Cardiomyocyte Cell-Cycle Regulation in Neonatal Large Mammals: Single Nucleus RNA-Sequencing Data Analysis *via* an Artificial-Intelligence–Based Pipeline

Thanh Nguyen[1†], Yuhua Wei[1†], Yuji Nakada[1], Yang Zhou[1] and Jianyi Zhang[1,2]*

[1]Department of Biomedical Engineering, University of Alabama at Birmingham, Birmingham, AL, United States, [2]Cardiovascular Diseases, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

Adult mammalian cardiomyocytes have very limited capacity to proliferate and repair the myocardial infarction. However, when apical resection (AR) was performed in pig hearts on postnatal day (P) 1 (AR_{P1}) and acute myocardial infarction (MI) was induced on P28 (MI_{P28}), the animals recovered with no evidence of myocardial scarring or decline in contractile performance. Furthermore, the repair process appeared to be driven by cardiomyocyte proliferation, but the regulatory molecules that govern the AR_{P1}-induced enhancement of myocardial recovery remain unclear. Single-nucleus RNA sequencing (snRNA-seq) data collected from fetal pig hearts and the hearts of pigs that underwent AR_{P1}, MI_{P28}, both AR_{P1} and MI, or neither myocardial injury were evaluated *via* autoencoder, cluster analysis, sparse learning, and semisupervised learning. Ten clusters of cardiomyocytes (CM1–CM10) were identified across all experimental groups and time points. CM1 was only observed in AR_{P1} hearts on P28 and was enriched for the expression of T-box transcription factors 5 and 20 (TBX5 and TBX20, respectively), Erb-B2 receptor tyrosine kinase 4 (ERBB4), and G Protein-Coupled Receptor Kinase 5 (GRK5), as well as genes associated with the proliferation and growth of cardiac muscle. CM1 cardiomyocytes also highly expressed genes for glycolysis while lowly expressed genes for adrenergic signaling, which suggested that CM1 were immature cardiomyocytes. Thus, we have identified a cluster of cardiomyocytes, CM1, in neonatal pig hearts that appeared to be generated in response to AR injury on P1 and may have been primed for activation of CM cell-cycle activation and proliferation by the upregulation of TBX5, TBX20, ERBB4, and GRK5.

Keywords: single-nucleus RNA-sequencing, heart, infarct, cardiomyocytes, cell cycle, autoencoder

INTRODUCTION

Mammalian cardiomyocytes exit the cell cycle during the perinatal period and lose the ability to proliferate; thus, the hearts of mammals are unable to repair the damage caused by myocardial injury that occurs more than 2 days after birth (Porrello et al., 2011; Lam and Sadek, 2018; Ye et al., 2018; Zhu et al., 2018). However, when apical resection (AR) was performed in pig hearts on postnatal day (P) 1 (AR_{P1}), and acute myocardial infarction (MI) was induced on P28 (MI_{P28}), the animals completely recovered with no evidence of myocardial scarring or decline in contractile performance

by P56, whereas the hearts of animals that underwent MI on P28 without previous AR injury displayed significant fibrosis and declines in contractile activity (Zhao et al., 2020). Furthermore, the repair process appeared to be driven by the proliferation of cardiomyocytes, and comparative analyses of bulk and single-nuclei RNA sequencing (snRNA-seq) data from the hearts of animals that underwent AR_{P1} only, MI_{P28} only, or both AR_{P1} and MI_{P28} ($AR_{P1}MI_{P28}$), as well as uninjured (CTL) and fetal pig hearts, suggested that signaling pathways associated with cell-cycle activity, glycolytic metabolism, and declines in DNA damage were upregulated in the cardiomyocytes of $AR_{P1}MI_{P28}$ hearts (Zhang et al., 2020; Zhao et al., 2020; Nakada et al., 2022). Also, in our previous study using snRNA-seq data (Nakada et al., 2022), we found a novel cardiomyocyte subpopulation marked by coupregulation of Nebulin (NEB) and Pyruvate Kinase M1/2 (PKM), which uniquely appeared in regenerative $AR_{P1}MI_{P28}$ heart on postnatal day P35. On the other hand, how AR_{P1} cardiomyocytes differed from CTL ones such that they responded differently following MI_{P28} injury was not thoroughly examined. For the studies presented in this report, we collected more snRNA-seq data from the cardiac tissues of additional animals and then analyzed our expanded dataset with an artificial-intelligence–based pipeline for deeper understanding on which regulatory molecules and signaling pathways contributed to the AR_{P1}-associated enhancement of myocardial regeneration, and which cardiomyocyte subsets highly utilizes these regulators. We expanded the snRNA-seq dataset by obtaining a new MI_{P28}-only group on postnatal days P30, P35, and P42 and more AR_{P1}-MI_{P28} pigs at P30, P35, and P42. In addition, an artificial-intelligence technique was developed and applied for high-dimensional snRNA-seq data using a much smaller number of dimensions (Wang et al., 2016) and consequently improved snRNA-seq data clustering findings.

RESULTS

Autoencoding and Cluster Analysis Identified 10 Cardiomyocyte Populations in the Hearts of Fetal, AR_{P1}, MI_{P28}, and $AR_{P1}MI_{P28}$ Animals

Our complete dataset encompassed the results from snRNA-seq analyses of myocardial tissues in animals that underwent AR_{P1} only and were sacrificed on P28 and P56 (AR_{P1}-P28 and ARP1-P56, respectively); animals that underwent MI_{P28} only and were sacrificed on P30, P35, P42, and P56 (MI_{P28}-P30, -P35, -P42, and -P56, respectively); animals that underwent both AR_{P1} and MI_{P28} and were sacrificed on P30, P35, P42, and P56 ($AR_{P1}MI_{P28}$-P30, -P35, -P42, and -P56, respectively); CTL animals that were sacrificed on P1, P28, and P56 (CTL-P1, -P28, and -P56, respectively); and embryos obtained on embryonic day 80 (Fetal) (Supplementary Table S1). Tissues were collected from the border zone of infarction or the corresponding region of non-infarcted hearts, and nuclei with <500 or >25,000 unique molecular identifiers (UMIs), or with >25% mitochondrial UMIs were excluded from subsequent analyses. A total of

283,421 nuclei from 41 hearts were included in our analyses; 1,786 (median) genes and 31,736 (median) UMIs were captured per nucleus (Supplementary Table S2), and 129,991 of the nuclei were from cardiomyocytes. Data were aligned and normalized with the Seurat v.4 toolkit (Hao et al., 2021), and then embedded with an autoencoder before clustering and visualization via Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018; Meehan, 2021); this deep-learning–based method has outperformed other state-of-the-art approaches for unsupervised cluster analysis (Yang et al., 2019).

When data for all cell types were analyzed (Supplementary Figure S1), most of the clusters contained cells from multiple injury groups and time points, indicating that the results were not influenced by between-batch variation or sampling error, and each cluster was composed primarily of a single cell type (cardiomyocytes, smooth muscle cells, endothelial cells, fibroblasts, immune cells, or skeletal muscle–like cells. The skeletal muscle–like cluster uniquely expressed exclusive-skeletal-markers myosin light chain 3 (MYL3) (Hailstones and Gunning, 1990) and Myosin light chain kinase 2 (MYLK2) (Stull et al., 2011); meanwhile, it expresses cardiac Actin Alpha Cardiac Muscle 1 (ACTC1) and Myosin Heavy Chain 7 (MYH7). Cardiac muscle populations expressing both cardiomyocyte and skeletal muscle markers were reported in Clément et al. (1999). Also, the cardiac/skeletal muscle–like cluster upregulated both Nebulin (NEB) and Pyruvate Kinase M1/2 (PKM), which were consistent with the PKM+/NEB + cardiomyocyte subpopulation in our previous report (Nakada et al., 2022). Cardiomyocytes were distributed into 10 clusters (CM1–CM10) (Figures 1A, B), each of which was associated with a set of explicitly expressed marker genes (Figure 1C, Supplementary Tables S3–S4).

Cardiomyocytes in the CM1 cluster were found almost exclusively in AR_{P1}-P28 hearts, where they comprised 62.91% of the total cardiomyocyte population (Figure 1D). The CM2 cluster included a substantial proportion of cardiomyocytes from CTL (22.53%) and AR_{P1} (25.11%) hearts at P28, as well as $AR_{P1}MI_{P28}$ hearts at P30 (31.77%) and P35 (21.17%). CM3 cardiomyocytes were present only in CTL-P56 hearts, where they comprised just 2.63% of all cardiomyocytes and expressed elevated levels of genes associated with cell differentiation. Small numbers of CM4 cardiomyocytes, which expressed high levels of collagen [Collagen Type V Alpha 2 Chain (COL5A2) and Collagen Type III Alpha 1 Chain (COL3A1)] and genes that regulate the pluripotency of stem cells, including APC (Kielman et al., 2002), PIK3CA (Jeong et al., 2017), MAPK1 (Lu et al., 2008), and JARID2 (Landeira et al., 2010) (Supplementary Figure S2A), were present in all hearts. Cardiomyocytes from CM5 were only in CTL-P56 hearts and explicitly expressed genes that drive cardiomyocyte maturation, such as Ankyrin Repeat And SOCS Box Containing 18 (ASB18), Yip1 Domain Family Member 7 (YIPF7), Creatine Kinase, M-Type (CKM), and Creatine Kinase, Mitochondrial 2 (CKMT2) (Uosaki et al., 2015; Guo and Pu, 2020). CM6 included the majority (57.16%–69.69%) of cardiomyocytes from all injury groups (AR_{P1}, MI_{P28}, and $AR_{P1}MI_{P28}$) on P56 and was enriched for
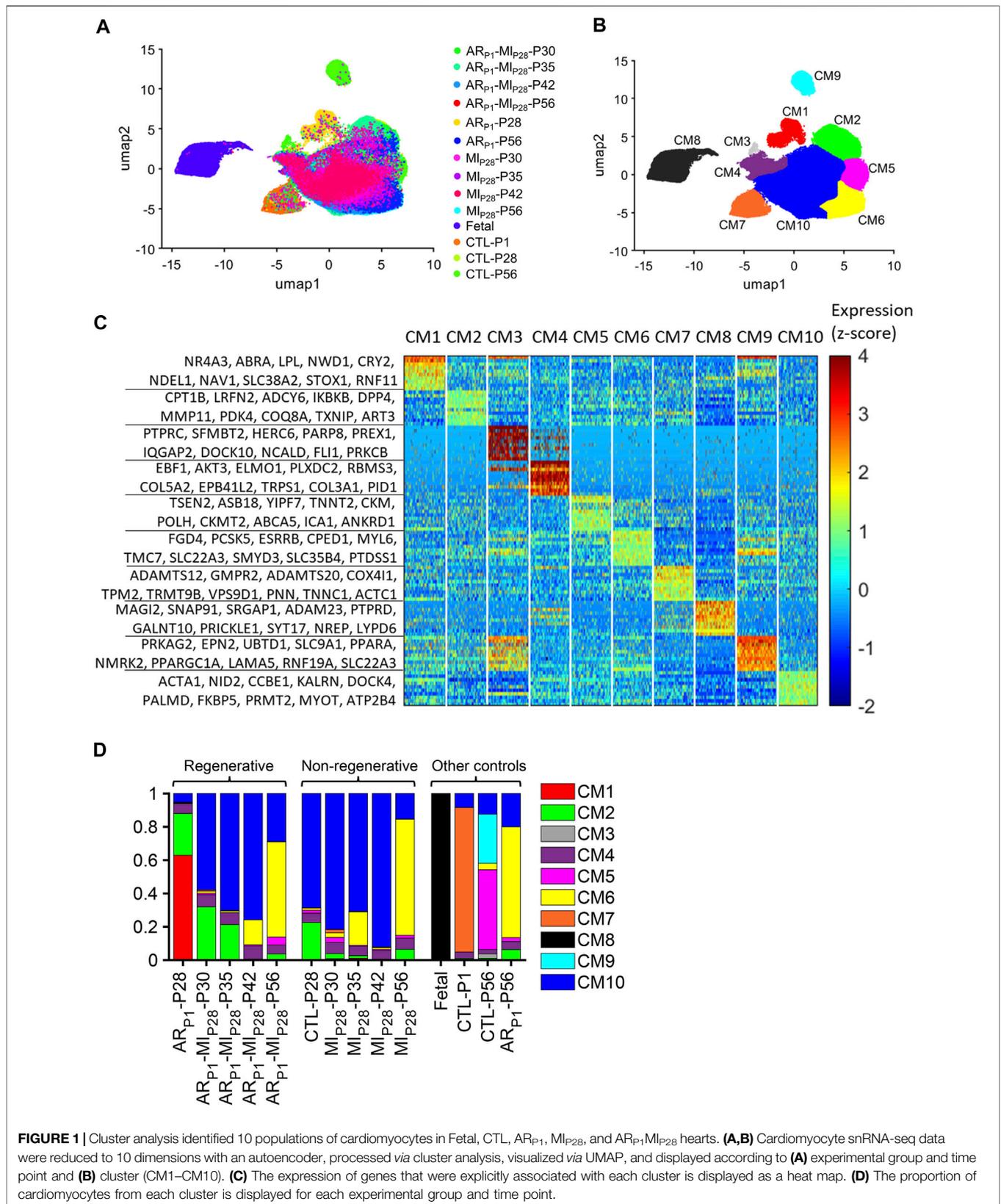
FIGURE 1 | Cluster analysis identified 10 populations of cardiomyocytes in Fetal, CTL, AR_{P1}, MI_{P28}, and $AR_{P1}MI_{P28}$ hearts. (A,B) Cardiomyocyte snRNA-seq data were reduced to 10 dimensions with an autoencoder, processed via cluster analysis, visualized via UMAP, and displayed according to (A) experimental group and time point and (B) cluster (CM1–CM10). (C) The expression of genes that were explicitly associated with each cluster is displayed as a heat map. (D) The proportion of cardiomyocytes from each cluster is displayed for each experimental group and time point.
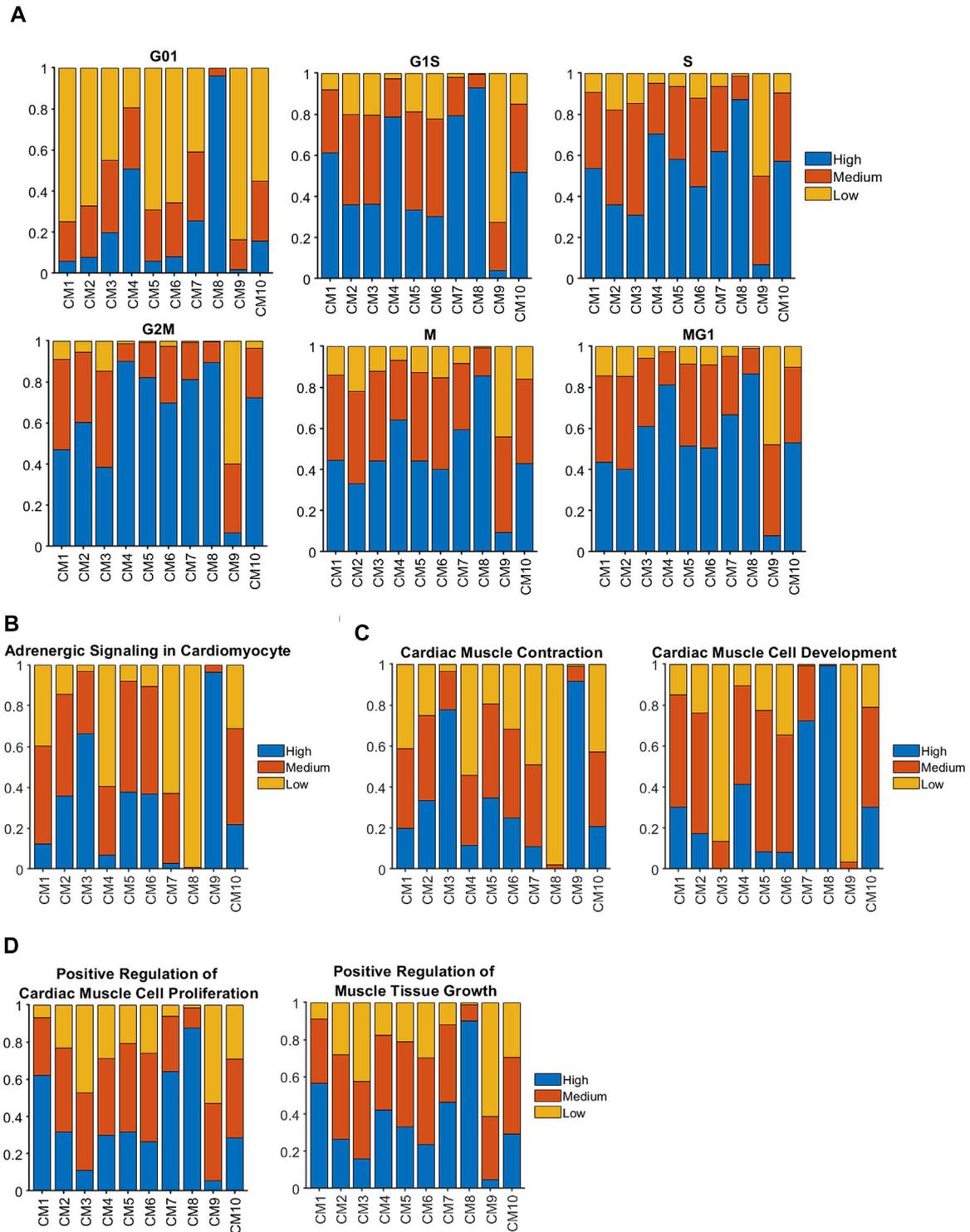
FIGURE 2 | Genes that contribute to cell-cycle activity and muscle growth tended to be more highly expressed in clusters associated with regenerating hearts than in CTL hearts at P28 and afterward. Sparse analysis was conducted for the expression of genes associated with **(A)** the cell-cycle, **(B)** adrenergic signaling, **(C)** the contractile activity and development of cardiac muscle, and **(D)** cardiac-muscle cell proliferation and muscular growth; then, the proportion of cardiomyocytes with high, medium, or low levels of expression was calculated for each cluster.

