Single Cell Multi-Omics of an iPSC Model of Human Sinoatrial Node Development Reveals Genetic Determinants of Heart Rate and Arrhythmia Susceptibility

James L. Engel[1,2,3,*], Xianglong Zhang[4,*], Daniel R. Lu[5], Olaia F. Vila[5], Vanessa Arias[5], Jasper Lee[5], Christopher Hale[5], Yi-Hsiang Hsu[4], Chi-Ming Li[5], Roland S. Wu[1], Vasanth Vedantham[2,3,6,#], Yen-Sin Ang[1,†,#]

[1]Amgen Research, Cardiometabolic Disorders, South San Francisco, CA; [2]Cardiology Division, Department of Medicine; [3]Cardiovascular Research Institute University of California, San Francisco, CA; [4]Amgen Research, Center for Research Acceleration by Digital Innovation, South San Francisco, CA and Cambridge, MA; [5]Amgen Research, Research Biomics, South San Francisco, CA and [6]Cardiovascular Genetics Center, University of California, San Francisco, CA

*These authors contributed equally

#Co-corresponding authors

†Current address: Cardiovascular Therapeutic Area, BioMarin Pharmaceutical Inc.

Corresponding Authors:

Yen-Sin Ang, PhD[†]
BioMarin Pharmaceutical Inc.
770 Lindaro Street
San Rafael, CA 94901
yensin.ang@bmrn.com

Vasanth Vedantham, MD, PhD
Smith Cardiovascular Research Building
555 Mission Bay Blvd South, 352M
San Francisco, CA 94158
vasanth.vedantham@ucsf.edu

ABSTRACT

Cellular heterogeneity within the sinoatrial node (SAN) is functionally important but has been difficult to model *in vitro*, presenting a major obstacle to studies of heart rate regulation and arrhythmias. Here we describe a scalable method to derive sinoatrial node pacemaker cardiomyocytes (PCs) from human induced pluripotent stem cells that recapitulates differentiation into distinct PC subtypes, including SAN Head, SAN Tail, transitional zone cells, and sinus venosus myocardium. Single cell (sc) RNA-sequencing, sc-ATAC-sequencing, and trajectory analyses were used to define epigenetic and transcriptomic signatures of each cell type, and to identify novel transcriptional pathways important for PC subtype differentiation. Integration of our multi-omics datasets with genome wide association studies uncovered cell type-specific regulatory elements that associated with heart rate regulation and susceptibility to atrial fibrillation. Taken together, these datasets validate a novel, robust, and realistic *in vitro* platform that will enable deeper mechanistic exploration of human cardiac automaticity and arrhythmia.

INTRODUCTION

Heart rhythm originates in the sinoatrial node (SAN), an array of several thousand specialized pacemaker cardiomyocytes (PCs) and non-myocyte cells located at the junction of the systemic venous inflow with the right atrium[1]. At its cranial end, the SAN head region wraps anteriorly around the superior vena cava, while an SAN tail region extends inferiorly along the inter-caval groove and a transition zone occupies the interface of the SAN and atrial myocardium[2,3]. Head PCs, Tail PCs, and Transition Zone PCs have distinct phenotypes and gene expression programs that allow for optimal SAN impulse generation, impulse transmission to the atrium, and autonomic response[4,5].

For example, during normal sinus rhythm, the site of origin of the cardiac impulse migrates cranially towards the SAN head under high sympathetic tone and shifts caudally into the SAN tail when parasympathetic tone predominates, reflecting greater autonomic sensitivity of SAN Head as compared to SAN Tail. Cells occupying the transition zone between the SAN and the atrial myocardium shield SAN automaticity from electrotonic suppression by atrial myocardium while promoting impulse transmission.

These PC subtypes are distinguishable by differential expression of key transcription factors including TBX18, ISL1, NKX2-5. TBX18 is restricted to the SAN head (where NKX2-5 is absent), while NKX2-5 is expressed at low levels in the SAN tail and at higher levels in transition zone cells and may partly explain the more atrial-like phenotype of the latter cell type. While ISL1 is enriched in the SAN Head, it is not known whether it plays a role in specifying SAN Head versus SAN Tail identity, and whether other factors are involved. In addition, although *Isl1*$^{+}$/*Tbx18*$^{+}$ SAN precursors arise near the right inflow region shortly after murine heart tube formation, the timing and inducing signals for specification of mammalian PCs *in vivo* are not known[6,7,8]. Detailed characterization of the differentiation programs of each human PC subtype has been challenging due to the low cell numbers and inaccessibility of human SAN tissue.

