Machine Learning Identification of Transcriptomic Signatures of Cardiomyocyte Maturation

Syed Umer Hannan[1], Syed Abubakar Hannan[1*], Atheer Alboloshi[2*], Norah Alghamdi[3]
[1]*College of Medicine, Alrayan Colleges, Medina, Saudi Arabia*
[2]*College of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia*
[3]*Deaprtment of Bioinformatics, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*
**Contributed equally to the study*

1 INTRODUCTION

Human induced pluripotent stem cell-derived cardiomyocytes (iPSC-CMs) offer a virtually unlimited, patient-specific source of cardiac tissue for disease modeling, drug screening, and regenerative medicine.(1,2) However, iPSC-CMs generated by current differentiation protocols retain a fetal-like phenotype rather than recapitulating the structural, metabolic, and electrophysiological properties of adult cardiomyocytes.(3) Because iPSC-CMs are typically studied only weeks after differentiation, they do not undergo the years of postnatal development that endogenous cardiomyocytes require to reach maturity, and this persistent immaturity constrains their reliability for modeling late-onset or adult cardiac phenotypes and for predicting drug responses in a clinically relevant cellular context.

A wide range of strategies have been developed to push iPSC-CMs toward an adult-like state, including metabolic substrate switching to fatty-acid-based maturation media,(4,5) mechanical and electrical stimulation,(6,7) and three-dimensional engineered tissue culture, alongside prolonged culture duration and non-myocyte coculture.(8,9) Despite this progress, even combined maturation protocols fall short of the maturity achieved by native adult heart tissue,(10) and the field lacks a consensus, quantitative standard for determining where a given iPSC-CM preparation falls along the developmental continuum,(11) since existing assessments rely heavily on labor-intensive physiological assays that vary substantially between laboratories.

Among the best-characterized molecular hallmarks of cardiomyocyte maturation are coordinated isoform switches in sarcomeric genes. During normal heart development, expression of the myosin heavy chain gene MYH6 (α-MHC) gives way to MYH7 (β-MHC),(12) a transition already established in the human fetal heart by five weeks of gestation and one that newly differentiated iPSC-CMs have largely not undergone.(13) In parallel, expression of slow skeletal troponin I (TNNI1) is progressively replaced by cardiac troponin I (TNNI3) as cardiomyocytes mature; in humans this switch begins shortly after birth, and maturation is not considered largely complete until TNNI3 becomes the dominant isoform, with TNNI1 protein and transcript becoming undetectable by approximately nine months of age.(14,15) Because these switches track developmental stage independent of any single maturation protocol, MYH7/MYH6 and TNNI3/TNNI1 expression ratios are widely used as quantitative proxies for cardiomyocyte maturity.(16,17)

Several groups have proposed transcriptome-wide scoring schemes to quantify maturation more comprehensively than individual marker genes, including entropy-based pseudotime metrics benchmarked against endogenous cardiac tissue and maturation scores trained on developmental time-course transcriptomes.(18–20) These efforts underscore that comparing maturity across studies remains difficult: existing transcriptomic scores are typically built from single studies, single species, or single time-course designs rather than from an integrated, multi-cohort human dataset spanning embryonic, maturing, and adult reference states, and few explicitly account for confounding sources of variation—such as tissue composition or treatment-associated batch effects—that can masquerade as developmental signal.(19)

Here we integrate five publicly available human cardiac RNA-seq cohorts spanning embryonic, maturing, and adult developmental conditions to

characterize genome-wide gene expression changes across iPSC-CM maturation stages and to construct a cardiac maturation clock anchored to the two most established sarcomeric isoform switches, MYH7/MYH6 and TNNI3/TNNI1. Together, this work provides a quantitative, reproducible framework for benchmarking iPSC-CM maturation protocols against the human developmental trajectory and for guiding future efforts to close the persistent maturity gap.

2 MANUSCRIPT PREPARATION

2.1 Problem Formulation and Study Design

Let denote expression profiles for n samples over p genes, an ordinal maturity label, and the study each sample originated from. We seek a classifier f: ℝ^p → {1, …, C} Let X ∈ ℝ^(n×p) denote gene-expression profiles for (n) samples across (p) genes, y ∈ {1, …, C} the maturity stage, and s ∈ {1, …, M} the study of origin. Our study integrates transcriptomic analyses with predictive modelling to identify maturation-associated molecular programs and assess whether these signals generalise across independent studies. We seek a classifier f: ℝ^p → {1, …, C} whose performance on an unseen study remains comparable to that on studies represented during training. This tests whether the learned signal reflects generalisable maturation biology rather than study-specific variation.

A major challenge is that study identity is confounded with maturity: studies differ in library preparation, sequencing depth, tissue sourcing, and the maturity stages sampled. We therefore treat (s) only as an evaluation-time grouping variable and withhold it from the classifier throughout model development. Grouped cross-validation assesses cross-study generalisation, while differential expression, pathway, temporal, co-expression, and regulatory analyses provide biological interpretation of the predictive signal.

2.2 Data Sources and Sample Composition

RNA-seq from four GEO series (GSE122380, GSE117192, GSE131236, GSE141910) and GTEx was integrated into 515 samples across three conditions (Embryonic n=297, Maturing n=48, Adult n=170). GSE122380 (days 0–15) constituted Embryonic; Maturing combined GSE117192 (91.7%) and GSE131236 (8.3%); Adult combined GTEx (LV+RA) and GSE141910 (LV, non-failing donor). The ~32,600-gene matrix was ComBat-batch-corrected (R sva) prior to all downstream analysis.(21)

Study (GSE)	N samples	Condition(s)	Day / timepoint
GSE122380(22)	297	Embryonic	Days 0–15 (daily timepoints, 16–19 samples/day)
GSE141910(23)	166	Adult	N/A
GSE117192(24)	44	Maturing	Day 60
GSE85331 / GSE85332(25)	18	Embryonic (13), Maturing (5)	Days 0, 2, 4 (Embryonic); Day 30 (Maturing)
GSE239918(26)	6	Embryonic (3), Maturing (3)	Day 15 (Embryonic); Day 30 (Maturing)
GSE131236(27)	4	Maturing	Days 30, 90 (2 each)
GTEx(28)	4	Adult	N/A

Table 1. Per-study composition of the machine-learning validation dataset. Sample counts, developmental condition(s), and timepoints for the seven GEO/GTEx studies comprising the 539-sample panel used for cross-study classifier validation (Section 3), distinct from the 515-sample core analysis cohort. Study sizes range from 4 (GTEx) to 297 (GSE122380) samples, with class labels of Embryonic, Maturing, and Adult assigned per study/timepoint.

2.3 Differential Expression Analysis

DESeq2 was run with one model fit across all 515 samples.(29) Background gene universe = full DESeq2-tested set (21,924–21,927 genes, background_genes.csv). Significance: padj < 0.05; |log2FC| > 1 additionally applied for GO/GSEA input, Venn/overlap, and marker-panel analyses.

2.4 Enrichment Analysis

GO ORA via goatools (GOEnrichmentStudyNS); annotations from mygene.info, split by namespace with separate per-namespace BH correction (α=0.05); propagate_counts enabled; DAG from go-basic.obo.(30–32) Redundant terms collapsed via rrvgo (calculateSimMatrix, org.Hs.eg.db, method="Rel", scored by −log10 padj,

reduceSimMatrix); term lists >300/namespace capped to the top 300 before similarity computation.(33) Reduced sets visualized as treemaps. TF-target ORA via Enrichr.(34)

gseapy.prerank against MSigDB Hallmark 2020, ranked by DESeq2 Wald statistic; symbol-indexed, duplicates collapsed to max |statistic|.(35,36) min_size=15, max_size=500, permutation_num=1000, seed=42. Significance: FDR q < 0.25. Leading-edge genes were carried into the network and connectivity-mapping analyses below. Hallmark terms were classified "trusted" vs. confound-associated by cross-referencing against independent marker/WGCNA/regulon evidence; this classification informed downstream gene-set construction, not the GSEA statistics themselves.

gseapy.ssgsea, rank-based sample normalization was carried out on the ComBat-corrected symbol-indexed matrix (duplicate mappings collapsed by mean).(37) Each of 515 samples was scored against Glycolysis, Oxidative Phosphorylation, and Fatty Acid Metabolism.

2.5 Time-Course Trajectory Modeling

Restricted to GSE122380, the only series with a genuine per-sample day gradient (other series' nominal day_estimate values were unrelated quantities — a fixed treatment duration, a sentinel, or chronological age — and were excluded). DESeq2 paired design (~individual + day, individual IDs extracted by regex from sample characteristics); individuals with fewer than two timepoints excluded. Day effect tested via likelihood-ratio test (full vs. ~individual); the day coefficient is per-day log2FC.

2.6 Co-Expression Network Analysis (WGCNA)

Signed network on the top 5,000 most-variable genes, standard QC (goodSamplesGenes). Soft-thresholding power scanned 1–30 targeting scale-free $R^2 > 0.85$; not achieved for this heterogeneous dataset (max $R^2 = 0.809$ at power = 30), so power = 30 was used and modules interpreted with that limitation noted. blockwiseModules (signed, minModuleSize=30, mergeCutHeight=0.25). Eigengenes correlated (Pearson, corPvalueStudent) against differentiation day, condition, and composite z-scored marker-panel scores (immune, neuroectodermal, fibroblast, endothelial).(38)

2.7 Transcription Factor / Regulator Activity

ChEA 2022 over-representation analysis (ORA) was applied to the trusted-pathway leading-edge gene lists to identify enriched transcription-factor target sets. ChEA3 provides curated TF–target gene-set libraries for transcription-factor enrichment analysis.(39) DoRothEA transcription-factor regulons were retrieved through the OmniPath REST API and restricted to high-confidence interactions (confidence levels A–C).(40–42) Regulatory mode was assigned using consensus stimulation/inhibition annotations; interactions lacking an unambiguous regulatory sign were excluded, and duplicate TF–target interactions were collapsed.

For each TF with ≥5 scoreable target genes, TF activity was estimated using a univariate linear model (ULM), regressing the DESeq2 Wald statistic against the signed TF–target interaction weights (stimulation = +1; inhibition = −1). The activity score was defined as the regression slope divided by its standard error (i.e., the slope t-statistic), consistent with the ULM implementation used by decoupler/decoupleR. This procedure was independently reimplemented using scipy.stats.linregress to avoid a decoupler/AnnData–NumPy dependency conflict.

2.8 Protein-Protein Interaction Network

STRING API (physical, confidence ≥400), batches of 300 symbols; networkx degree centrality (hubs) and betweenness centrality (bridge genes).(43) Two input gene sets, the per-contrast trusted-category leading-edge union, and the cross-contrast confound-excluded core set, were used for separate analyses.