the expression of Z-disc and insulin-signaling genes. CM7 encompassed most (86.75%) of the cardiomyocytes in CTL-P1 hearts and was associated with elevated levels of genes that participate in the morphogenesis of cardiac muscle and myofibril assembly. The CM8 cluster included nearly all (99.69%) of the cardiomyocytes in fetal hearts, none of those from any other group or time point, and was enriched for genes that contribute to embryonic development. CM9 cardiomyocytes were present only in CTL-P56 hearts, where they comprised 29.51% of all cardiomyocytes, and expressed high levels of genes associated with the Z-disc, focal-adhesion, and other structural components of muscle cells. Finally, cardiomyocytes from the CM10 cluster were found in all nonfetal groups at all time points and included the majority (57.85%–92.16%) of cardiomyocytes in MI_{P28} and $AR_{P1}MI_{P28}$ hearts from P30-P42. The pairwise similarities among these clusters were presented in **Supplementary Figure S3**.

AR_{P1} and MI_{P28}, Both Alone and in Combination, Promoted Cardiomyocyte Cell-Cycle Activity for Approximately Two Weeks After Myocardial Infarction Induction

Sparse modeling (Bi et al., 2003; Huang et al., 2010; Chkifa et al., 2015; Zhang et al., 2014) enables researchers to extract relevant data from datasets that contain a large number of variables that do not contribute to the property being studied. When sparse modeling was used to evaluate the expression of cell-cycle genes in our cardiomyocyte snRNA-seq dataset (**Supplementary Figure S4**), our results are consistent with our previous report in (Nakada et al., 2022). Briefly, the proportion of high-cell-cycle in each phase of the cell cycle was the greatest in Fetal and CTL-P1 hearts. Also, these proportions in CTL-P1 hearts were higher than in hearts from any other postnatal group, which is consistent with the perinatal occurrence of cardiomyocyte cell-cycle arrest. However, cycling cardiomyocytes were much more common in AR_{P1} hearts at P28, and in both MI_{P28} and $AR_{P1}MI_{P28}$ hearts from P30-P42, than in CTL hearts at P28 or P56. Cell-cycle activity was also significantly more common in cardiomyocytes from $AR_{P1}MI_{P28}$ hearts than from MI_{P28} hearts on P35 (G1S: $p = 1.06 \times 10^{-58}$; S: $p = 9.67 \times 10^{-12}$; G2M: $p = 7.44 \times 10^{-66}$; M: $p = 4.11 \times 10^{-7}$) but not on P42, which suggests that AR on P1 promoted cardiomyocyte proliferation for approximately 1 week after MI induction on P28. Notably, this period of AR_{P1}-enhanced proliferation coincided with a greater proportion of CM2 cardiomyocytes in $AR_{P1}MI_{P28}$ hearts.

AR_{P1}-P28 Hearts Contained a Cluster of Cardiomyocytes With a Latent Capacity for Myocardial Proliferation and Growth

When cell-cycle gene expression was evaluated for cardiomyocyte (CM1) clusters (**Figure 2A**), cycling activity tended to be highest in CM8 and lowest in CM9, which is consistent with our observation that these two clusters were almost exclusively associated with Fetal and CTL-P56 hearts, respectively. The proportion of cycling cells was also elevated among CM4

cardiomyocytes, which express high levels of pluripotency-maintenance genes (Mouse Genome Informatics, 2022) (**Supplementary Figure S2B**) and consequently, appear to have some progenitor-cell–like properties. However, CM1 cardiomyocytes, which were found only in AR_{P1} hearts at P28, were not especially more proliferative than CM9 (primarily in CTL-P56); for example, G2M (Mouse Genome Informatics, 2021a) [odds ratio (OR) = 2.66, $p = 3.38 \times 10^{-76}$] and MG1 (Cui et al., 2020) (OR = 1.45, $p = 1.59 \times 10^{-21}$) cells were less common in CM1 than in CM10, which comprised the bulk of cardiomyocytes in both MI_{P28} and $AR_{P1}MI_{P28}$ hearts during the 2 weeks after MI induction. Nevertheless, analyses of adrenergic signaling (KEGG: Adrenergic signaling in cardiomyocytes, 2021) (**Figure 2B**), cardiac-muscle contraction (Mouse Genome Informatics, 2021b), and cardiac-muscle–cell development (Gene Ontology: Cardiac Muscle Cell Development, 2021) (**Figure 2C**) indicated that CM1 cardiomyocytes were functionally immature, while genes associated with the proliferation (Mouse Genome Informatics, 2021c) (OR = 5.61, $p < 10^{-60}$) and growth (Mouse Genome Informatics, 2021d) (OR = 4.35, $p < 10^{-60}$) of cardiac muscle (**Figure 2D**) were more highly expressed in CM1 than in CM10. Collectively, these observations suggest that although CM1 cardiomyocytes themselves did not display exceptionally high levels of cell-cycle activity, they were still immature and might retain a latent capacity for proliferation that could have been reactivated by MI induction on P28.

CM1 Cardiomyocytes Express Elevated Levels of T-Box Transcription Factors 5 and 20 (TBX5 and TBX20), Erb-B2 Receptor Tyrosine Kinase 4, and G Protein-Coupled Receptor Kinase 5

CM1 cardiomyocytes were enriched for expression of the regulatory molecules TBX5 ($p = 8.65 \times 10^{-184}$), TBX20 ($p = 2.97 \times 10^{-225}$), and Erb-B2 receptor tyrosine kinase 4 (ERBB4) ($p = 4.86 \times 10^{-186}$) (**Figure 3A**). Notably, these three genes were also coupregulated in the clusters that were exclusively associated with Fetal hearts (CM8) and with CTL hearts on P1 (CM7), and previous reports have shown that disruptions of TBX5 (Misra et al., 2014; Maitra et al., 2009), TBX20 (Xiang et al., 2016; Chakraborty and Yutzey, 2012), or ERBB4 (Bersell et al., 2009) activity reduce cardiomyocyte proliferation in fetal and neonatal mouse hearts. Furthermore, the expression of G protein-coupled receptor kinase 5 (GRK5), which is associated with cardiac hypertrophy (Gold et al., 2012; Traynham et al., 2015), was upregulated in CM1 cardiomyocytes. Thus, we queried the TRRUSTv2 (Han et al., 2018) and STRING v11.5 (Szklarczyk et al., 2021) databases to identify the genes that are targeted by or interact with these four molecules, and then used the database for annotation, visualization, and integrated discovery (DAVID) bioinformatics resources (Huang et al., 2009) with the Gene Ontology Annotation (GOA) (Huntley et al., 2015), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017), and Reactome (Jassal et al., 2020) databases to characterize the network (**Figure 3B**) of biological processes and signaling pathways that may have been upregulated in CM1
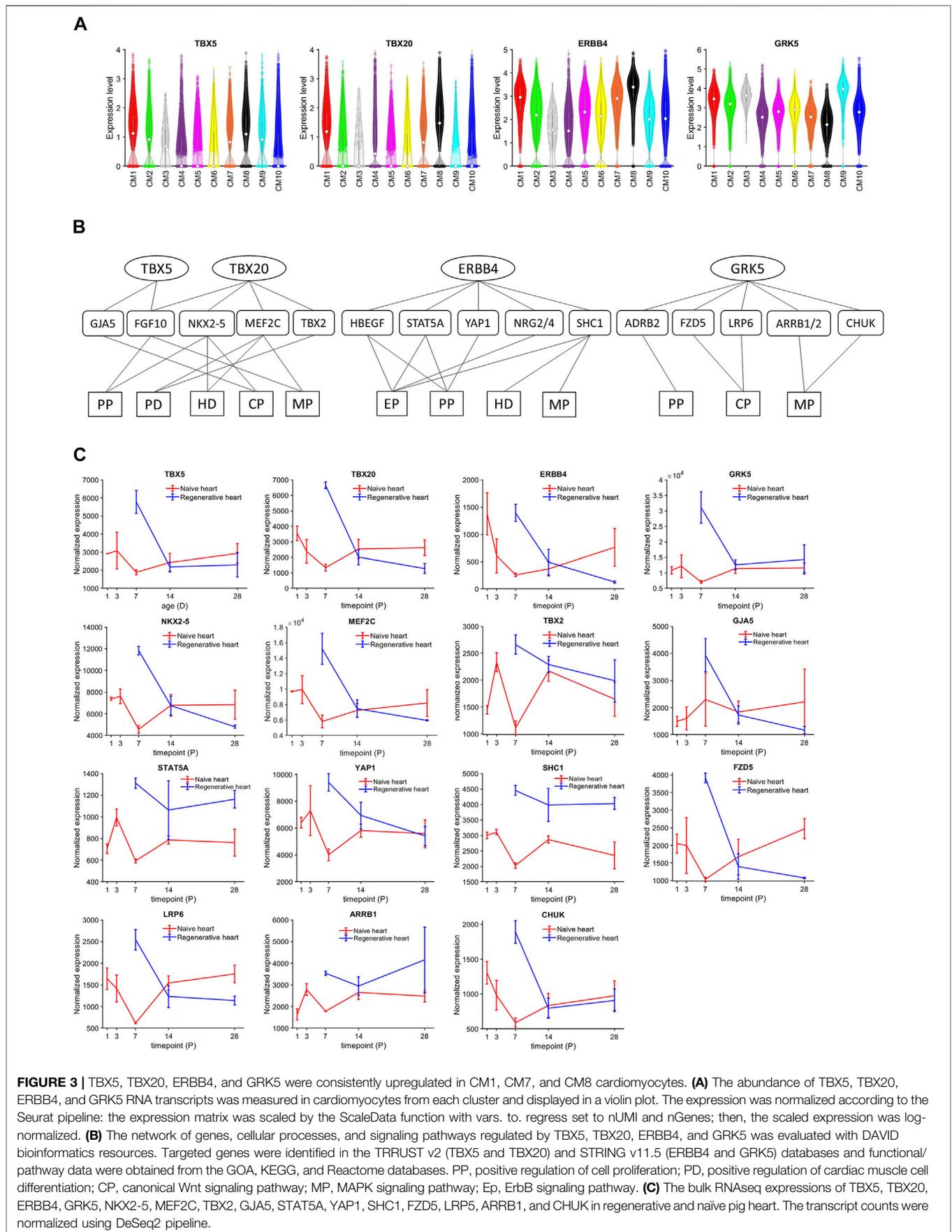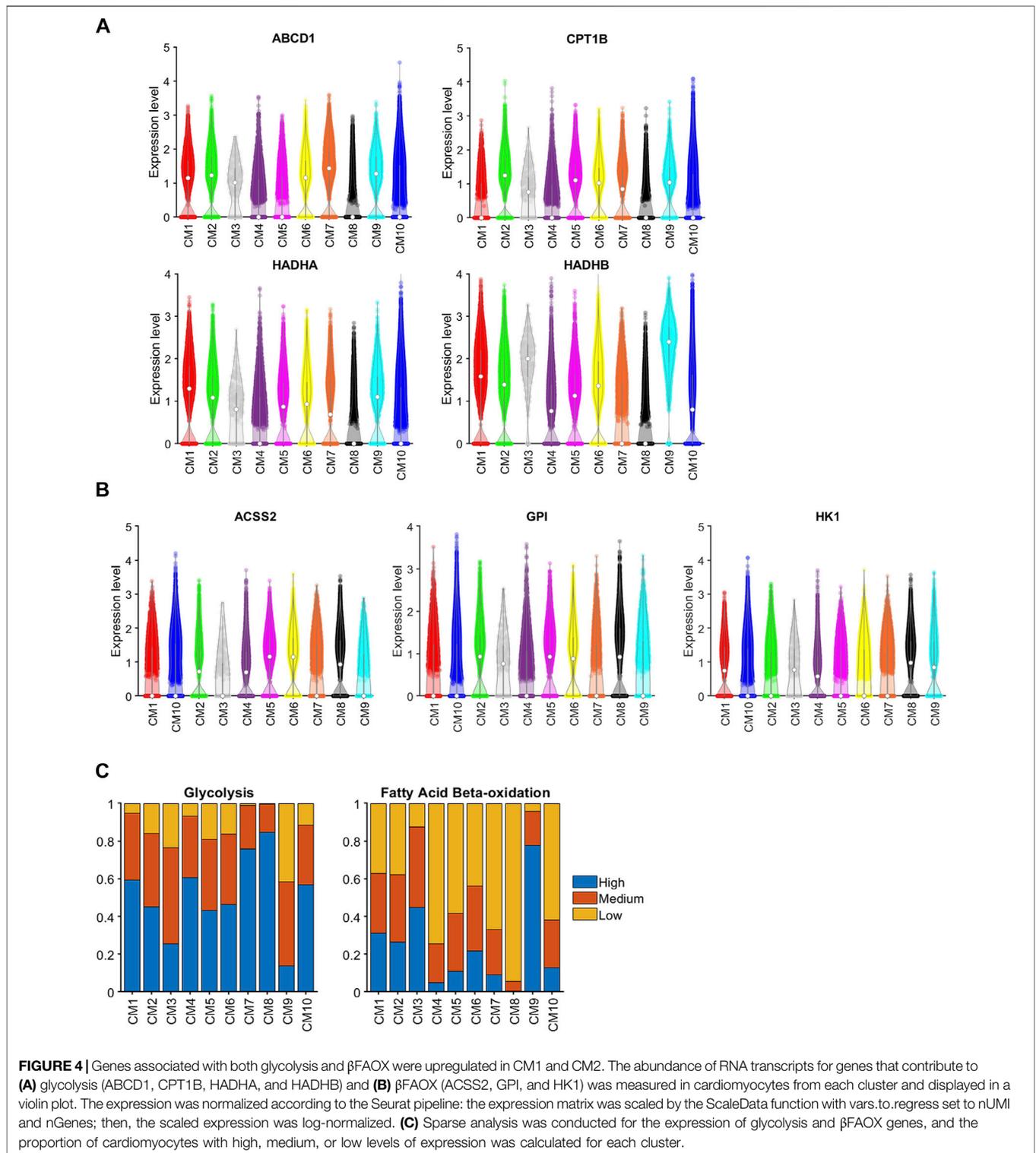
FIGURE 3 | TBX5, TBX20, ERBB4, and GRK5 were consistently upregulated in CM1, CM7, and CM8 cardiomyocytes. **(A)** The abundance of TBX5, TBX20, ERBB4, and GRK5 RNA transcripts was measured in cardiomyocytes from each cluster and displayed in a violin plot. The expression was normalized according to the Seurat pipeline: the expression matrix was scaled by the ScaleData function with vars. to. regress set to nUMI and nGenes; then, the scaled expression was log-normalized. **(B)** The network of genes, cellular processes, and signaling pathways regulated by TBX5, TBX20, ERBB4, and GRK5 was evaluated with DAVID bioinformatics resources. Targeted genes were identified in the TRRUST v2 (TBX5 and TBX20) and STRING v11.5 (ERBB4 and GRK5) databases and functional/pathway data were obtained from the GOA, KEGG, and Reactome databases. PP, positive regulation of cell proliferation; PD, positive regulation of cardiac muscle cell differentiation; CP, canonical Wnt signaling pathway; MP, MAPK signaling pathway; Ep, ErbB signaling pathway. **(C)** The bulk RNAseq expressions of TBX5, TBX20, ERBB4, GRK5, NKX2-5, MEF2C, TBX2, GJA5, STAT5A, YAP1, SHC1, FZD5, LRP5, ARRB1, and CHUK in regenerative and naïve pig heart. The transcript counts were normalized using DeSeq2 pipeline.