Sinus node dysfunction, which can present with symptomatic bradycardia and failure of heart rate to augment appropriately with activity, is among the most common heart rhythm disorders and is strongly associated with the development of atrial

fibrillation. However, how the different manifestations of SND are related to dysfunction of specific PC subtypes has not been established, nor have the molecular and genetic pathways that drive the association of SND with atrial fibrillation. An *in vitro* system that modeled the development and diversification of SAN precursors would have the potential to shed light on these poorly understood areas.

Human induced pluripotent stem cells (hiPSCs) can be differentiated into various cardiomyocyte subtypes, including atrial[9], ventricular[10], and pacemaker[11-13]. Existing protocols to generate hiPSC-SAN cells use a variety of methods, including selection from mixed populations of hiPSC-derived cardiomyocytes[14], directed differentiation using exposure to morphogens and growth factors[11,15], overexpression of transcription factors[16-18], and directed differentiation using small molecules. Recent work has shown that PC subtypes can be discerned from hiPSC-based differentiations using single cell (sc)-RNA sequencing and clustering analysis[19], raising the possibility of using *in vitro* systems to model SAN development, and to develop datasets that can be integrated with human genetic data to yield insights into SND and atrial fibrillation.

Here we present a novel and scalable monolayer hiPSC SAN differentiation protocol that yielded SAN head, SAN tail, transition zone cells, and sinus venosus myocardium. scRNA-seq and sc-ATAC-seq was performed on over 180,000 cells from these differentiations, allowing for detailed characterization of each of these PC subtypes and to perform differentiation trajectory analysis, providing new insights into the epigenetic and transcriptional determinants of PC subtype identity. Integration of multi-omics datasets with existing genome wide association studies allowed us to establish specific associations of discrete cell types with clinical phenotypes and to improve the power of existing genome wide association studies by identifying several novel genes associated with atrial fibrillation and heart rate regulation.

RESULTS

Differentiation of Sinoatrial Node-Like Cardiomyocytes for Multi-Omics Profiling.

To differentiate Sinoatrial Node-Like Cardiomyocytes (SANCMs), we modified established monolayer differentiation protocols for ventricular cardiomyocytes (VCMs) and atrial cardiomyocytes (ACMs) [20-22] by incorporating key aspects of an embryoid

body-based SANCM derivation protocol including supplementation with BMP4, inhibition of TGFβ and FGF, and lactate selection[11,23] (Fig. 1a). Multielectrode array recording at day 9 revealed SANCMs beating at 2.8 +/- 0.1 Hz, ACMs at 2.1 +/- 0.1 Hz, and VCMs at 1.3 +/- 0.1 Hz (n = 4 wells for each condition, Extended Data Fig. 1a). qPCR on RNA isolated from three independent differentiations showed that while all CMs expressed *TNNT2*, *NKX2-5*, *ACTN2*, *MYH6, and SCN5A* (Fig.1b and Extended Data Fig. 1b), expression levels of chamber-specific markers differed: VCMs were enriched for *MYL2*, *IRX4, MYH7* and *HEY2* (Fig. 1c and Extended Data Fig. 1c), while ACMs were enriched for *NPPA*, *MYL7*, *NR2F2* and *KCNA5* (Fig. 1d and Extended Data Fig. 1d). Compared to VCMs, SANCMs were enriched for PC genes *TBX18*, *SHOX2*, *ISL1, HCN1*, *HCN4*, and *KCNJ3* (Fig. 1e,f), while *NKX2-5* showed a non-significant reduction in the SANCMs. When compared to ACMs, SANCMs were significantly enriched for both *TBX18* and *SHOX2* but not for the other SAN marker genes, likely reflecting the presence of some PC-like cells in the ACM differentiation[9,24]