2.9 Cell-Type Composition and Confound Markers

Marker panels: endothelial (PECAM1), immune (PTPRC/CD45), fibroblast (COL1A1, DCN), neuroectodermal (TUBB3, PAX6, NES), selected from established single-cell/bulk literature.(44–46) Composite scores (used in WGCNA trait correlation) = row-mean of column z-scored expression per panel.

2.10 Core Maturation Gene Set Definitions

Broad set (Venn/overlap use): genes meeting padj < 0.05 and |log2FC| > 1 in all three pairwise contrasts simultaneously, consistent with standard differential-expression filtering.(47) Narrow set (PPI/connectivity use): the intersection, across all three contrasts, of leading-edge genes from trusted-category GSEA terms, representing the core genes driving enrichment rather than the full gene sets.(48) These sets are not interchangeable, as the former is defined by gene-level differential-expression

thresholds, whereas the latter is defined by pathway-level enrichment and leading-edge membership.

2.11 Software Environment

R (DESeq2, WGCNA, rrvgo, org.Hs.eg.db, GOSemSim; Bioconductor) for DE, co-expression, and GO simplification. Python (goatools, gseapy, mygene, networkx, scipy, pandas; Google Colab) for enrichment, GSEA, single-sample scoring, regulon scoring, network analysis, and connectivity mapping. Fixed DESeq2-tested background gene universe throughout; unmapped identifiers logged, not excluded.

3 EXPERIMENT DESIGN

3.1 Data Compilation for Validation

For the machine learning validation component, we compiled 539 bulk RNA-seq profiles from seven independent studies, intersected to a common panel of 5,000 genes. Each sample carries a three-level maturity label: Embryonic, Maturing, Adult, with class counts of 313, 56, 170 respectively. Study sizes range from 28 to 463 samples. Per-study sample counts and class composition are given in Table 1.

3.2 Preprocessing for Validation

To mitigate inter-study batch effects and sequencing-depth variation, raw counts were normalised row-wise to counts per million and log1p-transformed. For gene i in sample j with raw count x_ij and G total genes:

$CPM_{ij} = (x_{ij} / \Sigma_{k=1}^{G} x_{kj}) \times 10^6$

$y_{ij} = \log(1 + CPM_{ij})$

Features were then z-scored to satisfy the scale assumptions of linear and margin-based classifiers. Standardisation statistics were estimated on the training split of each fold and applied unchanged to the held-out split, so no held-out sample contributes to its own normalisation.

3.3 Study-Grouped Evaluation

We used stratified group k-fold cross-validation with k = 3, grouped by study. Study sizes are typically unbalanced, so a per-fold score is governed by whichever large study happens to be held out. We instead collect out-of-fold predictions for every sample, so that each sample is predicted exactly once by a model that never saw its study, and compute a single metric over the pooled prediction vector. This weights every sample equally and yields one estimate with a well-defined denominator.

3.4 Models

Four classifiers were trained on the full 5,000 standardised features. All linear models use balanced class weights to compensate for the minority Maturing class.

3.5 Feature Selectors

The primary selector was Lasso, chosen for its sparsity property. We trained it on the data using grouped stratified k-fold cross-validation and retained all genes with non-zero coefficients. To assess Lasso's credibility as a selector, we employed a nested validation scheme in which a Ridge model was trained on the Lasso-selected genes from each fold and scored on the corresponding held-out split, rather than evaluating on the full matrix, which would yield a falsely optimistic score. The other selector was principal component analysis (PCA), an unsupervised approach that simplifies large datasets by reducing the number of variables while preserving essential information. We contrasted the Lasso-retained genes with those having the highest loadings on the first principal component.

3.6 Evaluation Metric

We report macro-averaged F1 over the pooled out-of-fold prediction vector.

3.7 Evaluation Protocols (E1–E4)

E1, preliminary transferability screen: Before committing to an evaluation protocol we ran a leave-one-study-out screen, training on six studies and testing on the seventh, to establish whether any cross-study signal exists. This screen used a separate and broader model set (CatBoost, MLP, logistic regression, Random Forest) and is reported as a diagnostic only. Its per-study scores are not comparable to the pooled estimates in E2 to E4.

E2, full-feature comparison: The four classifiers of Section 3.4 are evaluated on all 5,000 features under the grouped protocol.

E3, feature selection: Two variants are run. In the non-nested variant, a Lasso is fitted within each fold and genes with non-zero mean absolute coefficient across folds are collected into a single set, on which an $\ell 2$ model is refit and re-evaluated under the same grouped protocol. This variant is diagnostic: the pooled set is informed by all folds, so its score is

optimistic and is reported only as an upper reference. In the nested variant, which is the reported result, the Lasso selects the top 100 genes by absolute coefficient from each training split and a Ridge evaluator is trained on those genes alone and scored on the held-out split. This nested procedure acts as a final check for the selected features, confirming that the observed performance is not an artifact of selecting on the evaluation data.

E4, selector agreement: We compute the core set and fold-specific remainder for each selector across the three folds, then the pairwise intersections between Lasso and PC1 loading at matched set size.

4 RESULTS

This study develops and applies a framework for separating biologically coherent cardiomyocyte maturation signal from tissue-composition and cross-study technical artifacts in heterogeneous public transcriptomic data. Because the public datasets spanning this developmental range differ systematically in cell-type composition, sample source, and experimental history, any pairwise comparison between them risks conflating true maturation biology with these confounds. The methodological contribution of the analysis below is a set of pre-specified, cross-validating criteria for distinguishing the two; these criteria are then used to recover a confound-resistant maturation signal and to generate a testable candidate intervention.

4.1 Canonical maturation markers validate the pipeline across the developmental trajectory

To confirm that the analytical pipeline captures genuine cardiomyocyte maturation biology prior to genome-wide analysis, eleven established maturation marker genes were examined across all three pairwise developmental transitions (Embryonic, Maturing, Adult; n = 297, 48, and 170 samples, respectively) (Figure 1, which reports the direction and magnitude of change for all eleven markers). Six markers exhibited a clear, monotonic increase toward the Adult condition — RYR2, CASQ2, MYL2, PLN, MYH7, and TNNI3 — consistent with progressive maturation of excitation-contraction coupling and sarcomeric composition.(10) The remaining markers are addressed individually below: the two isoform-switch ratios anchoring the maturation clock (MYH7/MYH6, TNNI3/TNNI1) showed more complex, contrast-dependent behavior, and two markers (ATP2A2, GJA1) moved in the direction opposite to expectation in every comparison.

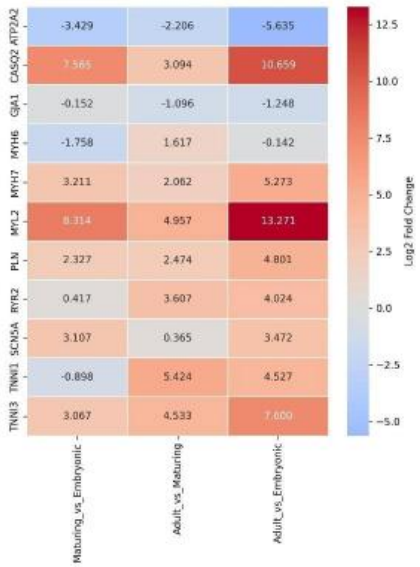


Figure 1. Expression changes of eleven canonical cardiomyocyte maturation markers across the three pairwise developmental contrasts. Heatmap of log2 fold-change for RYR2, CASQ2, MYL2, PLN, MYH7, MYH6, TNNI3, TNNI1, SCN5A, GJA1, and ATP2A2 in Maturing-vs-Embryonic, Adult-vs-Maturing, and Adult-vs-Embryonic contrasts. Six markers (RYR2, CASQ2, MYL2, PLN, MYH7, TNNI3) increase monotonically toward the Adult condition; ATP2A2 and GJA1 move in the direction opposite to expectation in every contrast.(10)

The two isoform-switch ratios underlying the maturation clock, MYH7/MYH6 and TNNI3/TNNI1, exhibited contrast-dependent behavior that was resolved upon further examination rather than indicating pipeline failure. In the Adult-versus-Maturing contrast, both MYH6 and MYH7 increased concurrently (log2 fold-change of +1.62 and +2.06, respectively), yielding a modest net ratio shift (Δlog2 = +0.445). A parallel likelihood-ratio-test model applied to the dense GSE122380 differentiation time course (days 0–15, individual-paired design) revealed the same co-induction pattern for MYH6 and MYH7, with near-identical per-day slopes (+0.781 and +0.783 log2 fold-change per day, respectively). This indicates that the postnatal MYH6-to-MYH7 isoform switch constitutes a distinct, later biological event, separable from the pan-sarcomeric gene induction captured within this 15-day differentiation window. The TNNI3/TNNI1 ratio exhibited a more pronounced discrepancy, moving in the direction opposite to expectation in the Adult-versus-Maturing

pairwise contrast (Δlog2 = −0.891) while moving in the expected direction within the confound-free GSE122380 trajectory (+0.05 log2 fold-change per day; TNNI3 padj = 1.8×10^{-40}). This divergence between a single-cohort time course and a between-study pairwise contrast constituted the initial indication of a compositional confound affecting the latter, addressed directly in Section 2.

Two markers, ATP2A2 and GJA1, exhibited expression changes opposite to the expected maturation direction in every comparison, including the confound-free GSE122380 time course (ATP2A2: −0.492 log2 fold-change per day, padj = 2.5×10^{-314}). This finding is reported as an open biological question rather than a pipeline artifact, given that it replicates in a design that is not subject to the cross-study confounds described below.

4.2 A pronounced tissue-composition confound dominates the Adult-versus-Maturing pairwise contrast

The Adult-versus-Maturing contrast, intended to isolate the terminal stage of cardiomyocyte maturation, instead exhibited an unusually large proportion of the transcriptome as differentially expressed: 85.5% of tested genes reached statistical significance (padj < 0.05), of which 62% exhibited |log2 fold-change| ≥ 1, indicating substantial effect sizes rather than a purely sample-size-driven statistical artifact.