FIGURE 4 | Genes associated with both glycolysis and βFAOX were upregulated in CM1 and CM2. The abundance of RNA transcripts for genes that contribute to **(A)** glycolysis (ABCD1, CPT1B, HADHA, and HADHB) and **(B)** βFAOX (ACSS2, GPI, and HK1) was measured in cardiomyocytes from each cluster and displayed in a violin plot. The expression was normalized according to the Seurat pipeline: the expression matrix was scaled by the ScaleData function with vars.to.regress set to nUMI and nGenes; then, the scaled expression was log-normalized. **(C)** Sparse analysis was conducted for the expression of glycolysis and βFAOX genes, and the proportion of cardiomyocytes with high, medium, or low levels of expression was calculated for each cluster.

cardiomyocytes. Collectively, our results identified increases in processes of cell proliferation (Mouse Genomic Informatics, 2021a) and cardiac-muscle–cell differentiation (Mouse Genome Informatics, 2021c), as well as in ErbB (Zhu et al., 2010), MAPK (Liu and Zhong, 2017), and canonical Wnt (Kwon

et al., 2007) signaling (**Supplementary Tables S5, S6**). Intermediate molecules in the network included: 1) Fibroblast Growth Factor 10 (FGF10), which is activated by both TBX5 and TBX20, phosphorylates Forkhead Box O3 (FOXO3), and downregulates the cell-cycle inhibitor p27 (Rochais et al.,

2014); 2) Gap Junction Protein Alpha 5 (GJA5), which is activated by TBX5 and contributes to the differentiation of induced-pluripotent stem cells into cardiomyocytes, endothelial cells, and other cardiac cell types (Paik et al., 2018); 3) Myocyte Enhancer Factor 2C (MEF2C) and NK2 Homeobox 5 (NKX2-5), which are upregulated by TBX20 and cooperatively participate in the embryonic development of mouse hearts (Vincentz et al., 2008; Materna et al., 2019); 4) the Hippo pathway effector Yes-associated Protein 1 (YAP1), which is activated by ERBB4 and regulates the cell cycle by interacting with transcription factors of the Transcriptional Enhanced Associate Domain (TEAD) family (Zhao et al., 2008; Yuan et al., 2020); and 5) Beta-2 Adrenergic Receptor (ADRB2), as well as downstream components of the canonical Wnt signaling pathway (Chen et al., 2009), which may reactivate proliferation in mature cardiomyocytes (Fan et al., 2018). Notably, bulk-RNA seq data (Zhang et al., 2020) cross-checking also showed the overexpression of TBX5, TBX20, ERBB4, GRK5, NKX2-5, MEF2C, TBX2, GJA5, STAT5A, YAP1, SHC1, FZD5, LRP6, ARRB1, and CHUK among regenerative hearts (**Figure 3C**). Here, these regenerative-hearts expressions on postnatal day 7 (P7) were even higher than they are in naïve-hearts P1 and P3.
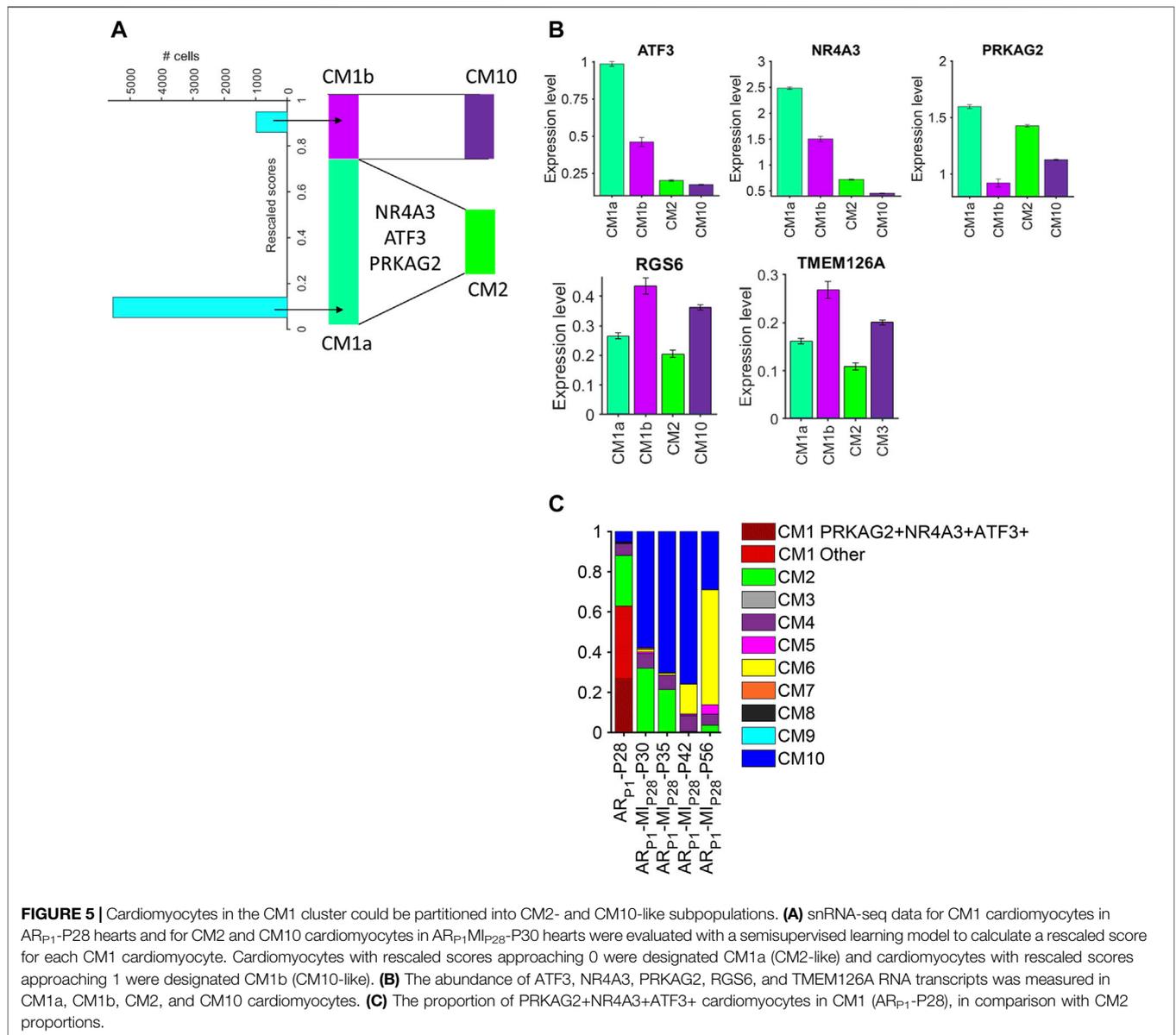
CM1 and CM2 Cardiomyocytes Were Enriched for the Expression of Genes That Contribute to Both Glycolysis and β Fatty Acid Oxidation

As cardiomyocytes mature, the primary mechanism for ATP generation switches from glycolysis to beta fatty acid oxidation (β-FAOX) (Lopaschuk and Jaswal, 2010), which is consistent with our observation that glycolysis genes [Acyl-CoA Synthetase Short Chain Family Member 2 (ACSS2), Glucose-6-Phosphate Isomerase (GPI), and Hexokinase 1 (HK1) (Glycolysis, 2021)] were highly expressed in the CM8 cluster (i.e., fetal cardiomyocytes) but not in CM9 (i.e., the CTL-P56–exclusive cluster), while genes involved in β-FAOX [ATP Binding Cassette Subfamily D Member 1 (ABCD1) (van Roermund et al., 2011), Carnitine Palmitoyltransferase 1B (CPT1B) (Angelini et al., 2021), and Hydroxyacyl-CoA Dehydrogenase Trifunctional Multienzyme Complex Subunits Alpha (Miklas et al., 2019) and Beta (Sekine et al., 2021) (HADHA and HADHB, respectively)] were downregulated in CM8 and upregulated in CM9 (**Figures 4A, B**). Sparse-model analysis indicated that glycolysis (Glycolysis, 2021) was the dominant metabolic pathway in most other cardiomyocyte clusters. Compared to CM9 (primarily CTL-P56 cardiomyocytes), glycolysis markers were upregulated in CM1 (OR = 9.21, $p < 10^{-60}$), CM2 (OR = 5.23, $p < 10^{-60}$), and CM10 (OR = 8.28, $p10^{-60}$), yet did not reach the CM7 (CTL-P1 exclusive) and CM8 (Fetal exclusive) level. In addition, compared to CM8, β-FAOX markers were upregulated in CM1 (OR = 128.77, $p < 10^{-60}$), CM2 (OR = 102.75, $p < 10^{-60}$), and in CM10 (OR = 41.68, $p < 10^{-60}$), yet did not reach CM9 level, (**Figure 4C**). β-FAOX upregulation in CM10 was significantly less than in CM1 and CM2. Notably, the CM2 and CM10 clusters comprised most (~90% or more) of the cardiomyocytes in $AR_{P1}MI_{P28}$ hearts during the first week after MI induction, and DAVID analysis indicated that the metabolism of CM2 but not CM10, cardiomyocytes was also supported by increases in insulin (KEGG: Insulin signaling pathway, 2021) and glucagon (KEGG: Insulin signaling pathway, 2021) signaling. Nevertheless, assessments of cell-cycle activity in the CM2 and CM10 clusters were similar (**Figure 3A**), so whether CM2 cardiomyocytes have a unique role in the AR_{P1}-induced enhancement of myocardial repair and regeneration remains unclear.