Next, quantitative immunocytochemistry for *TNNT2*, *NKX2-5*, *SHOX2*, *ISL1*, and *HCN4* was performed in randomly selected high-power fields from 2 replicates each of VCM and SANCM differentiations (Fig. 1g and Extended Data Fig 1e). As expected, more than 95% of cells measured in both differentiations were *TNNT2*+, (Extended Fig 1f). For SANCMs, 65% were SHOX2+, 75% were ISL1+ and 79% were HCN4+ as compared to 4%, 8% and 23% of VCMs, respectively ($p < 0.0001$ for each gene; Fig. 1h and Extended Data Fig 1e,f). For NKX2-5, 34% of SANCMs were positive while 85% of VCMs were positive and the average signal intensity per positive nucleus was reduced by 50% in SANCMs, a degree of downregulation that comports well with findings from RNA sequencing datasets of SAN from animal models (Fig. 1h and Extended Data Fig 1f).

When action potential waveforms from spontaneously beating VCM and SANCM cells were recorded using a voltage-sensitive dye and compared (n = 399 for VCMs and n = 169 for SANCMs), VCMs exhibited a steeper phase 0 and prolonged phase 2, while SANCMs exhibited more rapid diastolic depolarization, shorter cycle length, shorter APD, and smaller AP amplitude (Fig. 1i and Extended Data Fig. 1g). SANCMs were also sensitive to application of Zatebradine, a specific blocker of HCN channels (Fig. 1j).

Taken together, these results demonstrate that our scalable SANCM differentiation protocol yields large numbers of PC-like cardiomyocytes suitable for multi-omics profiling.

Single-Cell Transcriptomics Reveals Distinct Human Atrial, Ventricular, and SAN Populations *In Vitro*.

Next, we performed single cell RNA sequencing (scRNA-seq) on day 34, purified VCMs, ACMs, and SANCMs from 3 independent sets of differentiations (n = 21,000 – 22,000 cells per condition, Extended Data Fig. 2a). Unsupervised clustering and dimension reduction was carried out for all cells and plotted in the same UMAP space. Cells derived from VCM, ACM, and SANCM directed differentiations occupied distinct regions in the UMAP space with minimal overlap, indicating unique expression signatures (Fig. 2a). We identified a total of 26 clusters, with the largest number of cells representing *TNNT2*$^{+}$*, ACTN2*$^{+}$*,* and *MYH6*$^{+}$ cardiomyocytes (Fig. 2b,c). Using a cell composition cut-off of 70%, 8, 3 and 6 clusters originated from the ACM, SANCM, and VCM differentiation, respectively, with the remaining clusters representing cells derived from multiple differentiations (Extended Data Fig. 2b).

Individual clusters were assigned to cardiomyocyte subtypes by intersection their marker genes with established expression patterns (Fig. 2d,e and Extended Data Fig. 2c). Clusters 2, 7, 9, 15 represent VCMs (*MYL2, MYH7, HEY2, IRX4,* and *HOPX*), clusters 0, 6, 8 and 12 represent ACMs (*MYL7, NPPA, NR2F2. KCNJ3,* and *KCNA5*) and cluster 1 represents PCs (*SHOX2, ISL1, TBX18,* and *TBX3*). We also identified several non-cardiomyocyte clusters (C10, C13, C14, C17, C19, C20, C21, C23, C24 and C25). that included proepicardial cells (C25) marked by *WT1, ALDH1A2, BNC1,* and *PDPN*, endocardial cells (C13, C19) marked by *FOXC1, MSX1, NRG1, NPR3* and *NFATC1*, primitive mesendoderm marked by *FOXA2, EOMES, MESP1* and *MESP2* (C17, C21), and visceral neurons marked by *PHOX2B, PRPH,* and *CHRNA3* (C24) (Fig. 2b,c and Extended Data Fig. 2d-h)

Differential expression analysis between SANCMs and ACMs yielded 506 genes that recapitulated well-established differences between atrial and PCs[25] (Fig. 2f). GO analysis of these differentially expressed genes (DEGs) demonstrated enrichment of

biological processes related to sinoatrial node development, cardiac conduction, and regulation of heart contraction (Fig. 2g). Intersecting the list of DEGs with neonatal mouse[26] and human fetal[27] SAN versus right atrium comparative transcriptomic datasets demonstrated significant overlap among the three datasets (Extended Data Fig. 2i p = 0.02 for overlap between SANCMs and human dataset and $p = 6.4 \times 10^{-7}$ for overlap between SANCMs and mouse dataset).