Examination of cell-type composition marker genes identified the underlying cause (Figure 2). Endothelial (PECAM1: log2FC = +7.81, padj = 1.1×10^{-231}) and immune (PTPRC/CD45: log2FC = +2.51, padj = 1.9×10^{-33}) markers were markedly enriched in the Adult condition, consistent with Adult samples originating from whole heart tissue (GTEx, GSE141910) rather than purified cardiomyocyte culture. Conversely, fibroblast/mesenchymal markers (DCN: log2FC = +4.34; COL1A1: log2FC = −5.46) and, unexpectedly, neuroectodermal lineage markers (TUBB3: log2FC = −6.62, padj = 1.5×10^{-267}; PAX6; NES) were elevated in the Maturing condition, indicating off-target fibroblast and neuroectodermal differentiation within the constituent cultured iPSC-CM samples, a recognized limitation of standard monolayer cardiac differentiation protocols.(49,50)

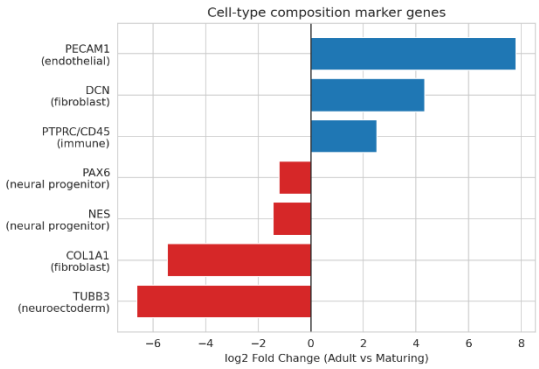


Figure 2. Cell-type composition marker genes reveal a tissue-source confound in the Adult-vs-Maturing contrast. Bar chart of log2 fold-change (Adult vs. Maturing) for endothelial (PECAM1), fibroblast (DCN, COL1A1), immune (PTPRC/CD45), and neural progenitor (PAX6, NES, TUBB3) marker genes. Endothelial and immune markers are enriched in Adult (whole-heart) tissue, while fibroblast and neuroectodermal markers are elevated in Maturing (cultured iPSC-CM) samples, indicating off-target differentiation and tissue-source mismatch rather than developmental signal.

Compositional analysis of the Maturing condition (n = 48) traced this finding to study-level structure: 91.7% of Maturing-condition samples derive from GSE117192, a hypoxia-treatment study in which only 25% of samples (11 of 44) represent true normoxic baseline, with the remainder reflecting hypoxia or re-oxygenation treatment histories. Gene Ontology enrichment analysis of the Adult-versus-Maturing gene lists (Figure 3) provided further support for a compositional, rather than purely developmental, interpretation: identical generic terms (mitochondrion, immune response, ribosomal subunit) were significantly enriched in both the up- and down-regulated gene sets, a pattern characteristic of oversized, heterogeneous gene lists rather than specific, directional biology.



Figure 3. Gene Ontology enrichment of the Adult-vs-

Maturing differentially expressed gene lists shows a non-specific, compositional pattern. Bubble plots of top enriched GO terms (by −log10 FDR, sized by gene count, colored by fold enrichment) among genes up-regulated (A) and down-regulated (B) in Adult vs. Maturing. Identical generic terms (mitochondrion, immune response, ribosomal subunit) appear in both directions, consistent with an oversized, heterogeneous gene list rather than directional maturation biology.

Analytical framework. To guard against a post-hoc interpretation of the compositional-confound finding above, corroboration was required, a priori, from multiple complementary analytical approaches before the confound explanation would be accepted and used to define a confound-excluded gene set for downstream analysis. These approaches are not independent experiments — they are different transformations of the same underlying expression matrix — but each interrogates a distinct structural property of the data (network topology, transcription-factor activity, pathway enrichment timing) that a compositional confound would be expected to leave a distinct signature on, so agreement across them is not guaranteed by construction. Specifically, the confound was treated as established only if it was corroborated by at least two of the following: co-expression network structure, transcription-factor regulon activity, and pathway-timing analysis, in addition to the marker-gene and Gene Ontology evidence already presented. The four analyses below were designed and run to test this pre-specified criterion, not to search post hoc for support of a conclusion already reached.

4.3 The confound is corroborated by convergent evidence from co-expression structure, regulator activity, and pathway timing

Four additional, complementary analyses of the same underlying data converged on the same conclusion, providing corroborating evidence beyond the marker-gene and Gene Ontology findings above.

Co-expression network structure (WGCNA). Among nine gene co-expression modules derived from the top 5,000 most variable genes across all 515 samples, no module tracked developmental day or condition independently of a cell-composition confound score (Figure 4). The module most strongly correlated with differentiation day (MEblue: r = 0.81) was similarly strongly correlated with endothelial composite marker scores (r = 0.87) and substantially, though somewhat less strongly, correlated with immune composite marker scores (r = 0.66). The module most strongly correlated with the Embryonic condition (MEturquoise: r = 0.55) exhibited comparable correlation with neural markers (r = 0.65). A third module (MEbrown) correlated specifically with the Maturing condition (r = 0.63) and neural markers (r = 0.72), independent of day, Adult status, or immune markers; Gene Ontology enrichment of this 430-gene module returned extracellular matrix and focal adhesion terms as the dominant signal (extracellular matrix organization, padj = 2.4×10^{-24}; focal adhesion, padj = 1.4×10^{-24}), identifying this module as fibroblast/mesenchymal rather than neuronal in character and refining the marker-based hypothesis. The module's concurrent correlation with neural marker scores is interpreted as reflecting co-elevation of fibroblast and neuroectodermal off-target lineages within the same constituent iPSC-CM samples, rather than indicating a neuronal component of the module itself, consistent with the dual fibroblast/neuroectodermal off-target differentiation pattern identified from marker-gene analysis above.

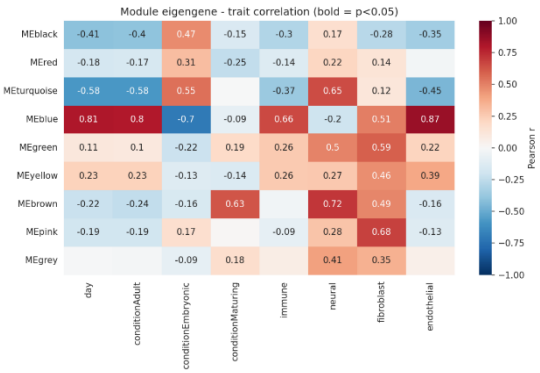


Figure 4. WGCNA module–trait correlations link co-expression structure to cell-composition confounds rather than developmental progression. Heatmap of Pearson correlations (bold = p<0.05) between nine module eigengenes and differentiation day, condition (Adult/Embryonic/Maturing), and composite immune, neural, fibroblast, and endothelial marker scores, across all 515 samples. The day-correlated module (MEblue) tracks endothelial/immune scores; the Embryonic-correlated module (MEturquoise) tracks neural markers; MEbrown correlates with the Maturing condition and neural markers independently of day.

Transcription factor regulon activity. Regulon activity scoring using DoRothEA (338 transcription factors, confidence levels A–C) identified RELA (activity score = 3.06, p = 0.0026), NFKB1 (2.44), RELB (2.51), and RFX5 (3.39; a regulator of MHC class II expression) as significantly activated in the Adult condition relative to Maturing, consistent with

a canonical NF-κB/immune-signaling regulatory program.(51,52) ERG, the master transcriptional regulator of endothelial identity, showed a directionally consistent, marginally significant increase in activity (2.16, p = 0.056; nominal, not reaching the conventional 0.05 threshold), lending additional, if weaker, support to the PECAM1 finding at the regulatory level (Figure 5). Reported DoRothEA p-values are nominal and are not corrected for the 338 transcription factors tested; the regulator-level findings in this study are interpreted as corroborating evidence alongside the marker-gene, Gene Ontology, and pathway-timing results rather than as independently significant discoveries.

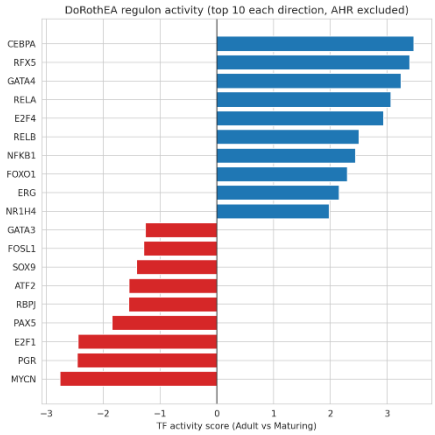


Figure 5. DoRothEA transcription-factor regulon activity distinguishes confound-associated and maturation-associated regulators (Adult vs. Maturing). Bar chart of the top 10 activated and top 10 repressed regulons (AHR excluded) by ULM activity score. Confound-linked immune/endothelial regulators (RELA, RFX5, NFKB1, RELB, ERG) are activated in Adult relative to Maturing.

Pathway timing across the developmental trajectory. Hallmark gene set enrichment analysis (GSEA) performed across all three pairwise contrasts (Figure 6) demonstrated that the Hypoxia and Epithelial-Mesenchymal Transition gene sets were selectively enriched in comparisons involving the Maturing condition (Hypoxia normalized enrichment score [NES] = 2.7 versus Embryonic, −1.5 versus Adult; Epithelial-Mesenchymal Transition NES = 3.1 versus Embryonic, the strongest NES observed among the 41 significant Hallmark terms identified in the Maturing-versus-Embryonic contrast specifically), rather than exhibiting the monotonic, direction-consistent pattern expected of genuine developmental progression. This isolated, condition-specific enrichment pattern is inconsistent with a developmental signal and instead provides direct evidence that the Maturing condition is compositionally distinct from a simple biological midpoint between Embryonic and Adult states.

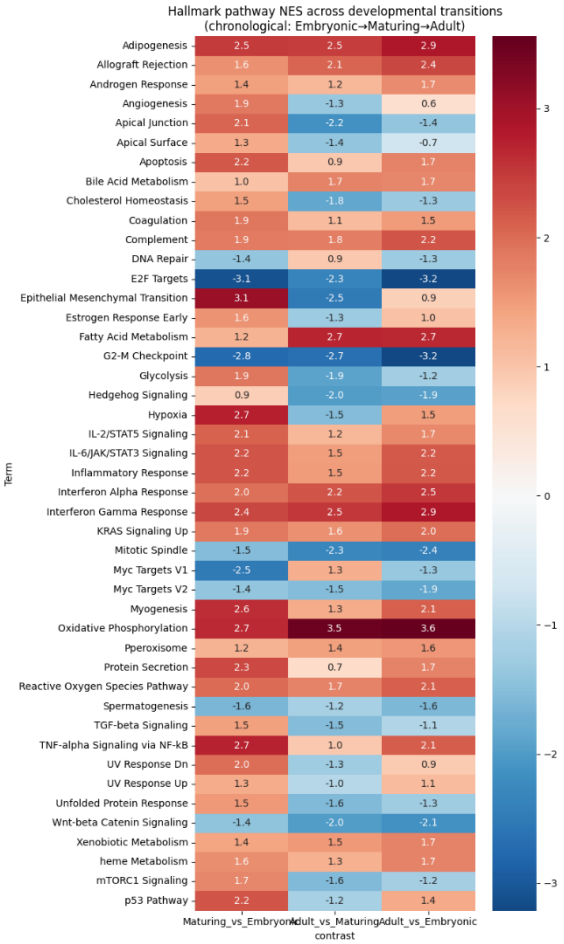


Figure 6. Hallmark pathway enrichment across the three developmental transitions. Heatmap of normalized enrichment scores (NES) for all significant MSigDB Hallmark gene sets across Maturing-vs-Embryonic, Adult-vs-Maturing, and Adult-vs-Embryonic contrasts, ordered chronologically. Hypoxia and Epithelial-Mesenchymal Transition are selectively enriched only in contrasts involving the Maturing condition, rather than showing the monotonic pattern expected of true developmental progression.