The Fate of CM1 Cardiomyocytes After Myocardial Infarction on P28 May Have Been Regulated by Protein Kinase AMP-Activated Noncatalytic Subunit Gamma 2, Nuclear Receptor Subfamily 4 Group a Member 3, and Activating Transcription Factor 3

Because cardiomyocytes in the CM1 cluster appeared to retain some latent capacity for proliferation and were present almost exclusively in AR_{P1} hearts at P28, whereas the CM2 and CM10 clusters collectively encompassed most of the cardiomyocytes in $AR_{P1}MI_{P28}$ hearts at P30 and P35 and were more similar to CM1 than other $AR_{P1}MI_{P28}$ clusters (**Supplementary Figure S3**), we used a semisupervised machine-learning technique to investigate whether the induction of MI in hearts that had recovered from previous AR surgery could have triggered the transformation of CM1 cardiomyocytes into CM2 and CM10 cardiomyocytes. snRNA-seq data for each cardiomyocyte in the CM1 cluster were compared to data for CM2 and CM10 cardiomyocytes from $AR_{P1}MI_{P28}$ hearts on P30; then, the CM1 cardiomyocytes were distributed by semisupervised machine-learning into two subpopulations: CM1a or CM1b, which more closely resembled cardiomyocytes from the CM2 or CM10 clusters. (**Figure 5A**). Transcription factors that were differentially expressed between the CM1a and CM1b subpopulations (**Figure 5B**, **Supplementary Table S7**) included protein kinase AMP-activated noncatalytic subunit gamma 2 (PRKAG2), nuclear receptor subfamily 4 group a member 3 (NR4A3), and activating transcription factor 3 (ATF3), which were also more highly expressed in CM2 than in CM10 cardiomyocytes, and the expression of all three genes declined in ARP1MIP28 hearts after P35, which coincided with the disappearance of CM2 cardiomyocytes. Notably, PRKAG2 regulates both glycolysis and fatty acid oxidation (Hinson et al., 2016), which were uniquely coupregulated in CM1 and CM2 cardiomyocytes, and both NR4A3 and ATF3 appear to be cardioprotective: NR4A3 was associated with improvements in infarct size and cardiac function, as well as declines in inflammation when overexpressed in infarcted mouse hearts (Jiang et al., 2019), and measures of cardiac fibrosis, hypertrophy, and function, as well as glucose tolerance, were worse in mice carrying a cardiac-specific ATF3-knockout mutation than in wild type mice when the animals were fed a high-fat diet (Kalfon et al., 2017). There are 38.15% CM1 (**Figure 5C**) cells co-expressing PRKAG2, NR4A3, and ATF3. Then, multiplying the percentage of PRKAG2 + NR4A3 + ATF3 + CM1 cells by the proportion of CM1 cells (38.15%) in

FIGURE 5 | Cardiomyocytes in the CM1 cluster could be partitioned into CM2- and CM10-like subpopulations. **(A)** snRNA-seq data for CM1 cardiomyocytes in AR_{P1}-P28 hearts and for CM2 and CM10 cardiomyocytes in $AR_{P1}MI_{P28}$-P30 hearts were evaluated with a semisupervised learning model to calculate a rescaled score for each CM1 cardiomyocyte. Cardiomyocytes with rescaled scores approaching 0 were designated CM1a (CM2-like) and cardiomyocytes with rescaled scores approaching 1 were designated CM1b (CM10-like). **(B)** The abundance of ATF3, NR4A3, PRKAG2, RGS6, and TMEM126A RNA transcripts was measured in CM1a, CM1b, CM2, and CM10 cardiomyocytes. **(C)** The proportion of PRKAG2+NR4A3+ATF3+ cardiomyocytes in CM1 (AR_{P1}-P28), in comparison with CM2 proportions.
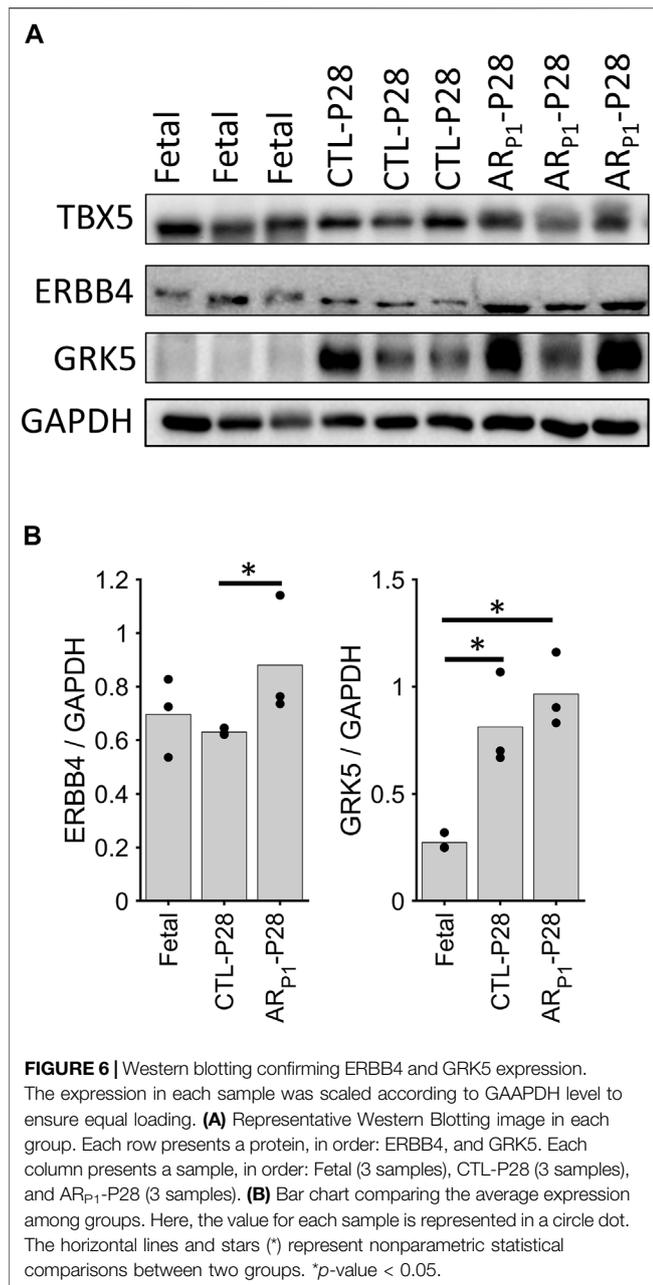
ARP1P28 cardiomyocytes (62.91%) yields 24.00%. This PRKAG2 + NR4A3 + ATF3 + percentage is similar to CM2 percentage in ARP1MIP28P30 and ARP1MIP28P35. Collectively, these observations suggest that CM1 may be composed of the mix-transition states. Also, PRKAG2, NR4A3, and ATF3 may function as molecular switches that trigger the transformation of CM1 cardiomyocytes into CM2 cardiomyocytes. Meanwhile, from the same analysis, transformation from CM1 into CM10 cardiomyocytes may associate with the expressions of RGS6 and TMEM126A, which consistently upregulated in CM1b and CM10.

CM1 Markers ERBB4 and GRK5 Are Highly Expressed in Regenerative Hearts

The roles of TBX5 (Maitra et al., 2009; Misra et al., 2014), TBX20 (Chakraborty and Yutzey, 2012; Xiang et al., 2016), and ERBB4

(Bersell et al., 2009) in cardiomyocyte proliferation had been reported in previous independent studies; therefore, we focused on validating ERBB4 and GRK5, which localize on the cell surface, protein expression in our pig tissue. The western blotting of ERBB4 and GRK5 showed increased protein levels in AR_{P1}-P28, which is consistent with our snRNA-seq data (**Figure 6**). Western blotting measures the protein expression in the whole tissue, including cardiomyocytes and noncardiomyocytes; therefore, we performed further immunofluorescence validation (cardiomyocyte-specific).

We are interested in investigating the specific expression of GRK5 during heart regeneration as existing GRK5 studies (Martini et al., 2008; Islam et al., 2013; Traynham et al., 2016; de Lucia et al., 2022) have not found the role of GRK5 in cardiomyocyte proliferation. Therefore, additional immunofluorescence analysis (**Figure 7**) was performed to

FIGURE 6 | Western blotting confirming ERBB4 and GRK5 expression. The expression in each sample was scaled according to GAAPDH level to ensure equal loading. **(A)** Representative Western Blotting image in each group. Each row presents a protein, in order: ERBB4, and GRK5. Each column presents a sample, in order: Fetal (3 samples), CTL-P28 (3 samples), and AR$_{P1}$-P28 (3 samples). **(B)** Bar chart comparing the average expression among groups. Here, the value for each sample is represented in a circle dot. The horizontal lines and stars (*) represent nonparametric statistical comparisons between two groups. *p-value < 0.05.
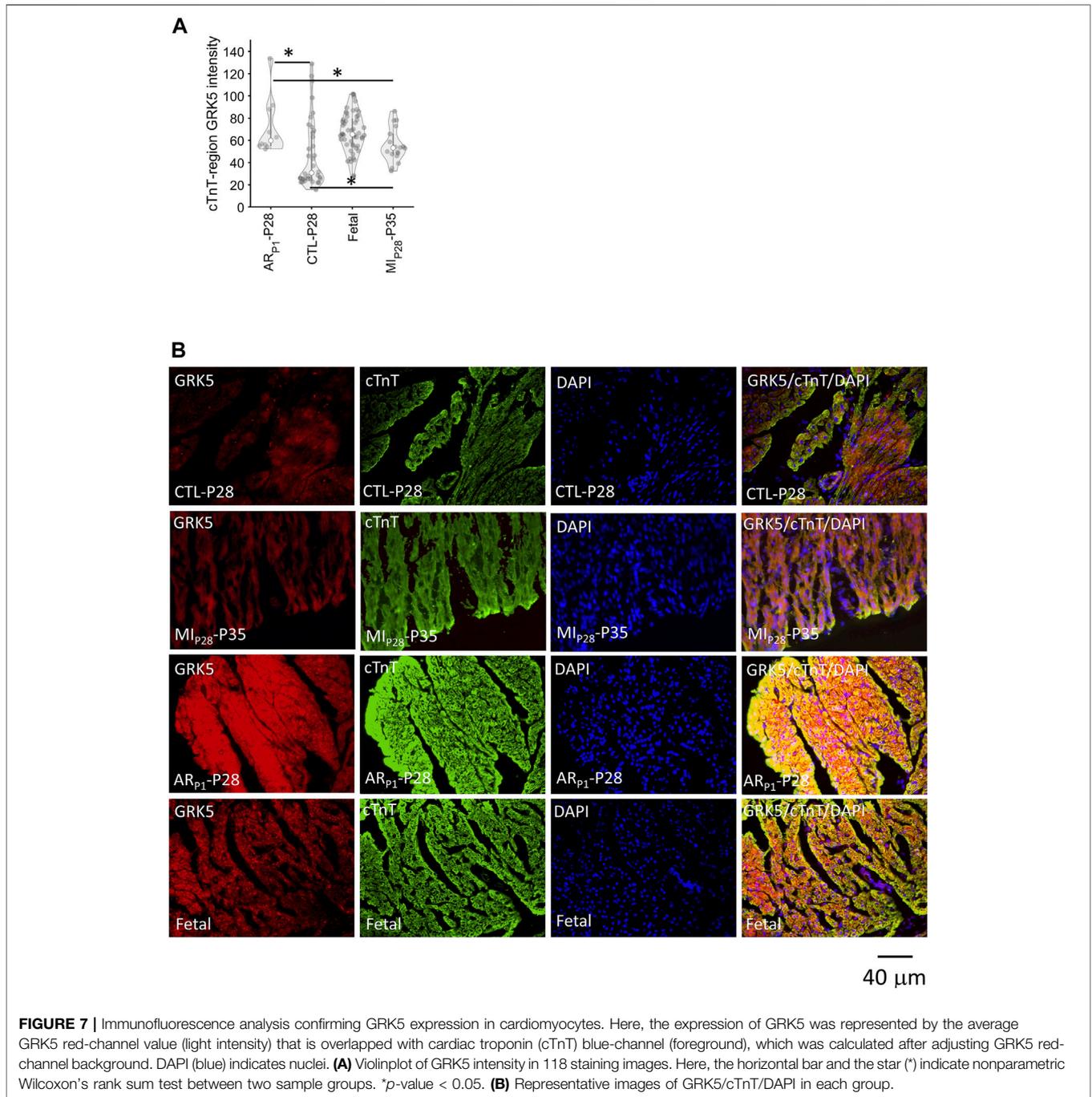
show the GRK5 expression in cardiomyocytes undergoing the regenerative or non-regenerative process. We found that GRK5 expression increases ($p = 0.03$) 7 days after MIP28 injury in the control (CTL-P28 and MI$_{P28}$-P35) group. But the increased GRK5 level in the control group is still less than its level in AR$_{P1}$-P28 samples ($p = 0.03$). This result is consistent with snRNA-seq analysis, where GRK5 expression is the most elevated in CM1 (exclusively for AR$_{P1}$-P28) among the injured-heart cardiomyocytes. Surprisingly, immunofluorescence analysis shows that GRK5 is also highly expressed in the Fetal group, where cardiomyocytes are expected to actively proliferate.

DISCUSSION

Although the proliferative capacity of mammalian cardiomyocytes is extremely limited, the meager amount of myocardial regeneration that occurs after MI in the adult heart appears to require the presence of cycling cardiomyocytes (Laflamme and Murry, 2011). A number of studies have shown that mammalian cardiomyocytes remain proliferative for only a few days after birth (Porrello et al., 2011; Lam and Sadek, 2018; Ye et al., 2018; Zhu et al., 2018) but when AR surgery was performed in pigs on P1, cardiomyocytes proliferated in response to MI injury that occurred on P28, and the animals fully recovered with little or no evidence of contractile dysfunction or myocardial scarring (Zhao et al., 2020). The snRNA-seq analyses presented here build upon these previous observations by showing that the proportion of cycling cardiomyocytes increased in response to either AR$_{P1}$ or MI$_{P28}$, and in animals that underwent both surgical procedures, AR$_{P1}$ appeared to further upregulate cardiomyocyte cell-cycle activity for 1 week after MI induction.

