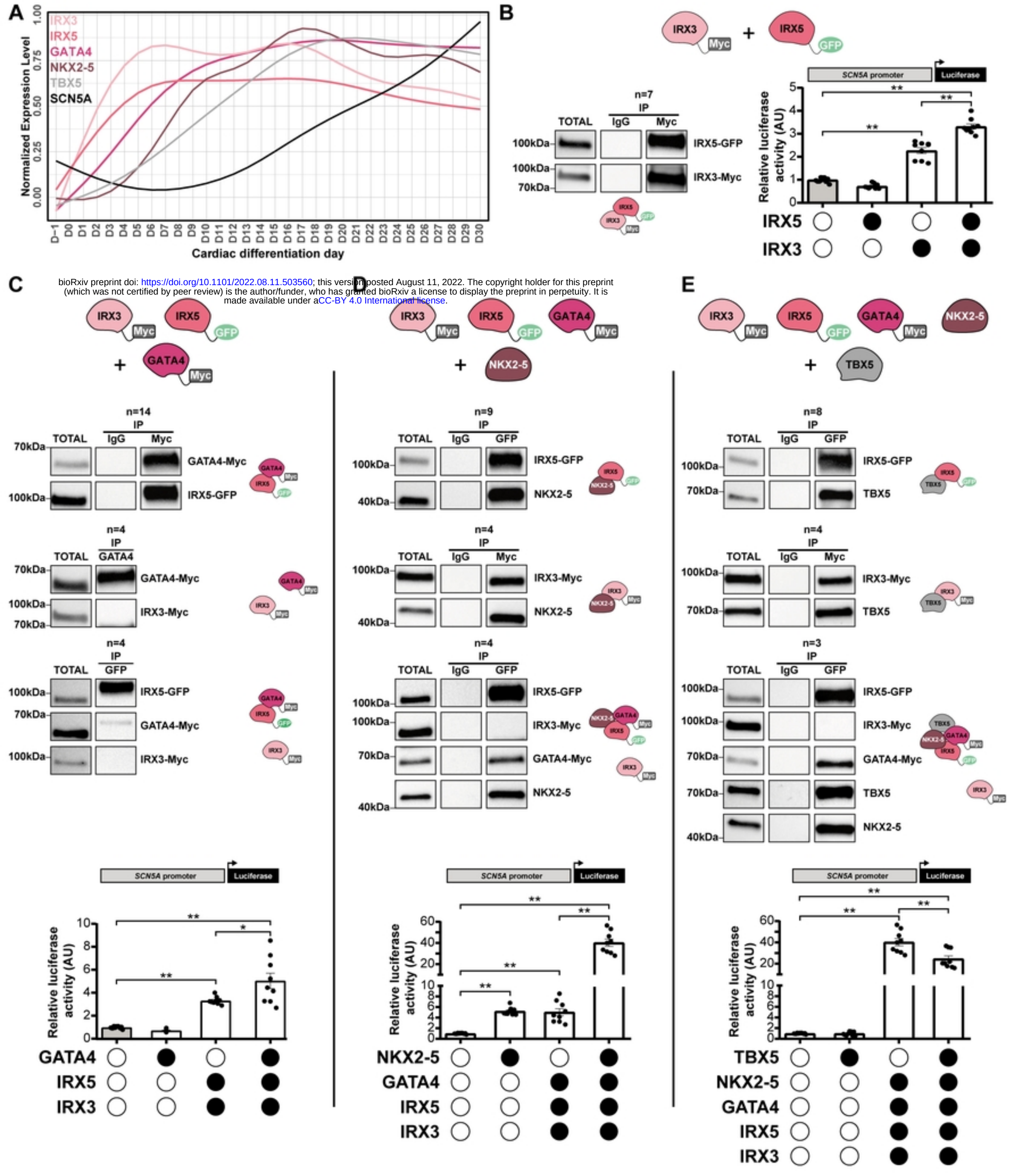


Fig 6