Extended regulator activity, all three contrasts. Extension of regulon activity scoring to all three pairwise contrasts demonstrated that confound-associated regulators (RELA, NFKB1, RELB, ERG) exhibited activity scores near zero in the Maturing-versus-Embryonic comparison (RELA = 0.16, ERG = −0.37) and became strongly activated only in comparisons involving the Adult condition, mirroring

the pathway-timing pattern described above at the level of individual named regulators.

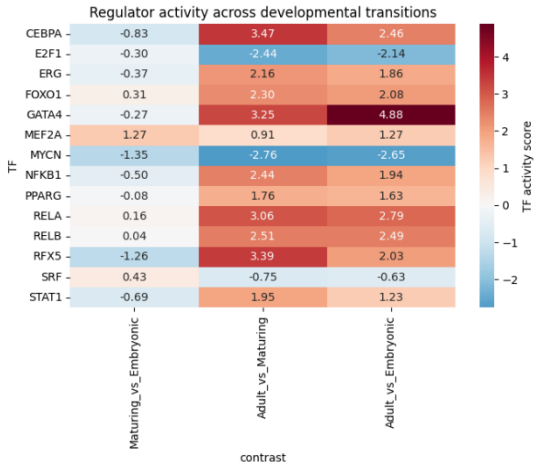


Figure 7. Confound-associated transcription factor regulon activity is restricted to comparisons involving the Adult condition. Heatmap of TF activity scores for 14 selected regulators (including RELA, NFKB1, RELB, ERG, RFX5) across all three pairwise contrasts, showing near-zero activity in Maturing-vs-Embryonic and strong activation only when the Adult condition is involved.

4.4 A biologically coherent maturation signal is separable from the confound

Despite the compositional confound described above, a maturation signal that is separable from that confound and internally consistent was recoverable using analytical approaches that isolate signal by pathway membership rather than raw differential expression.

Structural and cell-cycle signal (Gene Ontology). Simplification of Gene Ontology terms (rrvgo, semantic similarity threshold of 0.7) applied to the Adult-versus-Maturing gene lists resolved 1,144 (up-regulated) and 1,022 (down-regulated) significant terms into a small number of interpretable thematic clusters. Sarcomere and contractile-apparatus terms (sarcomere, M band, I band, A band, Z disc, contractile fiber) formed a coherent, exclusively up-regulated cluster. Cell-cycle terms (mitotic cell cycle, spindle checkpoint signaling) formed a coherent, exclusively down-regulated cluster, consistent with the established post-mitotic character of mature cardiomyocytes.(53)

Pathway-level confirmation (GSEA). Full-transcriptome GSEA against the MSigDB Hallmark collection (Figure 8) provided converging evidence for both signals: Oxidative Phosphorylation (NES = +3.47, FDR < 0.001), Fatty Acid Metabolism (+2.69), and Adipogenesis (+2.46) constituted the strongest positive enrichments; G2-M Checkpoint (−2.67), E2F Targets (−2.35), and Mitotic Spindle (−2.30) constituted the strongest negative enrichments.

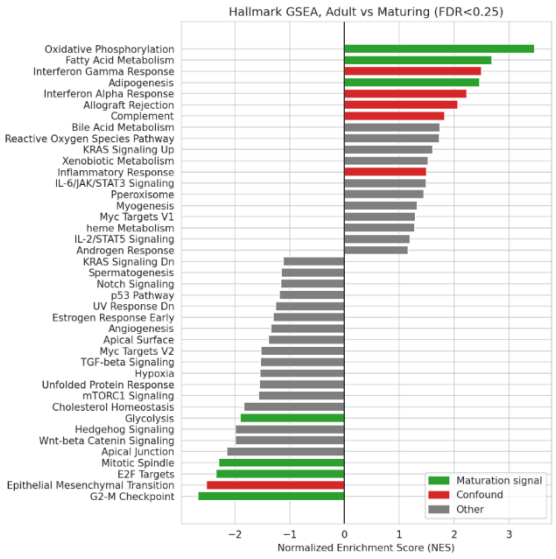


Figure 8. Hallmark GSEA of the Adult-vs-Maturing contrast separates maturation signal from confound-associated pathways. Bar chart of NES for all Hallmark terms reaching FDR<0.25, color-coded as maturation signal (green; e.g., Oxidative Phosphorylation, Fatty Acid Metabolism, Adipogenesis, G2-M Checkpoint, E2F Targets, Mitotic Spindle), confound (red; e.g., Hypoxia, Epithelial-Mesenchymal Transition, immune/interferon terms), or other (gray).

Regulator-level confirmation. E2F1 regulon activity was significantly repressed in the Adult condition relative to Maturing (activity score = −2.44, p = 0.015, n = 322 target genes), with MYCN similarly repressed (−2.76, p = 0.013), providing regulator-level evidence convergent with the cell-cycle-exit signal identified by Gene Ontology and GSEA analyses above. Activation of CEBPA (3.47, p = 0.0007) and PPARG (1.76) supported the metabolic maturation signal at the regulatory level.

Network-level confirmation. A protein-protein interaction network was constructed from a confound-excluded core gene set, defined as the intersection, across all three pairwise developmental contrasts, of leading-edge genes from trusted-category GSEA terms — defined as Hallmark terms not among those independently identified as confound-associated in Section 3 (Hypoxia, Epithelial-Mesenchymal Transition) — comprising Oxidative Phosphorylation, Fatty Acid Metabolism, Adipogenesis, Glycolysis, G2-M Checkpoint, E2F Targets, and Mitotic Spindle. Trusted-category

leading-edge gene sets were recovered in all seven terms for each contrast (Maturing-versus-Embryonic: 512 genes; Adult-versus-Maturing: 512 genes; Adult-versus-Embryonic: 541 genes), yielding 248 genes common to all three contrasts — this confound-excluded set, rather than the larger, less specific 3,562-gene set introduced for completeness in Section 6, constitutes the headline confound-resistant maturation gene set reported in this study. Of these, 191 genes mapped to the STRING physical interaction database at the applied confidence threshold (≥400) and formed a network of 946 edges (Figure 9). This network resolved into two visually and topologically distinct clusters: a mitochondrial Complex I/III structural clique, with the highest-degree hub genes corresponding to core respiratory chain subunits (UQCRC2, degree = 36; CYC1, degree = 34; UQCRFS1, COX4I1, NDUFB8, degree = 33; additional Complex I/III subunits including COX5B, COX6B1, UQCRQ, NDUFS2, NDUFS3, NDUFAB1, UQCRC1, UQCR11, COX7C, and COX7B each with degree ≥ 31), and a cell-cycle clique. Betweenness centrality, which identifies genes bridging distinct network communities rather than genes embedded within a single dense cluster, ranked TP53 (betweenness = 0.227), CDCA8 (0.155), VDAC1 (0.155), and CDK1 (0.149, degree = 28) as the principal bridge nodes connecting the metabolic and cell-cycle clusters, with additional bridging contribution from ATP5F1B, CYCS, STAT5A, AURKB, BRCA2, PLK1, H2AX, INCENP, and ATP5F1C. The identification of TP53 and CDK1 as high-betweenness bridge nodes, rather than high-degree hub nodes, is consistent with established literature describing mechanistic coupling between cardiomyocyte cell-cycle exit and metabolic maturation via p53/E2F-mitochondrial biogenesis crosstalk, and indicates that these genes are more plausible candidate coordinating regulators than the structurally obligate, high-degree respiratory-complex subunits, which are densely interconnected by definition and are not expected to exhibit independent regulatory function.(54) Restriction of network construction to this confound-excluded gene set was necessary to recover this cell-cycle/metabolism-specific structure: a parallel network constructed from the broader, differential-expression-significance-based core gene set (Section 6; 3,562 genes) instead returned a fibroblast/adhesion-dominated network, with distinct hub genes (ESR1, FN1, CDH1, CD44), underscoring that confound exclusion at the level of pathway membership, rather than statistical significance alone, was required for an interpretable network analysis.

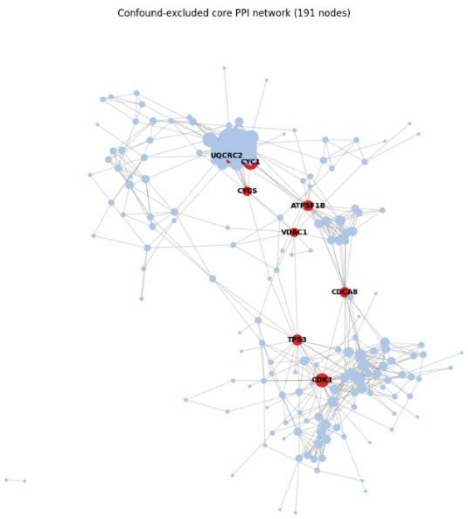


Figure 9. Confound-excluded core protein–protein interaction network. STRING physical-interaction network (confidence ≥400) built from the 191 mapped genes of the 248-gene confound-excluded core set, forming 946 edges. The network resolves into a mitochondrial Complex I/III structural clique (high-degree hubs, e.g., UQCRC2, CYC1) and a cell-cycle clique, bridged by high-betweenness nodes TP53, CDCA8, VDAC1, and CDK1 (highlighted).

4.5 The glycolysis-to-oxidative-phosphorylation metabolic switch exhibits two temporally distinct components

Single-sample GSEA (ssGSEA) scoring of Glycolysis, Oxidative Phosphorylation, and Fatty Acid Metabolism gene sets across all 515 samples (Figure 10) yielded a composite Metabolic Switch Index (Glycolysis score minus Oxidative Phosphorylation score) that decreased monotonically across the three conditions (Embryonic: −1,872; Maturing: −2,078; Adult: −5,222), consistent with progressive metabolic maturation. The magnitude of this change was concentrated almost entirely within the Maturing-to-Adult transition; Oxidative Phosphorylation and Fatty Acid Metabolism scores changed only marginally between the Embryonic and Maturing conditions (+508 and +54, respectively) prior to substantial increases toward the Adult condition (+2,037 and +2,122, respectively).