Cluster analysis of our snRNA-seq data indicated that cardiomyocytes in the hearts of Fetal pigs, CTL (uninjured) neonatal pigs, and neonatal pigs that underwent AR$_{P1}$, MI$_{P28}$, or both (AR$_{P1}$MI$_{P28}$) can collectively be grouped into 10 different subpopulations (CM1–CM10), and that the abundance of each cluster varied depending on the injury group and time-point analyzed. Cell-cycle gene expression tended to be highest in CM8, which is consistent with our observation that this cluster comprised >99% of the cardiomyocytes in fetal hearts but was absent in all other groups, and cell cycle activity was greater in CM1–7 and CM10 than in CM9, which was almost exclusively associated with CTL-P56 hearts. Notably, the CM1 cluster comprised >60% of the cardiomyocytes in AR$_{P1}$-P28 hearts but was absent in all other groups, including CTL-P28 hearts, which suggests that CM1 cardiomyocytes may have a prominent role in the AR$_{P1}$-induced enhancement of myocardial regeneration. As in **Figure 1B**, CM1 is very separated from CM8 (representing fetal cardiomyocytes) and CM7 (representing neonatal day 1 cardiomyocyte). Therefore, CM1 is more likely associated with the later postnatal period regeneration rather than the neonatal period. In addition, we hypothesize that CM1 responds to the second myocardial infarction injury on the postnatal day 28 (MI$_{P28}$) by transiting to CM2 or CM10. This response may change the transcription profile of CM1 cardiomyocytes; therefore, the analysis may not represent this "CM1—the same cluster" in AR$_{P1}$-MI$_{P28}$-P30 and AR$_{P1}$-MI$_{P28}$-P30. Therefore, we performed semi-supervised learning to find which regulators might determine the transition from CM1 into CM2 or CM10. Interestingly, proliferative regulators, TBX5 and GRK5 continued expressing highly in CM2. Different from our previous publication (Nakada et al., 2022), which only detected six cardiomyocyte subpopulations, the newly added data, and AI-based pipeline enabled significantly deeper analysis, resulting in ten subpopulations. Beyond reconfirming the PKM2+/NEB + subpopulation as in (Nakada et al., 2022), this work characterized four cardiomyocyte subpopulations in injured-

FIGURE 7 | Immunofluorescence analysis confirming GRK5 expression in cardiomyocytes. Here, the expression of GRK5 was represented by the average GRK5 red-channel value (light intensity) that is overlapped with cardiac troponin (cTnT) blue-channel (foreground), which was calculated after adjusting GRK5 red-channel background. DAPI (blue) indicates nuclei. (A) Violinplot of GRK5 intensity in 118 staining images. Here, the horizontal bar and the star (*) indicate nonparametric Wilcoxon's rank sum test between two sample groups. *p-value < 0.05. (B) Representative images of GRK5/cTnT/DAPI in each group.

heart cardiomyocytes (CM1, CM2, CM10, and CM6). Furthermore, transitions among these subpopulations (CM1, CM2, and CM10) were investigated, which had not been examined in our previous work (Nakada et al., 2022).

Cell-cycle activity was not substantially greater in CM1 cardiomyocytes than in cardiomyocytes from any other cluster except CM9. However, TBX5, TBX20, ERBB4, and GRK5, which have been linked to the proliferation of fetal and neonatal cardiomyocytes [TBX5 (Misra et al., 2014; Maitra et al., 2009), TBX20 (Xiang et al., 2016; Chakraborty and Yutzey, 2012),

ERBB4 (Bersell et al., 2009)] and cardiac hypertrophy [GRK5 (Gold et al., 2012; Traynham et al., 2015)], were upregulated in the CM1 cluster, as well as in Fetal cardiomyocytes. Bioinformatics analysis (Figure 3B) showed that TBX5 and TBX20 transcription factors amplified the expression of positive regulation of cell proliferation, cardiac muscle cell differentiation, canonical Wnt signaling pathway, and MAPK signaling pathway makers. Furthermore, ERBB4 and GRK5, which are surface receptor proteins, may activate ERBB, canonical Wnt, and MAPK signaling pathways at downstream.

Thus, AR on P1 may have primed cardiomyocytes for cell-cycle reactivation by maintaining the signaling mechanisms associated with these four molecules through at least P28. TBX5 and GRK5 were also highly expressed in CM2 cardiomyocytes, which comprised ~20% of cardiomyocytes in AR_{P1}-P28 hearts and a substantially greater proportion of cardiomyocytes in $AR_{P1}MI_{P28}$ than in MI_{P28} hearts during the first week after MI induction. This 7-day window coincided with the period when cycling cardiomyocytes were significantly more common in $AR_{P1}MI_{P28}$ hearts, and although we did not observe differences between cardiomyocyte clusters, injury groups, or time-points in the expression of downstream effectors of these signaling pathways, the results from bulk RNA sequencing assessments suggested that MI induction on P1 upregulated the expression of two molecules identified in our network analysis, MEF2C and NKX2-5, as well as TBX5 (Zhang et al., 2020). CM1 and CM2 cardiomyocytes also displayed substantial evidence of both glycolytic and βFAOX metabolism, while most of the other cardiomyocytes (i.e., CM10 and CM4) in $AR_{P1}MI_{P28}$ and MI_{P28} hearts on P30 and P35 relied primarily on glycolysis for energy production, but whether TBX5 or GRK5 activate βFAOX in neonatal cardiomyocytes, and if so, whether βFAOX upregulation improves the regenerative response to myocardial injury, has yet to be investigated.

The TBX5+/TBX20+/ERBB4+/GRK5+ cardiomyocyte subpopulation (CM1) and its markers were not detected in our previous report (Nakada et al., 2022). This subpopulation was detected in this study by adding new snRNA-seq data and applying the artificial-intelligence techniques. Pipelines to analyze snRNA-seq data include unsupervised clustering (Kiselev et al., 2019), which means the technical result depends on the data being used. Given the same data, it is known that different pipelines can lead to different results (Vieth et al., 2019). Therefore, whether the novel CM1 subpopulation appeared in this study primarily due to adding new data or using a different pipeline is yet to be investigated. From the computational perspective, compared to the Seurat-based pipeline (Hao et al., 2021) used in (Nakada et al., 2022), the artificial-intelligence-based pipeline gives the dimensional reduction advantage. Before clustering, artificial intelligence (autoencoder) reduced the high-dimensional snRNA-seq data into 10 dimensions; meanwhile, Seurat, applying Principal Component Analysis still reduced into 2,000 dimensions. Notably, both of them use the same principle to compute lower-number dimensions: when using the lower-number dimensions to reconstruct the data, the difference between the reconstructed data and original high-dimensional data is minimized. In computing, given the similar optimization criteria, using a lower number of dimensions is better to "curse of dimensionality" issue (Trunk, 1979): the calculation is much less accurate if too many data dimensions are used.

As support of our finding, the roles of TBX5, TBX20, and ERBB4 in promoting cardiomyocyte proliferation had been reported in other studies. In addition, we confirmed the evaluated protein levels of ERBB4 and GRK5 in regenerative heart tissue. Therefore, immunofluorescence analysis was performed to quantify GRK5 expression within cardiomyocytes, where we confirmed the signal intensity of the cTnT region/GRK5 in AR_{P1}-P28 and Fetal groups were higher. Overall, our results suggest that GRK5 may contribute to cardiomyocyte proliferation, whose mechanisms are yet to be further confirmed in future studies.

In conclusion, the results presented here identified a cluster of cardiomyocytes, CM1, in neonatal pig hearts that appeared to be generated in response to AR injury on P1, which in turn, results in CMs cell-cycle activation, and has improved recovery from subsequent AMI on P28. Although CM1 cardiomyocytes did not appear to be substantially more proliferative than cardiomyocytes from other clusters that were present in injured and uninjured hearts through P42, they may have been primed for cell-cycle reactivation by the upregulation of regulatory molecules that contribute to cardiomyocyte proliferation (TBX5, TBX20, and ERBB4) and cardiac hypertrophy (GRK5). Collectively, these observations support future investigations of the roles of these four regulatory molecules in cardiomyocyte proliferation and myocardial repair.

MATERIALS AND METHODS

Animals

All experimental protocols were approved by The Institutional Animal Care and Use Committee (IACUC) of the University of Alabama, Birmingham, and performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH publication No 85-23). Pigs were purchased from Prestage Farm, Inc. (West Point, MS) and cared for as described previously (Zhao et al., 2020). Pigs were housed in an incubator at ~85°F with room air through P14 and then transferred to the normal room and grown under regular pig feeding and ~72°F temperature. Animals were fed bovine colostrum every 4 h for the first 2 days of life, a 1:1 mixture of colostrum: sow's milk on day 3 of life, and sow's milk thereafter. Supplemental iron was provided on day 7.

Aapical Resection and Myocardial Infarction–Induction Surgery

Pigs were anesthetized with isoflurane and placed on a heating pad in a dorsal recumbent position; then, a median sternotomy was performed to expose the heart. AR was performed on P1 by removing 4–5 mm of tissue from the ventricular apex, and MI was induced on P28 by permanently occluding the left-anterior descending coronary artery with a suture; then, the sternum was reapproximated, and the chest was closed in layers, and the air was evacuated from the mediastinum.

Nuclei Isolation

Tissues were cut while submerged in cold phosphate-buffered saline (PBS) or UW solution, washed to remove blood, and transferred into 1 ml lysis buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Nonidet™ P40 Substitute, and 50 U/ml RNase inhibitor in DEPC-treated water), cut into smaller pieces, and aspirated with the lysis buffer into a 50-ml tube; then, 10 ml of lysis buffer was added,

the tissues were ground for 20–30 s, and more lysis buffer was added. The mixture was placed on ice for 10 min, filtered with 100-μm and 70-μm strainers, and centrifuged for 5 min at 700 rcf and 4°C; then, the supernatant was removed, and the pellet was resuspended in 10 ml nuclei wash and resuspension buffer [1 × PBS, 1.0% bovine serum albumin (BSA), and 50 U/ml RNase inhibitor]. The suspension was passed through a 40-μm strainer, and the nuclei were centrifuged again for 5 min at 700 rcf and 4°C; then, the supernatant was removed, the pellet was resuspended in 1 ml nuclei wash, and resuspension buffer and the nuclei were centrifuged a third time for 5 min at 700 rcf and 4°C. The supernatant was removed; then, the pellet was resuspended in 5 ml sucrose cushion buffer I (2.7 ml Nuclei PURE 2M Sucrose Cushion Solution and 300 μl Nuclei PURE Sucrose Cushion Buffer) and mixed with 10 ml sucrose buffer. The mixture was layered over 5 ml of sucrose cushion buffer in a second Eppendorf tube and then centrifuged for 60 min at 13,000 g and 4°C. All but 100 μl of the supernatant was removed, the nuclei were resuspended in nuclei wash and resuspension buffer, and the solution was passed through a 40-μm strainer; then, the nuclei concentration was determined with a cell counter or hemocytometer and adjusted to 1,000 nuclei/μl. The nuclei were placed on ice, stained with propidium iodide for 5 min, and then immediately processed *via* the 10× Genomics® Single-Cell Protocol.

Pre-Processing of snRNA-Seq Data

Sample demultiplexing, barcode processing, and gene counting were performed with Cell Ranger Single-Cell Software v.3.10 (https://support.10xgenomics.com/single-cell-gene expression/software). Reads were aligned to the Sscrofa11.1 pre-mRNA reference genome (Pig, 2021), and the pre-mRNA portion of the reference genome was extracted for single-nuclei UMI mapping with Cell Ranger mkref pre-mRNA. Only confidently mapped reads with valid barcodes and UMIs were used to generate the gene-barcode matrix. Cross-sample integration and quality-control were performed with the Seurat R package. Doublets were identified by using Seurat's standard pipeline (https://satijalab.org/seurat/v3.2/pbmc3k_tutorial.html), and barcodes were removed if they had fewer than 500 UMIs, more than 30,000 UMIs, or >5% mitochondrial UMIs. Nuclei were removed if they had <200 detected genes or if >25% of the transcripts were from mitochondrial genes. Mitochondrial genes and other transcript identifiers were removed without mapping to the official gene symbols from later analyses. Data were normalized as directed in the online Seurat tutorial (https://satijalab.org/seurat/v3.2/pbmc3k_tutorial.html); total expression was multiplied by a factor of 10,000 and then log-transformed. Variations in the number of genes and UMIs detected per cell were accommodated by normalizing the scaled expression matrix *via* the ScaleData function with vars. to. regress set to nUMI and nGenes. Normalization returned two gene-cell matrices: the first was in log scale, and the second was the adjusted gene-cell count.

Autoencoder

An autoencoder (Kramer, 1991) is a deep-neural-network artificial-intelligence technique used to synthesize (Sheikh et al., 2021), denoise (Eraslan et al., 2019), or translate (Cho et al., 2014) data. A data-synthesis autoencoder has at least three layers: an input layer, which consists of the original high-dimensional dataset, a central embedded layer with fewer dimensions (typically 10–20), and a synthetic output layer whose dimensionality is equivalent to the input layer. The autoencoding procedure is performed by encoding the input layer into the embedded layer, decoding the embedded layer into the output layer, and then evaluating the extent of similarity between the input and output layers. When the output layer matches the input layer with maximum fidelity, the embedded layer is considered an accurate low-dimensional representation of the input data. Notably, the immense datasets generated *via* single-cell and single-nucleus genomic analyses can require a prohibitively large amount of computer memory, so a number of recent studies (Wang and Gu, 2018; Eraslan et al., 2019; Tran et al., 2021) have reduced the dimensionality of the input data before the autoencoding procedure is initiated. However, the transcriptional heterogeneity of cardiac cells is exceptionally high (Nahrendorf and Swirski, 2013; Vidal et al., 2019; Tsedeke et al., 2021) and likely to be exacerbated by the physiological changes that occur in response to AR surgery and MI induction. Thus, since any attempt to reduce the dimensionality of the input data before autoencoding could mask this complexity, the input layer used for the studies reported here included the complete list of 14,753 genes (i.e., 14,753 dimensions) with at least 1,000 UMIs in our dataset.

Autoencoding was performed in Matlab (trainAutoencoder, 2021) with self-built models; the architecture of the autoencoder was restricted to three layers, and the embedded layer contained 10 dimensions (**Supplementary Figure S5**). Optimization was achieved by minimizing the function

$$E = \frac{1}{N} \sum_{i}^{N} \sum_{j}^{N} (x_i - y_j)^2 + 0.001 ||W||^2 + Q, \tag{1}$$

where N denotes the number of data points, x_i denotes an arbitrary input data point, y_j denotes an arbitrary synthetic data point, $||W||^2$ represents the regularization of autoencoder weights, and Q represents the sparsity parameters (trainAutoencoder, 2021).

Data Visualization and Clustering

After autoencoding, the embedded data was reduced from 10 to 2 dimensions *via* Uniform Manifold Approximation (UMAP) toolkit (McInnes et al., 2018; Meehan, 2021), and clustering was performed with the UMAP toolkit density-based clustering (dbscan) algorithm (McInnes et al., 2018; Meehan, 2021; Ester et al., 1996; dbscan, 2021); the 30th-distance graph (minpts, or n_neighbors = 30) was plotted, and epsilon (or min_dist) was set to 0.3 (Meehan, 2021). Cardiomyocytes were identified *via* expression of the cardiomyocyte-specific markers ACTC1 and MYH7 (**Supplementary Note S1**); then, the cardiomyocyte data was extracted, and the autoencoding, visualization (UMAP), and clustering (dbscan) procedures were repeated. The assignment of cardiomyocyte clusters (CM1–CM10) was based on visualization and the distribution

of cardiomyocytes in each subgroup (**Supplementary Note S2**), and marker genes for each cluster were identified according to the following criteria: 1) a cluster p-value [Fisher's Exact test (Fisher, 1922)] of less than 10^{-6}, 2) expression by at least 50% of cells in the cluster, and 3) mean abundance at least 1.3-fold greater among cells in the cluster than in the total cardiomyocyte population.