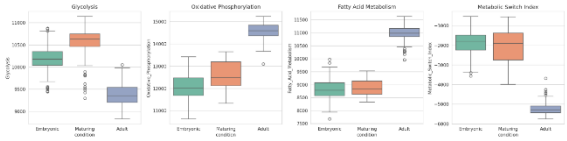


Figure 10. Single-sample GSEA metabolic pathway scores across the developmental trajectory. Boxplots of ssGSEA scores for Glycolysis, Oxidative Phosphorylation, and Fatty Acid Metabolism, plus the composite Metabolic Switch Index (Glycolysis − Oxidative Phosphorylation), across Embryonic, Maturing, and Adult samples (n=515). The Metabolic Switch Index decreases monotonically, with the largest shift occurring between Maturing and Adult.

This raw-score pattern initially suggested that metabolic maturation constitutes a comparatively late event. However, full-transcriptome GSEA comparing all three pairwise contrasts (Figure 11) demonstrated that Oxidative Phosphorylation was already strongly and significantly enriched in the earliest transition (Maturing-versus-Embryonic NES = 2.7, comparable to the Adult-versus-Maturing NES of 3.5), indicating that a statistically robust, if smaller-magnitude, shift toward oxidative metabolism is detectable from the earliest stage of differentiation. The raw-score and NES-based findings are interpreted as complementary rather than contradictory: ssGSEA raw scores reflect absolute per-sample effect magnitude, whereas GSEA NES reflects the statistical concentration of coordinated gene-level shifts across a ranked list, which may achieve strong statistical significance even when the underlying magnitude of change is modest.

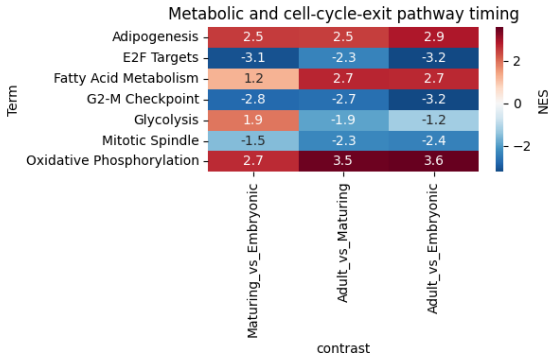


Figure 11. Pathway-level timing of metabolic and cell-cycle-exit signatures across developmental transitions. Heatmap of NES for Adipogenesis, E2F Targets, Fatty Acid Metabolism, G2-M Checkpoint, Glycolysis, Mitotic Spindle, and Oxidative Phosphorylation across all three pairwise contrasts, showing that Oxidative Phosphorylation is already strongly enriched in the earliest (Maturing-vs-Embryonic) transition.

The Glycolysis score exhibited a non-monotonic pattern across the developmental trajectory (NES = +1.9 for Maturing versus Embryonic; NES = −1.9 for Adult versus Maturing), consistent with two biologically distinct, sequential processes: an initial decrease in glycolytic dependence upon exit from pluripotency, a phenomenon independently documented in the pluripotency literature (pluripotent stem cells exhibit constitutively elevated aerobic glycolysis independent of lineage-specific maturation), followed by the well-characterized glycolysis-to-oxidative-phosphorylation switch specifically associated with postnatal cardiomyocyte maturation.(55,56)

4.6 A broader, differential-expression-based candidate gene set spans all three developmental transitions

This candidate gene set is presented for completeness and should be interpreted with caution given the confound described in Section 3. To identify genes exhibiting consistent and substantial expression change across the full developmental trajectory, differentially expressed genes (padj < 0.05, |log2 fold-change| > 1) were intersected across all three pairwise contrasts (Figure 12). Effect-size filtering was applied uniformly given the large sample sizes involved (up to n = 297 per group), which in its absence yielded an uninterpretably large proportion of the transcriptome as nominally significant (61.5% at padj < 0.05 alone). Following filtering, 7,761 genes were significant in the Maturing-versus-Embryonic contrast, 12,738 in the Adult-versus-Embryonic contrast, and 11,739 in the Adult-versus-Maturing contrast; 3,562 genes (16.2% of the tested transcriptome) were significant and directionally consistent across all three comparisons. This set is presented as a broad, statistically reproducible candidate maturation gene list. It is noted explicitly that this set was constructed on the basis of differential-expression significance alone and, consistent with the network analysis in Section 4, is expected to retain a proportion of confound-associated genes rather than representing a confound-free set; the stricter, confound-excluded gene set (248 trusted-pathway genes, 191 of which mapped to the STRING network) described in Section 4 is recommended for analyses requiring mechanistic specificity. Accordingly, the 3,562-gene set below should be treated as a broad, lower-specificity candidate pool provided for completeness rather than as this study's primary result (see Section 4 for the confound-excluded 248-gene set).

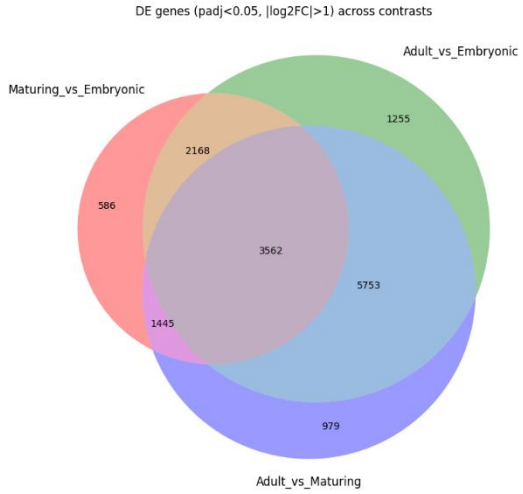


Figure 12. Overlap of differentially expressed genes across the three pairwise developmental contrasts. Venn diagram of genes meeting padj<0.05 and |log2FC|>1 in Maturing-vs-Embryonic (7,761 genes), Adult-vs-Embryonic (12,738 genes), and Adult-vs-Maturing (11,739 genes) contrasts, yielding a 3,562-gene broad core set significant and directionally consistent across all three comparisons.

4.7 Connectivity mapping generates a coherent, hypothesis-level candidate: MEK/ERK pathway inhibition

To generate translationally relevant hypotheses, the LINCS L1000 chemical perturbation database was queried (via Enrichr) using directional gene signatures. An initial query using the undirected union of trusted-pathway leading-edge genes returned a diffuse and mechanistically incoherent hit list dominated by cytotoxic and broadly antiproliferative compounds, including mitoxantrone, a DNA-damaging chemotherapeutic agent with no established maturation-specific mechanism of action. This outcome is attributable to the known tendency of generic cell-cycle-arrest transcriptional signatures to produce non-specific results in connectivity-mapping analyses when input gene lists are not stratified by direction of effect.

A directional query strategy was subsequently applied: genes increasing with maturation (leading-edge genes from Oxidative Phosphorylation, Fatty Acid Metabolism, and Adipogenesis; n = 267) were queried against LINCS up-regulation signatures, and genes decreasing with maturation (leading-edge genes from G2-M Checkpoint, E2F Targets, and Mitotic Spindle; n = 174) were queried separately against LINCS down-regulation signatures. This yielded a substantially more coherent and mechanistically interpretable result (Figure 13). Of 208 compound perturbations reaching significance in both directions independently, the fifteen highest-ranked hits were dominated by a single coherent pharmacological class: inhibitors of the RAS/RAF/MEK/ERK signaling cascade. Five distinct MEK1/2 inhibitors were independently recovered among the top hits — PD-0325901 (combined score ≈ 79, the highest-ranked compound overall, recovered independently across multiple cell lines and doses including MCF10A and HCC515 at 3.33 μM), trametinib (recovered across three independent dose points in the HCC515 cell line: 1.11, 3.33, and 10 μM), AZD-8330, PD-184352, and selumetinib — together with the upstream BRAF inhibitor SB590885 and multiple receptor tyrosine kinase inhibitors converging on the same cascade (afatinib, dasatinib, NVP-TAE684, foretinib, vemurafenib). A secondary, related cluster of PI3K/mTOR pathway inhibitors was also recovered (NVP-BEZ235, GSK-2126458, GDC-0980, PIK-93), consistent with parallel growth-factor-driven signaling contributing to the same phenotype. The recovery of five mechanistically related MEK1/2 inhibitors, rather than a single compound, together with pathway-upstream and pathway-parallel hits, indicates that this result reflects a coherent pathway-level signal rather than an artifact of any single compound-cell-line combination. This finding is consistent with the established role of MAPK/ERK signaling as a driver of cardiomyocyte proliferation and a barrier to maturation,(57) and identifies a transcriptional perturbation in LINCS that resembles aspects of the inferred maturation signature, yielding MEK/ERK pathway inhibition as a coherent, hypothesis-generating candidate for promoting induced pluripotent stem cell-derived cardiomyocyte maturation. This candidate is pending experimental validation and is not presented as a defensible therapeutic strategy in its own right; it directly addresses the project's original translational objective at the level of hypothesis generation only.

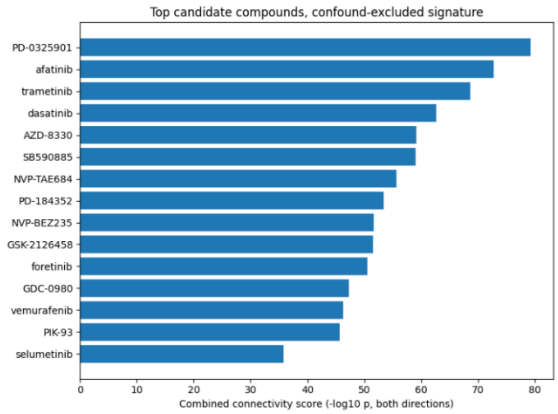


Figure 13. Top candidate compounds from direction-stratified LINCS L1000 connectivity mapping. Bar chart of combined connectivity scores (−log10 p, both directions) for the fifteen top-ranked compound perturbations queried against the confound-excluded, direction-stratified maturation gene signature, dominated by MEK1/2 inhibitors (PD-0325901, trametinib, AZD-8330, PD-184352, selumetinib) and related RAS/RAF/MEK/ERK and PI3K/mTOR pathway inhibitors.