Algorithm Quality Analysis

After being clustered UMAP toolkit (McInnes et al., 2018; Meehan, 2021), the snRNA-seq clusters were validated and manually adjusted (if needed) by the cell-groups' localization in each sample group. In **Supplementary Figure S6**, the cell-landscape in each group revealed specific regions that were significantly enriched by or were exclusive for AR_{P1}-P28, AR_{P1}-MI_{P28}-P30/P35, injured P56, Fetal, CTL-P1, and CTL-P56 cardiomyocytes. There were regions mapped to CM1, CM2, CM6, CM8, CM7, and CM5/CM9, respectively (**Supplementary Figure S7B**). This result explained and justified our clustering parameter settings. This strategy was also used in other single-cell analyses, such as in (Cui et al., 2020).

In addition to the (minpts = 30, epsilon = 0.3) parameter, combination was manually examined according to the instruction in (DBSCAN, 2022). Briefly, for each cell-point, the distances between the point and 30 other nearest points were calculated; then, among these 30 distances, the largest distance was selected as "30th nearest distance." Repeating this process for all cell-points, the list of "30th nearest distances" for all cell-points between 0.1 and 1.2 was obtained, as displayed on the y-axis of **Supplementary Figure S7A**. At any threshold t on the y-axis, the number of cell-points m having "30th nearest distances" < t was counted, and the (m, t) dot was drawn in **Supplementary Figure S7A**. Repeating the (m, t) process as t increased from 0.1 to 1.2 and connecting these (m, t) dots, the "30th nearest distance curve" was completed as shown in **Supplementary Figure S7A**. On "30th nearest distance curve," "the elbow" (m, t) point corresponds to $t \approx 0.3$. This result further justified our decision to use the parameter combination (minpts = 30, epsilon = 0.3).

On the other hand, we manually examined the clustering result when slightly changing minpts = 25, 30, and 35, and epsilon = 0.25, 0.3, and 0.35 with the Matlab UMAP toolkit (Meehan, 2021). Changing these parameters may yield different numbers of clusters. However, using the cell groups localization to manually merge the smaller clusters, we reconstructed a nearly identical cluster landscape to what was obtained using (minpts = 30, epsilon = 0.3) parameter combination. This result justified the robustness of the cluster algorithm and toolkit.

We also justified the criteria by the number of marker genes and their percentage over the entire pig genome, which consists of 25,800 genes. **Table 1** shows that the number of genes passing two criteria (to qualify as cluster markers) is always less than 2% of the genome. Furthermore, we combined the marker genes from all 10 CM clusters, yielding 1,069 marker genes. Then, for the gene, we counted in how many CM clusters such that the gene was a marker. There were 636 genes (59.50%) that were markers of only one CM cluster; 312 genes (29.19%) that were markers of two CM clusters; 98 genes (9.17%) that were markers of three clusters; 21

genes (1.96%) that were markers of four clusters; and only 2 genes (0.19%) that were markers of five clusters. There were no genes that were markers of six or more clusters. Therefore, the criteria to select cluster markers ensured that the markers were very specific for the cluster.

DAVID Functional Annotation

Cluster-specific markers were analyzed with the DAVID functional annotation tool (Huang et al., 2009). Only terms present in the manually annotated Gene Ontology (Huntley et al., 2015), KEGG (Kanehisa et al., 2017), and Reactome (Jassal et al., 2020) categories were selected. To avoid false discoveries caused by multi-hypothesis testing, the selected results were required to have $p < 0.01$ or Benjamini-adjusted (Li and Barber, 2019) $p < 0.05$.

Sparse Modeling

Because the markers selected *via* our DAVID functional analysis were restricted to those present in manually annotated databases, complementary analyses were conducted without the selected markers *via* Sparse modeling. Briefly, for each cell expression data vector **x**, the sparse model estimates a score y from the linear formula

$$y = \mathbf{w}\mathbf{x} + b. \tag{2}$$

Only genes known to be associated with the ontology or pathway being evaluated were considered [e.g., G2M scores were calculated using only genes with known G2M ontology (Mouse Genome Informatics, 2021a)], and higher y implied that the cell was more likely to undergo the corresponding process. Sparse analysis also requires predefined "positive" ($y = 1$) and "negative" ($y = -1$) cell groups (**Table 2**); thus, cell-cycle gene scores (for example) were calculated with Fetal and CTL-P56 cardiomyocytes as the positive and negative groups, respectively, because Fetal cardiomyocytes are highly proliferative, whereas CTL-P56 cardiomyocytes have exited the cell cycle.

The **w** and b (bias) parameters of **Eq. 1** were computed by minimizing

$$\frac{1}{2}|\mathbf{w}| + C\sum_{\forall i}\epsilon_i. \tag{3}$$

Subject to

$$\begin{cases} y_i(\mathbf{w}\mathbf{x}_i + b) + \epsilon_i \geq 1 \\ \epsilon_i \geq 0 \end{cases} \forall i, \tag{4}$$

where i denotes an arbitrary "positive" or "negative" cell and the "margins" of the model are defined by $\mathbf{w}\mathbf{x} + b = 1$ and $\mathbf{w}\mathbf{x} + b = -1$. A good model will have a very high percentage (e.g., 90%) of cells with $\epsilon_i = 0$.

Scores were computed for all cells; then, cells with $y > 1$ (from **Eq. 2**) were categorized "high," cells with $y < -1$ were categorized "low," and cells with $-1 \leq y \leq 1$ were categorized "middle." Thus, a "high" G2M categorization (for example) indicated that the cell was more Fetal- than CTL-P56–like and, consequently, more likely to execute the G2M phase transition. Significance was evaluated by calculating the OR and p-value [Fisher's Exact

TABLE 1 | Number and proportion of genes passing the cluster marker criteria: expressing in more than 50% of the cluster cell and having at least 1.3-fold abundance.

Cardiomyocyte (CM) cluster	Criterion 1: genes expressed in more than 50% of the cluster cell		Criterion 2: expression is 1.3-fold higher (than other clusters)		Criterion 1 and Criterion 2	
	#Genes	%Genome (%)	#Genes	%Genome (%)	#Genes	%Genome (%)
CM1	881	3.40	2,263	8.74	126	0.49
CM2	964	3.73	1,221	4.72	74	0.29
CM3	1,541	5.95	5,610	21.68	419	1.62
CM4	1,435	5.54	9,670	37.36	274	1.06
CM5	1,090	4.21	1,935	7.48	109	0.42
CM6	1,073	4.15	1,209	4.67	89	0.34
CM7	862	3.33	4,438	17.15	102	0.39
CM8	807	3.12	6,544	25.29	206	0.80
CM9	855	3.30	1,277	4.93	241	0.93
CM10	912	3.52	1,391	5.38	8	0.03

TABLE 2 | Positive and negative cell groups for Sparse analysis.

Analysis	Positive cell group	Negative cell group
Cell cycle phases (Cui et al., 2020) (G01, G1S, S, G2M, M, and MG1)	Fetal cardiomyocytes	Naïve-P56 cardiomyocytes
Cardiomyocyte adrenergic signaling (KEGG: Adrenergic signaling in cardiomyocytes, 2021)	CM9*	CM8*
Cardiac muscle contraction (Mouse Genome Informatics, 2021b)	CM9	CM8
Cardiac muscle cell development (Gene Ontology: Cardiac Muscle Cell Development, 2021)	CM8	CM9
Positive regulation of cardiac muscle cell proliferation (Mouse Genome Informatics, 2021c)	CM8	CM9
Positive regulation of cardiac muscle tissue growth (Mouse Genome Informatics, 2021d)	CM8	CM9
Glycolysis (Glycolysis, 2021)	CM8	CM9
Beta fatty acid oxidation (Mouse Genomic Informatics, 2021b)	CM9	CM8

The CM8 and CM9 clusters were exclusively associated with cardiomyocytes from Fetal and CTL-P56 hearts, respectively.

Test (Fisher, 1922)] for the proportion of "high" cells in each group, and $p < 10^{-6}$ was considered significant.

Network Analysis for TBX5, TBX20, ERBB4, and GRK5

Genes targeted by TBX5 and TBX20 were identified in the TRRUST v2 database (Han et al., 2018), and genes that function downstream of ERBB4 and GRK5 were identified in the STRING v11.5 (Szklarczyk et al., 2021) database. Identified genes were combined into a single set and analyzed *via* DAVID (Huang et al., 2009).

Semisupervised Learning Analysis

The semisupervised learning model was built with the Matlab "fitsemiself" function (Matlab, 2021); CM2 cardiomyocytes from $AR_{P1}MI_{P28}$-P30 hearts were used as the class 1 cells, and CM10 cardiomyocytes from $AR_{P1}MI_{P28}$-P30 hearts were used as the class 2 cells. For each CM1 cardiomyocyte from AR_{P1}-P28 hearts, the model calculated a rescaled score between 0 and 1; cells with scores approaching 0 or 1 were categorized an CM1a (i.e., CM2-like) or CM1b (i.e., CM10-like), respectively.

Cardiomyocyte Pairwise Cluster Similarity Analysis

The similarity between two cardiomyocyte clusters, denoted as CM*x* and CM*y*, was calculated by averaging 1,000 similarity scores between 1,000 randomly selected cells in CM*x* and 1,000 randomly selected cells in CM*y*. The cell-cell similarity was computed using the following equation

$$\sqrt{\sum_{j=1}^{m}\left(sign(x_j) - sign(y_j)\right)^2}. \quad (5)$$

Here, *x,y* represents the gene expression of a randomly selected cell in CM*x* and CM*y*, correspondingly, *m* is the number of genes, and *j* represents the gene index. The "sign" function converts the expression into the binary (0 or 1) format, whereas $x(i) = 1$ means the *i*th gene expresses in cell *x*, while $x(i) = 0$ means the *i*th gene does not express in cell *x*.

Bulk-RNA Sequencing Analysis

The regenerative pig heart bulk-RNA sequencing data was obtained from the Gene Expression Omnibus database, accession number GSE144883, and processed according to (Zhang et al., 2020). Briefly, the paired-end fastq files were trimmed using TrimGalore (Krueger, 2019), then mapped to Ensembl Sscrofa11.1 pre-mRNA reference pig genome (Pig, 2021) using STAR v.2.5.2 (Dobin et al., 2013), then the transcripts for each gene (raw expression) were counted using HtSeq v.0.6.1 (Anders et al., 2015). Then, the raw expression matrix was normalized using Deseq2 (Love et al., 2014). Expressions between 3 regenerative hearts, which underwent

myocardial infraction on postnatal day 1, and 3 naïve hearts were compared, and error-bar plotted.

Western Blotting

The Western blotting protocol, which quantified protein expression of ERBB4, GRK5, and GAPDH was completed according to our previous publication (Zhang et al., 2020). Tissues were lysed in PIPA Lysis and Extraction Buffer (Thermo Scientific, 89,901) with Halt™ Protease and phosphatase Single-Use Inhibitor Cocktail (Thermo Scientific, 78,442); then, the lysates were denatured at 95°C for 6°min, separated in a 4%–20% precast gel (Bio-rad, 4568093), and transferred onto a PVDF membrane (Bio-rad, 1704156). The membrane was incubated with 5% nonfat dry milk (Bio-rad, 1706404) for 30 min, with primary antibodies at 4°C overnight, and then with horseradish-peroxidase (HRP)–conjugated secondary antibodies for 30 min. Protein bands were detected with the chemiluminescent HRP substrate (Millipore, WBKLS0500) in a ChemiDoc™ Imaging System (Bio-rad).

We performed western blotting in three Fetal, AR_{P1}-P28, and three CTL-P28 samples. In each sample, protein expression was scaled according to GAPDH to confirm equal loading. Then, statistical comparison and testing were performed using the nonparametric test, according to (Zhao et al., 2021), due to the small sample size. A p-value < 0.05 indicates statistical significance.

Histology

The immunofluorescence analyses were conducted similar to our previous work in Nakada et al. (2022). Hearts were cut into transverse blocks (thickness: 1 cm), and myocardium from the anterior-apical zone (AAZ) were either snap-frozen with liquid nitrogen or processed with 10% formalin fixation and dehydration with 10%- 30% sucrose overnight. Samples were cut into transverse sections (thickness: 10 μm) and stained with antibodies against GRK5 (1:100, rabbit polyclonal, Invitrogen, PA5-106484) and cardiac troponin T (1:50, mouse monoclonal, R&D System, MAB 1874) overnight was followed by blocking in Ultravision Protein Block (Epredia, TA125PBQ) for 7 min. For each group, at least 10 sections of border zone myocardium were analyzed, and a total of 118 images from subendomyocardium, and subepimyocardium were counted. Anti-rabbit and anti-mouse secondary antibodies conjugated to Alexa Fluor 555 and 488 were used for visualization by microscopy. DAPI was used for nuclei staining.

To quantify the GRK5 light intensity at the cardiac troponin T area, the following image processing pipeline was performed (**Supplementary Figure S8**). First, each staining image (including GRK5 and cTnT) was represented in Matlab by the Red-Green-Blue channel matrices. Here, in the red-channel matrix, the number between 0 (totally no red) and 255 (maximum red) represents the red color intensity, so does in the green-channel and blue-channel. Then, image segmentation was performed in the cTnT image green channel: green >10 implied the "foreground" (with cTnT) areas, while green <10 implied the "background" (**Supplementary Figure S8B**). Then, the "segmented" image

was mapped to the GRK5 red channel, where this channel was adjusted by subtracting with the background red channel baseline (**Supplementary Figure S8C**). GRK5 intensity was calculated by the mean of the foreground red-channel (adjusted) number, which was between 0 and 255. Since only the overlap between GRK5 (red) and cTnT (green) areas was used in GRK5 intensity calculation, this approach was better to quantify GRK5 in cardiomyocytes. Then, statistical comparison among the groups was completed using nonparametric Wilcoxon's Rank sum test. A p-value < 0.05 indicates statistical significance.

The background red channel baseline was determined in each GRK5 image based on the distribution (histogram) of the background red channel numbers. Manually investigating all 118 images, we noted three background scenarios. First, when the background numbers follow a power-law distribution, where most of the numbers were around 0 and much fewer nonzero backgrounds, the baseline was set to 0, implying no background adjustment was needed. Second, when the background number appeared to be in a homogeneous distribution, which was either normal, uniform, or power-law, the baseline was set to be the average of the background number. Third, when the background numbers formed two or more distributions, we separated these distributions and set the baseline to be the average of the largest (most right) distribution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE185289.

ETHICS STATEMENT

The animal study was reviewed and all experimental protocol were approved by The Institutional Animal Care and Use Committee (IACUC) of the University of Alabama, Birmingham, and performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH publication No 85-23).