4.8 Cross-Study Signal Exists but LOSO Does Not Yield a Usable Estimate (E1)

The leave-one-study-out screen confirmed that maturity is predictable across studies, with macro-F1 of 0.870 for CatBoost, 0.836 for the MLP, 0.718 for logistic regression and 0.565 for Random Forest.
The screen was not usable as an evaluation protocol. It produces M separate scores rather than one estimate, and with study sizes as small as 28 samples the per-study scores carry wide and unequal uncertainty. Several held-out studies do not contain all three classes, so their macro-F1 is computed over a reduced label set and is not comparable across folds. Pooling studies into three balanced groups, as in Section 3.3, resolves both issues: every fold contains all three classes, and every sample receives exactly one out-of-fold prediction.

4.9 Linear Models Dominate in the High-Dimensional Regime (E2)

The three linear models cluster within 0.024 of each other while Random Forest trails by more than 0.3. The class boundary is close to linearly separable and a margin-based fit exploits this directly, whereas an axis-aligned tree ensemble must approximate the same boundary through many splits and overfits study-specific structure in the process. The narrow spread among the linear models indicates the result is driven by the geometry of the data rather than by any particular regulariser.

4.10 One Hundred Genes Recover Almost All of the Signal (E3)

The Lasso zeroed more than 4,400 of the 5,000 coefficients, retaining 542 genes with non-zero mean absolute coefficient across folds. Refitting an $\ell 2$ model, using the nested method to prevent leakage, on the highest 100 gene set each fold gave macro-F1 of 0.9547, against 0.966 for the same model on all 5,000 features.
Under the leakage-free nested protocol, a Ridge evaluator restricted to the top 100 genes per fold reached macro-F1 of 0.955. This is within 0.011 of the full-feature Ridge and within 0.001 of the pooled-selection reference, despite using 2% of the input dimensions and despite the selected genes differing between folds. Discriminative information for the maturity stage is therefore concentrated in a small gene set, and the nesting confirms this is not an artifact of selecting on the evaluation data.

4.11 Selected Genes Are Partly Reproducible Across Folds and Selectors (E4)

Across the three Lasso folds, 4 genes were selected in all three, 24 in exactly two, and 240 in exactly one out of a 100-gene budget per fold. The core set is the stable component of the signature and is what we carry forward for interpretation. The fold-specific remainder reflects the substitutability of correlated genes under $\ell 1$ penalisation, where near-duplicate predictors are selected arbitrarily from within a correlated block.
Comparing selectors at matched set size, Lasso and PC1 loading share 183 genes. The Lasso and permutation intersection is the strongest candidate list, since the two selectors agree despite resting on different assumptions, one on sparsity in a linear coefficient vector and one on model-agnostic degradation under permutation. The smaller overlap with PC1 indicates the maturation signature is not simply the highest-variance axis of the data, which is the outcome that would have undercut the supervised selection entirely.
Summary of unresolved and provisional findings
Several findings arising from this analysis remain provisional and are reported as candidates for targeted follow-up rather than established conclusions. The reversed expression trajectories of ATP2A2 and GJA1 are not explained by any confound identified in this study and warrant direct investigation, potentially through non-linear (rather than strictly monotonic)

modeling of the GSE122380 time course, given the possibility of a non-monotonic expression trajectory not captured by a linear day covariate. The apparent discrepancy between ssGSEA raw-magnitude and GSEA statistical-enrichment timing of the oxidative phosphorylation signal would benefit from an explicit per-sample trajectory analysis plotting pathway score against continuous differentiation day, extending the current three-condition comparison. Finally, the extent to which the confound-excluded core gene set versus the broader differential-expression-based core set differ in their downstream biological interpretation has not been directly quantified and is recommended as a discussion point rather than a resolved finding.
.pdf).

The percentage of references used as self-citations, i.e. the number of references where the authors' names appear, must not exceed 20% of the total.

For more instructions about the references and citations usage please see the appropriate link at the conference website.

5 DISCUSSION

The central aim of this study was to determine whether a genuine, generalizable cardiomyocyte maturation signal could be recovered from heterogeneous public transcriptomic cohorts, and to anchor that signal to the two best-established isoform-switch ratios, MYH7/MYH6 and TNNI3/TNNI1. The results support an affirmative answer, but only after a substantial tissue-composition confound — arising from the mismatch between whole-heart Adult reference tissue and cultured, hypoxia-treated Maturing-condition samples — was explicitly characterized and excluded at the level of pathway membership rather than differential-expression significance alone. Below we interpret the principal findings, reconcile the confound with the marker-based clock motivated in the Introduction, and outline the limitations that should guide follow-up work.

5.1 A pre-specified corroboration framework as a template for confound-aware meta-analysis

The analytical framework applied here required convergence from at least two of three orthogonal structural probes — co-expression topology, transcription-factor regulon activity, and pathway-enrichment timing — before a compositional explanation would be accepted, in addition to the marker-gene and Gene Ontology evidence already in hand. This bar was met comfortably: WGCNA identified no module tracking differentiation day independently of endothelial or neural composite scores, DoRothEA identified activation of canonical NF-κB/immune regulators and the endothelial master regulator ERG specifically in the Adult condition, and Hallmark GSEA localized Hypoxia and Epithelial-Mesenchymal Transition enrichment to comparisons involving the Maturing condition rather than the monotonic pattern expected of genuine developmental progression. That a very large nominal effect — 85.5% of tested genes reaching significance, 62% of those exceeding $|\log2FC| \geq 1$ — proved to be substantially compositional rather than developmental is itself an important methodological finding: effect-size filtering, the field's most common safeguard against significance inflation, is not by itself sufficient to distinguish a tissue-composition confound from developmental biology when the cohorts being compared were collected for unrelated original purposes (here, a hypoxia-treatment study and a whole-organ eQTL resource). The pre-specified, multi-method corroboration framework applied here offers a reusable template for other groups integrating public iPSC-CM datasets before accepting a between-study contrast at face value.

5.2 The isoform-switch ratios anchoring the maturation clock hold within a single cohort but are not robust across between-study contrasts

MYH7/MYH6 and TNNI3/TNNI1 motivated this study as the best-characterized quantitative markers of cardiomyocyte maturation, and both behaved as the underlying developmental biology predicts within GSE122380, the one cohort offering a genuine, confound-free per-sample day gradient: TNNI3 increased at 0.05 log2 fold-change per day (padj = 1.8×10^{-40}), and MYH6/MYH7 were co-induced at near-identical per-day slopes, consistent with pan-sarcomeric induction preceding the postnatal MYH6-to-MYH7 switch as a distinct, later event. Neither ratio reproduced this clean behavior in the Adult-versus-Maturing pairwise contrast: MYH6 and MYH7 rose concurrently rather than divergently, and the TNNI3/TNNI1 ratio inverted entirely ($\Delta\log2 = -0.891$). A ratio built from only two genes carries none of the internal redundancy that lets a pathway- or network-level signal average out a compositional confound; any tissue- or treatment-associated shift affecting the baseline expression of MYH6, MYH7,

TNNI3, or TNNI1 in either study propagates directly into the ratio. This does not undermine the developmental biology motivating the isoform switches (12–15) — the GSE122380 trajectory replicates it precisely — but it does mean a two-marker ratio computed across pooled, between-study public data is not, on its own, a reliable cross-cohort clock. The practical implication is that isoform-ratio staging should be validated within a single, compositionally homogeneous cohort or time-course design, with cross-study benchmarking of a preparation's maturity drawing instead on the broader, confound-corroborated readouts developed in Section 4.4. Encouragingly, six markers remained monotonic across all three contrasts despite the confound — RYR2, CASQ2, MYL2, PLN, MYH7, and TNNI3 — suggesting that excitation–contraction and core sarcomeric genes are comparatively insensitive to this dataset's off-target fibroblast, neuroectodermal, endothelial, and immune contamination, and remain useful first-pass indicators even in uncorrected cross-study comparisons.

5.3 Network topology separates coordinating regulators from structurally obligate hubs

Restricting PPI network construction to the 248-gene confound-excluded core, rather than the larger 3,562-gene differential-expression-based set, was necessary to recover an interpretable mitochondrial/cell-cycle architecture; the broader set instead returned a fibroblast/adhesion network (hubs ESR1, FN1, CDH1, CD44) that recapitulates the off-target lineage signal characterized in Section 4.2 rather than core maturation biology. Within the confound-excluded network, the distinction between degree centrality and betweenness centrality was biologically informative rather than merely descriptive: the highest-degree hubs (UQCRC2, CYC1, UQCRFS1, COX4I1, NDUFB8) are Complex I/III respiratory-chain subunits, densely interconnected by structural necessity and not obvious candidates for independent regulatory function, whereas the highest-betweenness genes — TP53, CDCA8, VDAC1, and CDK1 — sit at the interface between the metabolic and cell-cycle cliques. This pattern is consistent with the established mechanistic coupling between cardiomyocyte cell-cycle exit and mitochondrial biogenesis via p53/E2F signaling (54), and, together with the regulon-level repression of E2F1 and MYCN and activation of CEBPA and PPARG, nominates TP53, CDK1, E2F1, and MYCN as coordinating regulators worth prioritizing for functional follow-up ahead of the structurally obligate respiratory subunits.

5.4 Convergent, direction-stratified evidence nominates MEK/ERK inhibition as a translational hypothesis

The connectivity-mapping result is strengthened less by any single compound than by its internal structure: five independently profiled MEK1/2 inhibitors (PD-0325901, trametinib, AZD-8330, PD-184352, selumetinib), the upstream BRAF inhibitor SB590885, and a secondary cluster of PI3K/mTOR inhibitors were recovered only once the gene signature was stratified by direction of effect — the undirected query instead returned a non-specific, cytotoxicity-dominated hit list. This recovery pattern is consistent with the established role of MAPK/ERK signaling as a driver of cardiomyocyte proliferation and a barrier to maturation, and the co-occurrence of PI3K/mTOR inhibitors alongside the MEK1/2 cluster echoes prior work showing that dual MAPK and PI3K/AKT inhibition enhances iPSC-CM maturation (57). As in Section 4, we emphasize that this is a hypothesis generated from transcriptional similarity to LINCS L1000 perturbation signatures collected predominantly in non-cardiomyocyte cell lines; its translation to iPSC-CM culture requires direct experimental testing and is reported here as a coherent, mechanistically interpretable candidate rather than a validated intervention.