AUTHOR CONTRIBUTIONS

TN, YW, YN, YZ, and JZ conducted the experiments and analyzed the data. TN, YW, YN, YZ, and JZ designed the study and wrote the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2022.914450/full#supplementary-material

Supplementary Figure S1 | Cluster analysis identified six different cell types in hearts from Fetal, CTL, ARP1, MIP28, and ARP1MIP28 animals. snRNA-seq data for all cells from each experimental group and time point were processed via cluster analysis and visualized *via* UMAP. **(A)** Results are displayed for each experimental group and time point. **(B)** Results are displayed for each of six different cell types (cardiomyocytes, fibroblasts, immune cells, smooth muscle cells, endothelial cells, and cardiomyocyte/skeletal-like muscle cells). **(C)** The expression of cell-type–specific genes that were explicitly associated with each cluster is displayed as a heat map (cardiomyocytes: myosin heavy chain 7 [MYH7], Actin Alpha Cardiac Muscle 1 [ACTC1]; fibroblasts: Collagen Type I Alpha 1 [COL1A1], Collagen Type I Alpha 2 [COL1A2]; immune cells: Bridging Integrator 2 [BIN2], gamma-interferon-inducible lysosomal thiol reductase [IFI30]; smooth-muscle cells: Actin Alpha 2 [ACTA2], Myosin Heavy Chain 11 [MYH11]; endothelial cells: Platelet And Endothelial Cell Adhesion Molecule 1 [PECAM1], Kinase Insert Domain Receptor [KDR]; skeletal-muscle cells: Myosin Light Chain Kinase 2 [MYLK2], Nebulin [NEB]).

Supplementary Figure S2 | Upregulation of pluripotent genes in CM4 cardiomyocyte. **(A)** Regulators in KEGG signaling pathways regulating pluripotency of stem cells, including AKT Serine/Threonine Kinase 3 [AKT3], Janus Kinase 1 [JAK1], Janus Kinase 2 [JAK2], Jumonji And AT-Rich Interaction Domain Containing 2 [JARID2], Mitogen-Activated Protein Kinase 1 [MAPK1], Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha [PIK3CA], Signal Transducer And Activator Of Transcription 3 [STAT3], and Zinc Finger Homeobox 3 [ZFHX3]. **(B)** Genes responsible for stem cell population maintenance according to Gene Ontology annotation, including APC Regulator Of WNT Signaling Pathway [APC], AT-Rich Interaction Domain 1A [ARID1A], DEAD-Box Helicase 6 [DDX6], Forkhead Box O1 [FOXO1], Nuclear Receptor Coactivator 3 [NCOA3], Recombination Signal Binding Protein For Immunoglobulin Kappa J Region [RBPJ], SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 2 [SMARCA2], SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin Subfamily C Member 1 [SMARCC1], SS18 Subunit Of BAF Chromatin Remodeling Complex [SS18], and Transcription Factor 7 Like 2 [TCF7L2]. The expression was normalized according to the Seurat pipeline: the expression matrix was scaled by the ScaleData function with vars.to.regress set to nUMI and nGenes; then, the scaled expression was log-normalized.

Supplementary Figure S3 | Pairwise similarity among cardiomyocyte CM1-CM10 clusters. The cluster similarity was calculated by averaging 1000 cell-cell similarity scores, whereas the cells were randomly selected in each cluster. The left-side bar maps the color to the similarity score, where more red implies a higher degree of similarity.

Supplementary Figure S4 | Cardiomyocyte cell-cycle gene expression tended to be greater in ARP1, MIP28, and ARP1MIP28 hearts from P30-P42 than in CTL hearts on P28. Sparse analysis was conducted for the expression of genes associated with the G01, G1S, S, G2M, M, and MG1 phases of the cell cycle, and the proportion of cardiomyocytes with high levels of expression was calculated for each experimental group and time point.

Supplementary Figure S5 | Schematic representation of the autoencoder model. The autoencoder constructed for the analyses presented here consisted of an input layer, a central embedded layer, and an output layer. The input layer included 14,753 genes and was alternately encoded into a 10-dimensional embedded layer and then decoded into a synthetic output layer comprising the same genes present in the input layer. The accuracy of the embedded layer was optimized by minimizing the indicated function.

Supplementary Figure S6 | Cell localization in each sample group's cardiomyocyte. Here, the umap plots for all cells were drawn, and the cells were colored based on: whether the cell belongs to a specific sample group (red) or the cell belongs to other sample groups. The order, from left to right and from top to bottom, is: CTL-P28, ARP1-P28, MIP28-P30, ARP1-MIP28-P30, MIP28-P35, ARP1-MIP28-P35, MIP28-P42, ARP1-MIP28-P42, MIP28-P56, ARP1-MIP28-P56, Fetal, CTL-P1, CTL-P56, and ARP1-P56.

Supplementary Figure S7 | Algorithm-quality analysis confirming that the parameter combination (minpts = 30, epsilon = 0.3) in UMAP toolkit is robust. **(A)** The 30th-nearest-distance plot show that the "elbow" point, which determine the optimal epsilon parameter (according to [92]) is approximately 0.3. **(B)** The preserved cluster landscape draw by slightly changing the parameters: minpts = 25, 30, 35, and epsilon = 0.25, 0.3, 0.35.

Supplementary Figure S8 | Summary of Immunofluorescence image processing. **(A)** The staining images (red: GRK5, green: cardiac troponin cTnT) were represented in red(R)/green(G)/blue(B) channel matrices. **(B)** Image segmentation was performed on the blue channel, which divided the image into the background (no cTnT, black) and foreground (with cTnT, white). **(C)** The red channel background distribution was plotted to determine a background baseline. This distribution had three scenarios. Top: near-zero distribution, the baseline (indicated by the dashed line) = 0; middle: one distribution form, the baseline = average of red channel background; bottom: the two or more distributions appeared, then the baseline = average of the largest (most right) distribution. **(D)** the red channel (GRK5) was adjusted by subtracting this channel from the background baseline.

REFERENCES

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq--a Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics* 31, 166–169. doi:10.1093/bioinformatics/btu638

Angelini, A., Saha, P. K., Jain, A., Jung, S. Y., Mynatt, R. L., Pi, X., et al. (2021). PHDs/CPT1B/VDAC1 axis Regulates Long-Chain Fatty Acid Oxidation in Cardiomyocytes. *Cell Rep.* 37, 109767. doi:10.1016/j.celrep.2021.109767

Bersell, K., Arab, S., Haring, B., and Kühn, B. (2009). Neuregulin1/ErbB4 Signaling Induces Cardiomyocyte Proliferation and Repair of Heart Injury. *Cell* 138, 257–270. doi:10.1016/j.cell.2009.04.060

Bi, J., Bennett, K., Embrechts, M., Breneman, C., and Song, M. (2003). Dimensionality Reduction via Sparse Support Vector Machines. *J. Mach. Learn. Res.* 3, 1229–1243.

Chakraborty, S., and Yutzey, K. E. (2012). Tbx20 Regulation of Cardiac Cell Proliferation and Lineage Specialization during Embryonic and Fetal Development *In Vivo*. *Dev. Biol.* 363, 234–246. doi:10.1016/j.ydbio.2011.12.034

Chen, M., Philipp, M., Wang, J., Premont, R. T., Garrison, T. R., Caron, M. G., et al. (2009). G Protein-Coupled Receptor Kinases Phosphorylate LRP6 in the Wnt Pathway. *J. Biol. Chem.* 284, 35040–35048. doi:10.1074/jbc.m109.047456

Chkifa, A., Cohen, A., and Schwab, C. (2015). Breaking the Curse of Dimensionality in Sparse Polynomial Approximation of Parametric PDEs. *J. de Mathématiques Pures Appliquées* 103, 400–428. doi:10.1016/j.matpur.2014.04.009

Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *arXiv Prepr. arXiv* 1409, 1259.

Clément, S., Chaponnier, C., and Gabbiani, G. (1999). A Subpopulation of Cardiomyocytes Expressing Alpha-Skeletal Actin Is Identified by a Specific Polyclonal Antibody. *Circ. Res.* 85, e51–8. doi:10.1161/01.res.85.10.e51

Cui, M., Wang, Z., Chen, K., Shah, A. M., Tan, W., Duan, L., et al. (2020). Dynamic Transcriptional Responses to Injury of Regenerative and Non-regenerative Cardiomyocytes Revealed by Single-Nucleus RNA Sequencing. *Dev. Cell* 53, 102–116. doi:10.1016/j.devcel.2020.02.019

dbscan (2021). *Dbscan*. Natick, Massachusetts: MathWorks, Inc. https://www.mathworks.com/help/stats/dbscan.html

DBSCAN (2022). *Introduction to DBSCAN*. Natick, Massachusetts: Mathworks, Inc. https://www.mathworks.com/help/stats/dbscan-clustering.html

de Lucia, C., Grisanti, L. A., Borghetti, G., Piedepalumbo, M., Ibetti, J., Lucchese, A. M., et al. (2022). G Protein-Coupled Receptor Kinase 5 (GRK5) Contributes to

Impaired Cardiac Function and Immune Cell Recruitment in Post-ischemic Heart Failure. *Cardiovasc Res.* 118, 169–183. doi:10.1093/cvr/cvab044

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S., and Theis, F. J. (2019). Single-cell RNA-Seq Denoising Using a Deep Count Autoencoder. *Nat. Commun.* 10, 390. doi:10.1038/s41467-018-07931-2

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD'96. *Proc. Second Int. Conf. Knowl. Discov. Data Min.*, 226–231.

Fan, Y., Ho, B. X., Pang, J. K. S., Pek, N. M. Q., Hor, J. H., Ng, S.-Y., et al. (2018). Wnt/β-catenin-mediated Signaling Re-activates Proliferation of Matured Cardiomyocytes. *Stem Cell Res. Ther.* 9, 338. doi:10.1186/s13287-018-1086-8

Fisher, R. A. (1922). On the Interpretation of χ 2 from Contingency Tables, and the Calculation of P. *J. R. Stat. Soc.* 85, 87–94. doi:10.2307/2340521

Gene Ontology: Cardiac Muscle Cell Development (2021). *Gene Ontology: Cardiac Muscle Cell Development.* http://www.informatics.jax.org/vocab/gene_ontology/GO:0055013

Glycolysis.(2021). *Gluconeogenesis - Sus scrofa Pig.* https://www.genome.jp/kegg-bin/show_pathway?ssc00010

Gold, J. I., Gao, E., Shang, X., Premont, R. T., and Koch, W. J. (2012). Determining the Absolute Requirement of G Protein-Coupled Receptor Kinase 5 for Pathological Cardiac Hypertrophy. *Circ. Res.* 111, 1048–1053. doi:10.1161/circresaha.112.273367

Guo, Y., and Pu, W. T. (2020). Cardiomyocyte Maturation. *Circ. Res.* 126, 1086–1106. doi:10.1161/circresaha.119.315862

Hailstones, D. L., and Gunning, P. W. (1990). Characterization of Human Myosin Light Chains 1sa and 3nm: Implications for Isoform Evolution and Function. *Mol. Cell Biol.* 10, 1095–1104. doi:10.1128/mcb.10.3.1095-1104.1990

Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., et al. (2018). TRRUST V2: an Expanded Reference Database of Human and Mouse Transcriptional Regulatory Interactions. *Nucleic Acids Res.* 46, D380–D386. doi:10.1093/nar/gkx1013

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., 3rd, Zheng, S., Butler, A., et al. (2021). Integrated Analysis of Multimodal Single-Cell Data. *Cell* 184, 3573–3587. doi:10.1016/j.cell.2021.04.048

Hinson, J. T., Chopra, A., Lowe, A., Sheng, C. C., Gupta, R. M., Kuppusamy, R., et al. (2016). Integrative Analysis of PRKAG2 Cardiomyopathy iPS and Microtissue Models Identifies AMPK as a Regulator of Metabolism, Survival, and Fibrosis. *Cell Rep.* 17, 3292–3304. doi:10.1016/j.celrep.2016.11.066

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4, 44–57. doi:10.1038/nprot.2008.211

Huang, K., Zheng, D., Sun, J., Hotta, Y., Fujimoto, K., and Naoi, S. (2010). Sparse Learning for Support Vector Classification. *Pattern Recognit. Lett.* 31, 1944–1951. doi:10.1016/j.patrec.2010.06.017

Huntley, R. P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M. J., et al. (2015). The GOA Database: Gene Ontology Annotation Updates for 2015. *Nucleic Acids Res.* 43, D1057–D1063. doi:10.1093/nar/gku1113

Islam, K. N., Bae, J.-W., Gao, E., and Koch, W. J. (2013). Regulation of Nuclear Factor κB (NF-Kb) in the Nucleus of Cardiomyocytes by G Protein-Coupled Receptor Kinase 5 (GRK5). *J. Biol. Chem.* 288, 35683–35689. doi:10.1074/jbc.m113.529347

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 48, D498–D503. doi:10.1093/nar/gkz1031

Jeong, H.-C., Park, S.-J., Choi, J.-J., Go, Y.-H., Hong, S.-K., Kwon, O.-S., et al. (2017). PRMT8 Controls the Pluripotency and Mesodermal Fate of Human Embryonic Stem Cells by Enhancing the PI3K/AKT/SOX2 Axis. *Stem Cells* 35, 2037–2049. doi:10.1002/stem.2642

Jiang, Y., Feng, Y.-P., Tang, L.-X., Yan, Y.-L., and Bai, J.-W. (2019). The Protective Role of NR4A3 in Acute Myocardial Infarction by Suppressing Inflammatory Responses via JAK2-Stat3/nf-Kb Pathway. *Biochem. Biophysical Res. Commun.* 517, 697–702. doi:10.1016/j.bbrc.2019.07.116

Kalfon, R., Koren, L., Aviram, S., Schwartz, O., Hai, T., and Aronheim, A. (2017). ATF3 Expression in Cardiomyocytes Preserves Homeostasis in the Heart and Controls Peripheral Glucose Tolerance. *Cardiovasc Res.* 113, 134–146. doi:10.1093/cvr/cvw228

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45, D353–D361. doi:10.1093/nar/gkw1092

KEGG: Adrenergic signaling in cardiomyocytes (2021). *KEGG: Adrenergic Signaling in Cardiomyocytes.* https://www.genome.jp/entry/pathway+hsa04261

KEGG: Insulin signaling pathway (2021). *KEGG: Insulin Signaling Pathway.* https://www.genome.jp/entry/pathway+hsa04910

Kielman, M. F., Rindapää, M., Gaspar, C., van Poppel, N., Breukel, C., van Leeuwen, S., et al. (2002). Apc Modulates Embryonic Stem-Cell Differentiation by Controlling the Dosage of β-catenin Signaling. *Nat. Genet.* 32, 594–605. doi:10.1038/ng1045