5.5 An independently derived, compact gene signature confirms a cross-study-generalizable maturation signal

The machine-learning component was designed to test a claim the pathway-based analyses could not test directly: whether a maturity classifier trained without knowledge of study identity generalizes to genuinely unseen studies. Grouped, stratified cross-validation gave a macro-F1 of 0.966 on the full 5,000-gene feature set; a nested Lasso–Ridge procedure, with feature selection confined to each training fold, achieved 0.955 using only the top 100 genes per fold — within 0.011 of the full-feature result despite a fifty-fold reduction in dimensionality and despite the selected genes differing across folds. That this leakage-free estimate matched the diagnostic pooled-selection reference to within 0.001 indicates the result is not an artifact of selecting features on the evaluation data. The reproducibility structure is also informative: only 4 of 100 selected genes were common to all three folds, yet 183 genes were shared

between Lasso and PC1 loading at matched set size, suggesting cardiomyocyte maturity occupies a dominant, high-variance axis of the transcriptome recoverable by both a supervised, sparsity-penalized method and an unsupervised, variance-maximizing one, rather than a narrow, idiosyncratic gene set. This ML-derived signature was built independently of, and with different selection logic than, the pathway-based 248-gene confound-excluded set in Section 4.4. We did not quantify the overlap between the two here; doing so is a priority for follow-up, since convergence between a supervised, cross-study classifier and an orthogonal, biology-first, confound-aware pathway analysis would constitute unusually strong evidence that both are capturing the same underlying maturation program.

5.6 Limitations

Several limitations of the source data constrain interpretation. Most consequentially, the Adult condition is built from whole-heart tissue (GTEx, GSE141910) rather than isolated or purified cardiomyocytes, while the Embryonic and Maturing conditions are cultured iPSC-CM preparations; this tissue-composition mismatch is a structural feature of the available public data that cannot be fully removed by statistical correction, even after the corroboration performed here, and residual endothelial-, immune-, or fibroblast-associated signal in the confound-excluded gene set cannot be entirely ruled out. Replication against single-cardiomyocyte or single-nucleus adult reference data would directly address this. The Maturing condition is additionally small (n = 48) and 91.7% derived from a single hypoxia-treatment study, so any Maturing-specific finding not corroborated by the GSE122380 time course should be treated cautiously. WGCNA soft-thresholding did not reach the conventional scale-free topology criterion ($R^2 > 0.85$; maximum achieved was 0.809 at power = 30), reflecting genuine heterogeneity across the constituent cohorts rather than a fitting error, and module-trait correlations should be read with that ceiling in mind. DoRothEA regulon p-values are nominal and were not corrected for the 338 transcription factors tested; regulator-level findings throughout are presented as corroborating evidence for other analyses rather than as independently significant discoveries.

Two findings remain genuinely unresolved rather than attributable to the characterized confound. ATP2A2 and GJA1 moved opposite to the expected maturation direction in every comparison, including the confound-free GSE122380 time course, and warrant direct mechanistic investigation, potentially with non-linear trajectory models capable of capturing a non-monotonic expression path across differentiation day. The apparent tension between ssGSEA raw-score magnitude (indicating metabolic change concentrated late, in the Maturing-to-Adult transition) and GSEA enrichment-timing (indicating oxidative phosphorylation is already statistically enriched in the earliest Maturing-versus-Embryonic transition) is interpreted here as a difference in what each metric measures rather than a contradiction, but that interpretation would be strengthened by continuous per-sample trajectory analysis rather than the three-condition comparison used throughout. Finally, the broader 3,562-gene differential-expression-based candidate set and the 248-gene confound-excluded set were shown to yield materially different network biology when used as PPI inputs, but the extent to which they diverge in downstream functional interpretation more generally has not been systematically quantified and is left as an open question.

6 CONCLUSION

This study integrated 515 human cardiac transcriptomes across four GEO cohorts and GTEx to characterize genome-wide transcriptional change across cardiomyocyte maturation and to evaluate MYH7/MYH6 and TNNI3/TNNI1 isoform-switch ratios as a cross-cohort maturation clock. The central finding is methodological as much as biological: a tissue-composition confound, driven by whole-heart Adult reference tissue and a hypoxia-treatment-dominated Maturing cohort, was large enough (85.5% of the transcriptome nominally significant, 62% at $|\log2FC| \geq 1$) to distort the very isoform ratios the clock was intended to anchor, and was only distinguishable from genuine developmental signal through convergent, pre-specified corroboration across co-expression, regulon-activity, and pathway-timing evidence. Once characterized and excluded at the pathway-membership level, a coherent, biologically interpretable maturation signal remained: a 248-gene confound-resistant core resolving into mitochondrial-respiratory and cell-cycle-exit network modules bridged by TP53 and CDK1, a two-phase glycolysis-to-oxidative-phosphorylation metabolic switch, and a direction-stratified connectivity-mapping signature converging on MEK/ERK and PI3K/mTOR pathway inhibition as a translational hypothesis. Independently, a machine-learning classifier trained without access to

study identity generalized across cohorts and recovered essentially full-feature performance (macro-F1 0.955 versus 0.966) from a compact, 100-gene panel, providing external, cross-validated support that the recovered signal is not an artifact of any single analytical approach.

Taken together, these results indicate that MYH7/MYH6 and TNNI3/TNNI1 remain valid maturation indicators within a single, compositionally controlled cohort, but that robust cross-study benchmarking of iPSC-CM maturity — the standardized, quantitative assessment the field has lacked (10,11) — requires a confound-aware, multi-method framework of the kind developed here, rather than a two-gene ratio applied naively across pooled public data. We offer this framework, the accompanying confound-excluded gene panel and compact ML classifier, and the MEK/ERK inhibition hypothesis as concrete, reusable resources for future work; the reversed ATP2A2/GJA1 trajectories, the oxidative-phosphorylation timing discrepancy, and the relationship between the confound-excluded and broader differential-expression gene sets are identified as specific, tractable questions for that follow-up.

ACKNOWLEDGEMENTS

Should be placed before the references section without numbering.

- **Conflicts of Interest:** The authors have no conflicts of interest or competing interests to declare in the context of this paper.
- **Financial and other Support:** No funding was received for the work reported here.
- **Author Contributions:** The authors contributed equally.
- **Data Sharing:** A dataset was generated and analyzed during the current study, available at <url>.
- **AI Tools:** The paper used AI tools for text correction purposes only.

REFERENCES

1. Moriwaki T, Tani H, Tohyama S. Human induced pluripotent stem cell-derived cardiomyocytes for disease modeling and drug discovery. Front Bioeng Biotechnol. 2025 Oct 16;13:1687840. doi:10.3389/fbioe.2025.1687840
2. Tu C, Chao BS, Wu JC. Strategies for Improving the Maturity of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes. Circulation Research. 2018 Aug 17;123(5):512–4. doi:10.1161/CIRCRESAHA.118.313472
3. Simons E, Loeys B, Alaerts M. iPSC-Derived Cardiomyocytes in Inherited Cardiac Arrhythmias: Pathomechanistic Discovery and Drug Development. Biomedicines. 2023 Jan 25;11(2):334. doi:10.3390/biomedicines11020334 PubMed PMID: 36830871; PubMed Central PMCID: PMC9953535.
4. Feyen DAM, McKeithan WL, Bruyneel AAN, Spiering S, Hörmann L, Ulmer B, et al. Metabolic Maturation Media Improve Physiological Function of Human iPSC-Derived Cardiomyocytes. Cell Reports. 2020 Jul;32(3):107925. doi:10.1016/j.celrep.2020.107925
5. Horikoshi Y, Yan Y, Terashvili M, Wells C, Horikoshi H, Fujita S, et al. Fatty Acid-Treated Induced Pluripotent Stem Cell-Derived Human Cardiomyocytes Exhibit Adult Cardiomyocyte-Like Energy Metabolism Phenotypes. Cells. 2019 Sep 17;8(9):1095. doi:10.3390/cells8091095
6. Ronaldson-Bouchard K, Ma SP, Yeager K, Chen T, Song L, Sirabella D, et al. Advanced maturation of human cardiac tissue grown from pluripotent stem cells. Nature. 2018 Apr;556(7700):239–43. doi:10.1038/s41586-018-0016-3
7. Kroll K, Chabria M, Wang K, Häusermann F, Schuler F, Polonchuk L. Electro-mechanical conditioning of human iPSC-derived cardiomyocytes for translational research. Progress in Biophysics and Molecular Biology. 2017 Nov;130:212–22. doi:10.1016/j.pbiomolbio.2017.07.003
8. Ulmer BM, Stoehr A, Schulze ML, Patel S, Gucek M, Mannhardt I, et al. Contractile Work Contributes to Maturation of Energy Metabolism in hiPSC-Derived Cardiomyocytes. Stem Cell Reports. 2018 Mar;10(3):834–47. doi:10.1016/j.stemcr.2018.01.039
9. Seibertz F, Sutanto H, Dülk R, Pronto JRD, Springer R, Rapedius M, et al. Electrophysiological and calcium-handling development during long-term culture of human-induced pluripotent stem cell-derived cardiomyocytes. Basic Res Cardiol. 2023 Apr 5;118(1):14. doi:10.1007/s00395-022-00973-0
10. Ewoldt JK, DePalma SJ, Jewett ME, Karakan MÇ, Lin YM, Mir Hashemian P, et al. Induced pluripotent stem cell-derived cardiomyocyte in vitro models: benchmarking progress and ongoing challenges. Nat Methods. 2025 Jan;22(1):24–40. doi:10.1038/s41592-024-02480-7
11. Robertson C, Tran DD, George SC. Concise Review: Maturation Phases of Human Pluripotent Stem Cell-Derived Cardiomyocytes. Stem Cells. 2013 May 1;31(5):829–37. doi:10.1002/stem.1331
12. Karbassi E, Fenix A, Marchiano S, Muraoka N, Nakamura K, Yang X, et al. Cardiomyocyte maturation: advances in knowledge and implications

for regenerative medicine. Nat Rev Cardiol. 2020 Jun;17(6):341–59. doi:10.1038/s41569-019-0331-x
13. Hong Y, Zhao Y, Li H, Yang Y, Chen M, Wang X, et al. Engineering the maturation of stem cell-derived cardiomyocytes. Front Bioeng Biotechnol. 2023 Mar 22;11:1155052. doi:10.3389/fbioe.2023.1155052
14. Developmental expression of troponin I isoforms in fetal human heart. FEBS Letters. 1991 Nov 4;292(1–2):5–8. doi:10.1016/0014-5793(91)80820-S
15. Sasse S, Brand NJ, Kyprianou P, Dhoot GK, Wade R, Arai M, et al. Troponin I gene expression during human cardiac development and in end-stage heart failure. Circulation Research. 1993 May;72(5):932–8. doi:10.1161/01.RES.72.5.932
16. Sun Y, Guo C, Chen Z, Lin J, Yang L, Zhang Y, et al. Non-Cell-Autonomous Cardiomyocyte Regulation Complicates Gene Supplementation Therapy for Lmna-Associated Cardiac Defects in Mice. JACC: Basic to Translational Science. 2024 Nov;9(11):1308–25. doi:10.1016/j.jacbts.2024.06.004
17. Miao S, Zhao D, Wang X, Ni X, Fang X, Yu M, et al. Retinoic acid promotes metabolic maturation of human Embryonic Stem Cell-derived Cardiomyocytes. Theranostics. 2020;10(21):9686–701. doi:10.7150/thno.44146
18. Uosaki H, Cahan P, Lee DI, Wang S, Miyamoto M, Fernandez L, et al. Transcriptional Landscape of Cardiomyocyte Maturation. Cell Rep. 2015 Nov 24;13(8):1705–16. doi:10.1016/j.celrep.2015.10.032 PubMed PMID: 26586429; PubMed Central PMCID: PMC4662925.
19. Kannan S, Farid M, Lin BL, Miyamoto M, Kwon C. Transcriptomic entropy benchmarks stem cell-derived cardiomyocyte maturation against endogenous tissue at single cell level. Saucerman JJ, editor. PLoS Comput Biol. 2021 Sep 17;17(9):e1009305. doi:10.1371/journal.pcbi.1009305
20. Chanthra N, Murphy S, Miyamoto M, Masuyama N, Yachie N, Kwon C, et al. Transcriptome-based maturation assessment revealed pro-maturation transcription factors of cardiomyocytes. Front Bioeng Biotechnol. 2026 Feb 18;14:1717996. doi:10.3389/fbioe.2026.1717996
21. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genomics and Bioinformatics. 2020 Sep 1;2(3):lqaa078. doi:10.1093/nargab/lqaa078
22. Strober BJ, Elorbany R, Rhodes K, Krishnan N, Tayeb K, Battle A, et al. Dynamic genetic regulation of gene expression during cellular differentiation. Science. 2019 Jun 28;364(6447):1287–90. doi:10.1126/science.aaw0040
23. Flam E, Jang C, Murashige D, Yang Y, Morley MP, Jung S, et al. Integrated landscape of cardiac metabolism in end-stage human nonischemic dilated cardiomyopathy. Nat Cardiovasc Res. 2022 Sep;1(9):817–29. doi:10.1038/s44161-022-00117-6 PubMed PMID: 36776621; PubMed Central PMCID: PMC9910091.
24. Ward MC, Gilad Y. A generally conserved response to hypoxia in iPSC-derived cardiomyocytes from humans and chimpanzees. Elife. 2019 Apr 8;8:e42374. doi:10.7554/eLife.42374 PubMed PMID: 30958265; PubMed Central PMCID: PMC6538380.
25. Liu Q, Jiang C, Xu J, Zhao MT, Van Bortle K, Cheng X, et al. Genome-Wide Temporal Profiling of Transcriptome and Open Chromatin of Early Cardiomyocyte Differentiation Derived From hiPSCs and hESCs. Circ Res. 2017 Aug 4;121(4):376–91. doi:10.1161/CIRCRESAHA.116.310456 PubMed PMID: 28663367; PubMed Central PMCID: PMC5576565.
26. Li S, Yan B, Wu B, Su J, Lu J, Lam TW, et al. Integrated modeling framework reveals co-regulation of transcription factors, miRNAs and lncRNAs on cardiac developmental dynamics. Stem Cell Res Ther. 2023 Sep 13;14(1):247. doi:10.1186/s13287-023-03442-0 PubMed PMID: 37705079; PubMed Central PMCID: PMC10500942.
27. Ebert A, Joshi AU, Andorf S, Dai Y, Sampathkumar S, Chen H, et al. Proteasome-Dependent Regulation of Distinct Metabolic States During Long-Term Culture of Human iPSC-Derived Cardiomyocytes. Circ Res. 2019 Jun 21;125(1):90–103. doi:10.1161/CIRCRESAHA.118.313973 PubMed PMID: 31104567; PubMed Central PMCID: PMC6613799.
28. The GTEx Consortium, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020 Sep 11;369(6509):1318–30. doi:10.1126/science.aaz1776
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):550. doi:10.1186/s13059-014-0550-8
30. Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A, et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 2018 Jul 18;8(1):10872. doi:10.1038/s41598-018-28948-z
31. Xin J, Mark A, Afrasiabi C, Tsueng G, Juchler M, Gopal N, et al. High-performance web services for querying gene and variant annotation. Genome Biol. 2016 May 6;17(1):91. doi:10.1186/s13059-016-0953-9 PubMed PMID: 27154141; PubMed Central PMCID: PMC4858870.
32. Wu C, MacLeod I, Su AI. BioGPS and MyGene.info: organizing online, gene-centric information. Nucleic Acids Research. 2013 Jan 1;41(D1):D561–5. doi:10.1093/nar/gks1114
33. Sayols S. rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms [Internet]. 2023. doi:10.17912/MICROPUB.BIOLOGY.000811

34. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013 Dec;14(1):128. doi:10.1186/1471-2105-14-128
35. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417–25. doi:10.1016/j.cels.2015.12.004 PubMed PMID: 26771021; PubMed Central PMCID: PMC4707969.
36. Fang Z, Liu X, Peltz G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Lu Z, editor. Bioinformatics. 2023 Jan 1;39(1):btac757. doi:10.1093/bioinformatics/btac757
37. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009 Nov;462(7269):108–12. doi:10.1038/nature08460
38. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec;9(1):559. doi:10.1186/1471-2105-9-559
39. Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Research. 2019 Jul 2;47(W1):W212–24. doi:10.1093/nar/gkz446
40. Badia-i-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Kuijjer ML, editor. Bioinformatics Advances. 2022 Jan 10;2(1):vbac016. doi:10.1093/bioadv/vbac016
41. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019 Aug;29(8):1363–75. doi:10.1101/gr.240663.118
42. Türei D, Schaul J, Palacio-Escat N, Bohár B, Bai Y, Ceccarelli F, et al. OmniPath: integrated knowledgebase for multi-omics analysis. Nucleic Acids Research. 2026 Jan 6;54(D1):D652–60. doi:10.1093/nar/gkaf1126
43. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638–46. doi:10.1093/nar/gkac1000 PubMed PMID: 36370105; PubMed Central PMCID: PMC9825434.
44. Delcroix V, Mauduit O, Lee HS, Ivanova A, Umazume T, Knox SM, et al. The First Transcriptomic Atlas of the Adult Lacrimal Gland Reveals Epithelial Complexity and Identifies Novel Progenitor Cells in Mice. Cells. 2023 May 21;12(10):1435. doi:10.3390/cells12101435
45. McIntire E, Barr KA, Gonzales NM, Allen OL, Gilad Y. Guided Differentiation of Pluripotent Stem Cells into Heterogeneously Differentiating Cultures of Cardiac Cells. bioRxiv: The Preprint Server for Biology; 2025. doi:10.1101/2023.07.21.550072 PubMed PMID: 37502898; PubMed Central PMCID: PMC10370173.
46. Monavarfeshani A, Yan W, Pappas C, Odenigbo KA, He Z, Segrè AV, et al. Transcriptomic analysis of the ocular posterior segment completes a cell atlas of the human eye. Proc Natl Acad Sci U S A. 2023 Aug 22;120(34):e2306153120. doi:10.1073/pnas.2306153120 PubMed PMID: 37566633; PubMed Central PMCID: PMC10450437.
47. Pareja-Lorente E, Aloy P. Systematic assessment of microenvironment-dependent transcriptional patterns and intercellular communication. Genome Biol. 2025 Jul 6;26(1):193. doi:10.1186/s13059-025-03677-5 PubMed PMID: 40619422; PubMed Central PMCID: PMC12232842.
48. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545–50. doi:10.1073/pnas.0506580102 PubMed PMID: 16199517; PubMed Central PMCID: PMC1239896.
49. Thomas D, Cunningham NJ, Shenoy S, Wu JC. Human-induced pluripotent stem cells in cardiovascular research: current approaches in cardiac differentiation, maturation strategies, and scalable production. Cardiovascular Research. 2022 Jan 7;118(1):20–36. doi:10.1093/cvr/cvab115
50. Mills RJ, Hudson JE. Bioengineering adult human heart tissue: How close are we? APL Bioeng. 2019 Mar;3(1):010901. doi:10.1063/1.5070106 PubMed PMID: 31069330; PubMed Central PMCID: PMC6481734.
51. Emery P, Durand B, Mach B, Reith W. RFX Proteins, a Novel Family of DNA Binding Proteins Conserved in the Eukaryotic Kingdom. Nucleic Acids Research. 1996 Jan 1;24(5):803–7. doi:10.1093/nar/24.5.803
52. Sun SC. The non-canonical NF-κB pathway in immunity and inflammation. Nat Rev Immunol. 2017 Sep;17(9):545–58. doi:10.1038/nri.2017.52 PubMed PMID: 28580957; PubMed Central PMCID: PMC5753586.
53. Padula SL, Velayutham N, Yutzey KE. Transcriptional Regulation of Postnatal Cardiomyocyte Maturation and Regeneration. Int J Mol Sci. 2021 Mar 23;22(6):3288. doi:10.3390/ijms22063288 PubMed PMID: 33807107; PubMed Central PMCID: PMC8004589.
54. Mak TW, Hauck L, Grothe D, Billia F. p53 regulates the cardiac transcriptome. Proc Natl Acad Sci U S A. 2017 Feb 28;114(9):2331–6.

doi:10.1073/pnas.1621436114 PubMed PMID: 28193895; PubMed Central PMCID: PMC5338492.
55. Zhang J, Nuebel E, Daley GQ, Koehler CM, Teitell MA. Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. Cell Stem Cell. 2012 Nov 2;11(5):589–95. doi:10.1016/j.stem.2012.10.005 PubMed PMID: 23122286; PubMed Central PMCID: PMC3492890.
56. Cliff TS, Dalton S. Metabolic switching and cell fate decisions: implications for pluripotency, reprogramming and development. Curr Opin Genet Dev. 2017 Oct;46:44–9. doi:10.1016/j.gde.2017.06.008 PubMed PMID: 28662447; PubMed Central PMCID: PMC5842063.
57. Garay BI, Givens S, Abreu P, Liu M, Yücel D, Baik J, et al. Dual inhibition of MAPK and PI3K/AKT pathways enhances maturation of human iPSC-derived cardiomyocytes. Stem Cell Reports. 2022 Sep;17(9):2005–22. doi:10.1016/j.stemcr.2022.07.003

APPENDIX

If any, the appendix should appear directly after the references without numbering, and not on a new page.