Kiselev, V. Y., Andrews, T. S., and Hemberg, M. (2019). Challenges in Unsupervised Clustering of Single-Cell RNA-Seq Data. *Nat. Rev. Genet.* 20, 273–282. doi:10.1038/s41576-018-0088-9

Kramer, M. A. (1991). Nonlinear Principal Component Analysis Using Autoassociative Neural Networks. *AIChE J.* 37, 233–243. doi:10.1002/aic.690370209

Krueger, F. (2019). *Trim Galore.* Cambridge, United Kingdom: Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Kwon, C., Arnold, J., Hsiao, E. C., Taketo, M. M., Conklin, B. R., and Srivastava, D. (2007). Canonical Wnt Signaling Is a Positive Regulator of Mammalian Cardiac Progenitors. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10894–10899. doi:10.1073/pnas.0704044104

Laflamme, M. A., and Murry, C. E. (2011). Heart Regeneration. *Nature* 473, 326–335. doi:10.1038/nature10147

Lam, N. T., and Sadek, H. A. (2018). Neonatal Heart Regeneration. *Circulation* 138, 412–423. doi:10.1161/circulationaha.118.033648

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H. F., et al. (2010). Jarid2 Is a PRC2 Component in Embryonic Stem Cells Required for Multi-Lineage Differentiation and Recruitment of PRC1 and RNA Polymerase II to Developmental Regulators. *Nat. Cell Biol.* 12, 618–624. doi:10.1038/ncb2065

Li, A., and Barber, R. F. (2019). Multiple Testing with the Structure-adaptive Benjamini-Hochberg Algorithm. *J. R. Stat. Soc. B* 81, 45–74. doi:10.1111/rssb.12298

Liu, P., and Zhong, T. P. (2017). MAPK/ERK Signalling Is Required for Zebrafish Cardiac Regeneration. *Biotechnol. Lett.* 39, 1069–1077. doi:10.1007/s10529-017-2327-0

Lopaschuk, G. D., and Jaswal, J. S. (2010). Energy Metabolic Phenotype of the Cardiomyocyte during Development, Differentiation, and Postnatal Maturation. *J. Cardiovasc Pharmacol.* 56, 130–140. doi:10.1097/fjc.0b013e3181e74a14

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8

Lu, C.-W., Yabuuchi, A., Chen, L., Viswanathan, S., Kim, K., and Daley, G. Q. (2008). Ras-MAPK Signaling Promotes Trophectoderm Formation from Embryonic Stem Cells and Mouse Embryos. *Nat. Genet.* 40, 921–926. doi:10.1038/ng.173

Maitra, M., Schluterman, M. K., Nichols, H. A., Richardson, J. A., Lo, C. W., Srivastava, D., et al. (2009). Interaction of Gata4 and Gata6 with Tbx5 Is Critical for Normal Cardiac Development. *Dev. Biol.* 326, 368–377. doi:10.1016/j.ydbio.2008.11.004

Martini, J. S., Raake, P., Vinge, L. E., DeGeorge, B. R., Jr., Chuprun, J. K., Harris, D. M., et al. (2008). Uncovering G Protein-Coupled Receptor Kinase-5 as a Histone Deacetylase Kinase in the Nucleus of Cardiomyocytes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12457–12462. doi:10.1073/pnas.0803153105

Materna, S. C., Sinha, T., Barnes, R. M., Lammerts van Bueren, K., and Black, B. L. (2019). Cardiovascular Development and Survival Require Mef2c Function in the Myocardial but Not the Endothelial Lineage. *Dev. Biol.* 445, 170–177. doi:10.1016/j.ydbio.2018.12.002

Matlab (2021). *Fitsemiself.* Natick, Massachusetts: Mathworks, Inc. https://www.mathworks.com/help/stats/fitsemiself.html

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv Prepr. arXiv* 1802, 03426.

Meehan, S. (2021). Uniform Manifold Approximation and Projection (UMAP). *Comput. Biol.* 6.

Miklas, J. W., Clark, E., Levy, S., Detraux, D., Leonard, A., Beussman, K., et al. (2019). TFPa/HADHA Is Required for Fatty Acid Beta-Oxidation and Cardiolipin Re-modeling in Human Cardiomyocytes. *Nat. Commun.* 10, 4671. doi:10.1038/s41467-019-12482-1

Misra, C., Chang, S.-W., Basu, M., Huang, N., and Garg, V. (2014). Disruption of Myocardial Gata4 and Tbx5 Results in Defects in Cardiomyocyte Proliferation and Atrioventricular Septation. *Hum. Mol. Genet.* 23, 5025–5035. doi:10.1093/hmg/ddu215

Mouse Genome Informatics (2021). *Gene Ontolog Annotations: G2/M Transition of Mitotic Cell Cycles.*

Mouse Genome Informatics (2021). *Gene Ontology Annotations: Cardiac Muscle Contraction.*

Mouse Genome Informatics (2021). *Gene Ontology Annotations: Positive Regulation of Cardiac Muscle Cell Proliferation*.

Mouse Genome Informatics (2021). *Gene Ontology Annotations: Positive Regulation of Cardiac Muscle Tissue Growth*.

Mouse Genome Informatics (2022). *GO:0019827 Stem Cell Population Maintenance*.

Mouse Genomic Informatics (2021). *Gene Ontology Annotations: Positive Regulation of Cell Population Proliferation*.

Mouse Genomic Informatics (2021). *Gene Ontology: Annotations: Fatty Acid Beta-Oxidation*.

Nahrendorf, M., and Swirski, F. K. (2013). Monocyte and Macrophage Heterogeneity in the Heart. *Circ. Res.* 112, 1624–1633. doi:10.1161/circresaha.113.300890

Nakada, Y., Zhou, Y., Gong, W., Zhang, E. Y., Skie, E., Nguyen, T., et al. (2022). Single Nucleus Transcriptomics: Apical Resection in Newborn Pigs Extends the Time-Window of Cardiomyocyte Proliferation and Myocardial Regeneration. *Circulation* 145 (23), 1744–1747.

Paik, D. T., Tian, L., Lee, J., Sayed, N., Chen, I. Y., Rhee, S., et al. (2018). Large-Scale Single-Cell RNA-Seq Reveals Molecular Signatures of Heterogeneous Populations of Human Induced Pluripotent Stem Cell-Derived Endothelial Cells. *Circ. Res.* 123, 443–450. doi:10.1161/circresaha.118.312913

Pig (Sscrofa11.1) (2021). *Pig assembly and gene annotation*. Ensembl.

Porrello, E. R., Mahmoud, A. I., Simpson, E., Hill, J. A., Richardson, J. A., Olson, E. N., et al. (2011). Transient Regenerative Potential of the Neonatal Mouse Heart. *Science* 331, 1078–1080. doi:10.1126/science.1200708

Rochais, F., Sturny, R., Chao, C.-M., Mesbah, K., Bennett, M., Mohun, T. J., et al. (2014). FGF10 Promotes Regional Foetal Cardiomyocyte Proliferation and Adult Cardiomyocyte Cell-Cycle Re-entry. *Cardiovasc Res.* 104, 432–442. doi:10.1093/cvr/cvu232

Sekine, Y., Yamamoto, K., Kurata, M., Honda, A., Onishi, I., Kinowaki, Y., et al. (2021). HADHB, a Fatty Acid Beta-Oxidation Enzyme, Is a Potential Prognostic Predictor in Malignant Lymphoma. *Pathology* 54, 286–293.

Sheikh, T. S., Khan, A., Fahim, M., and Ahmad, M. (2021). *International Conference on Analysis of Images, Social Networks and Texts*. Springer, 270–281.

Stull, J. T., Kamm, K. E., and Vandenboom, R. (2011). Myosin Light Chain Kinase and the Role of Myosin Light Chain Phosphorylation in Skeletal Muscle. *Archives Biochem. Biophysics* 510, 120–128. doi:10.1016/j.abb.2011.01.017

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/measurement Sets. *Nucleic Acids Res.* 49, D605–D612. doi:10.1093/nar/gkaa1074

trainAutoencoder (2021). *trainAutoencoder*. Natick, Massachusetts: Mathworks, Inc.

Tran, D., Nguyen, H., Tran, B., La Vecchia, C., Luu, H. N., and Nguyen, T. (2021). Fast and Precise Single-Cell Data Analysis Using a Hierarchical Autoencoder. *Nat. Commun.* 12, 1029. doi:10.1038/s41467-021-21312-2

Traynham, C. J., Cannavo, A., Zhou, Y., Vouga, A. G., Woodall, B. P., Hullmann, J., et al. (2015). Differential Role of G Protein-Coupled Receptor Kinase 5 in Physiological versus Pathological Cardiac Hypertrophy. *Circ. Res.* 117, 1001–1012. doi:10.1161/circresaha.115.306961

Traynham, C. J., Hullmann, J., and Koch, W. J. (2016). "Canonical and Non-canonical Actions of GRK5 in the Heart". *J. Mol. Cell. Cardiol.* 92, 196–202. doi:10.1016/j.yjmcc.2016.01.027

Trunk, G. V. (1979). A Problem of Dimensionality: A Simple Example. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-1, 306–307. doi:10.1109/tpami.1979.4766926

Tsedeke, A. T., Allanki, S., Gentile, A., Jimenez-Amilburu, V., Rasouli, S. J., Guenther, S., et al. (2021). Cardiomyocyte Heterogeneity during Zebrafish Development and Regeneration. *Dev. Biol.* 476, 259–271. doi:10.1016/j.ydbio.2021.03.014

Uosaki, H., Cahan, P., Lee, D. I., Wang, S., Miyamoto, M., Fernandez, L., et al. (2015). Transcriptional Landscape of Cardiomyocyte Maturation. *Cell Rep.* 13, 1705–1716. doi:10.1016/j.celrep.2015.10.032

van Roermund, C. W. T., Visser, W. F., Ijlst, L., Waterham, H. R., and Wanders, R. J. A. (2011). Differential Substrate Specificities of Human ABCD1 and ABCD2 in Peroxisomal Fatty Acid β-oxidation. *Biochimica Biophysica Acta (BBA) - Mol. Cell Biol. Lipids* 1811, 148–152. doi:10.1016/j.bbalip.2010.11.010

Vidal, R., Wagner, J. U. G., Braeuning, C., Fischer, C., Patrick, R., Tombor, L., et al. (2019). Transcriptional Heterogeneity of Fibroblasts Is a Hallmark of the Aging Heart. *JCI Insight* 4. doi:10.1172/jci.insight.131092

Vieth, B., Parekh, S., Ziegenhain, C., Enard, W., and Hellmann, I. (2019). A Systematic Evaluation of Single Cell RNA-Seq Analysis Pipelines. *Nat. Commun.* 10, 4667. doi:10.1038/s41467-019-12266-7

Vincentz, J. W., Barnes, R. M., Firulli, B. A., Conway, S. J., and Firulli, A. B. (2008). Cooperative Interaction ofNkx2.5andMef2ctranscription Factors during Heart Development. *Dev. Dyn.* 237, 3809–3819. doi:10.1002/dvdy.21803

Wang, D., and Gu, J. (2018). VASC: Dimension Reduction and Visualization of Single-Cell RNA-Seq Data by Deep Variational Autoencoder. *Genomics, Proteomics Bioinforma.* 16, 320–331. doi:10.1016/j.gpb.2018.08.003

Wang, Y., Yao, H., and Zhao, S. (2016). Auto-encoder Based Dimensionality Reduction. *Neurocomputing* 184, 232–242. doi:10.1016/j.neucom.2015.08.104

Xiang, F.-l., Guo, M., and Yutzey, K. E. (2016). Overexpression of Tbx20 in Adult Cardiomyocytes Promotes Proliferation and Improves Cardiac Function after Myocardial Infarction. *Circulation* 133, 1081–1092. doi:10.1161/circulationaha.115.019357

Yang, X., Deng, C., Zheng, F., Yan, J., and Liu, W. (2019). Deep Spectral Clustering Using Dual Autoencoder Network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 4066-4075. IEEE.

Ye, L., D'Agostino, G., Loo, S. J., Wang, C. X., Su, L. P., Tan, S. H., et al. (2018). Early Regenerative Capacity in the Porcine Heart. *Circulation* 138, 2798–2808. doi:10.1161/circulationaha.117.031542

Yuan, Y., Park, J., Feng, A., Awasthi, P., Wang, Z., Chen, Q., et al. (2020). YAP1/TAZ-TEAD Transcriptional Networks Maintain Skin Homeostasis by Regulating Cell Proliferation and Limiting KLF4 Activity. *Nat. Commun.* 11, 1472. doi:10.1038/s41467-020-15301-0

Zhang, E., Nguyen, T., Zhao, M., Dang, S. D. H., Chen, J. Y., Bian, W., et al. (2020). Identifying the Key Regulators that Promote Cell-Cycle Activity in the Hearts of Early Neonatal Pigs after Myocardial Injury. *PLoS One* 15, e0232963. doi:10.1371/journal.pone.0232963

Zhang, Y., Zhou, G., Jin, J., Zhao, Q., Wang, X., and Cichocki, A. (2014). Aggregation of Sparse Linear Discriminant Analyses for Event-Related Potential Classification in Brain-Computer Interface. *Int. J. Neur. Syst.* 24, 1450003. doi:10.1142/s0129065714500038

Zhao, B., Ye, X., Yu, J., Li, L., Li, W., Li, S., et al. (2008). TEAD Mediates YAP-dependent Gene Induction and Growth Control. *Genes Dev.* 22, 1962–1971. doi:10.1101/gad.1664408

Zhao, M., Nakada, Y., Wei, Y., Bian, W., Chu, Y., Borovjagin, A. V., et al. (2021). Cyclin D2 Overexpression Enhances the Efficacy of Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes for Myocardial Repair in a Swine Model of Myocardial Infarction. *Circulation* 144, 210–228. doi:10.1161/circulationaha.120.049497

Zhao, M., Zhang, E., Wei, Y., Zhou, Y., Walcott, G. P., and Zhang, J. (2020). Apical Resection Prolongs the Cell Cycle Activity and Promotes Myocardial Regeneration after Left Ventricular Injury in Neonatal Pig. *Circulation* 142, 913–916. doi:10.1161/circulationaha.119.044619

Zhu, W.-Z., Xie, Y., Moyes, K. W., Gold, J. D., Askari, B., and Laflamme, M. A. (2010). Neuregulin/ErbB Signaling Regulates Cardiac Subtype Specification in Differentiating Human Embryonic Stem Cells. *Circ. Res.* 107, 776–786. doi:10.1161/circresaha.110.223917

Zhu, W., Zhang, E., Zhao, M., Chong, Z., Fan, C., Tang, Y., et al. (2018). Regenerative Potential of Neonatal Porcine Hearts. *Circulation* 138, 2809–2816. doi:10.1161/circulationaha.118.034886