Sub-clustering of SAN Differentiation Identifies Specialized PC Subtypes.

The functional organization of the SAN, diagrammed in Fig. 3a, includes an autonomically responsive SAN head located cranially, an SAN tail extending inferiorly along the venous valves, and a transition zone that facilitates source-sink matching between SAN and atrium[3]. Different PC subtypes express different levels of key transcriptional regulators that guide their distinctive phenotypes[28] (Fig. 3b).

To test whether discrete PC subtypes could be identified within our SANCM population, we further sub-clustered on SANCMs, yielding 11 subclusters (Fig. 3c). Cluster 2 resembled SAN head cells (*TBX18*^{HI}/SHOX2^{+}/ISL1^{HI}/*TBX3*$^{+}$/*NKX2-5*^{LOW}), clusters 0 and 3 resembled SAN tail cells (TBX18^{-}/SHOX2^{+}/ISL1^{LOW}/TBX3^{+}/NKX2-5^{+}), clusters 1 and 4 were transitional cells (*TBX18*$^{-}$/SHOX2^{+}/ISL1^{-}/*TBX3*^{LOW}/*NKX2-5*^{HI}), and clusters 6, 7, and 8 represented sinus venosus myocardium (*TBX18*$^{+}$/SHOX2^{+}/ISL1^{LOW}/*TBX3*$^{-}$/*NKX2-5*^{LOW}). The two SAN tail cell clusters were distinguished by differential expression of the SAN transcriptional repressor *TBX3* and its downstream target genes such as *NPPA*, possibly reflecting residual heterogeneity within these populations or varying degrees of maturation (Fig. 3c). Notably, cluster 5 most resembled sinus venosus myocardium but was lower in *TBX18* expression than other sinus venosus clusters. We also identified small populations of proepicardial (cluster 9) and visceral neurons (cluster 10) (Fig. 3d).

Direct comparison identified several hundred DEGs between SAN head (C2) and SAN tail (C0+C3) populations, encompassing GO terms for numerous biological processes related to cellular morphogenesis, cytoskeletal organization, cell adhesion, and neuron protrusion/axon development (Fig. 3e and Extended Data Table 1). SAN Head and Tail cells had contrasting expression of *TBX18* and *NKX2-5,* a uniformly high

expression of *SHOX2* throughout and a gradient of *ISL1* expression from SAN head ($ISL1^{HI}$) to SAN tail ($ISL1^{LOW}$) cells (Fig. 3f,g).

Importantly, 15 out of 85 mouse SAN head vs tail DEGs identified in a recent mouse scRNA-seq study were also differentially expressed between C2 (SAN head-like) and C0+C3 (SAN tail-like), an odds ratio of 7.8 over SAN-expressed genes that were not on the mouse head versus tail DEG list ($p = 2.6 \times 10^{-8}$ by Fisher's exact test, Fig. 3h). The list of shared DEGs between mouse and human included canonical transcription factors *ISL1, TBX18, SHOX2* and *NKX2-5*, supporting the existence of an evolutionarily conserved blueprint not only for SAN development broadly but also for fine patterning of PC subtypes within the entire node. Other SAN Head vs Tail DEGs that were conserved between mouse and human included *GJA1*, an established downstream target of *NKX2-5* (enriched in SAN tail), modulators of signal transduction (*FHL2*, *VSNL1*, *BMP2*), and cytoskeletal organizing proteins (*MARCKS, TMSB4X*). In addition, we also observed conserved enrichment in SAN Head cells for *NF1A*, a transcription factor without a previously known role in SAN development.

To test for concordance with other SAN differentiation methodologies, we projected scRNA-seq data from a recently published hiPSC SAN differentiation[19] into our UMAP and found very close agreement in cluster assignments between the two protocols, such that 94% of SAN head cells mapped to either our SAN head cluster (C2) or to the narrow interface between our SAN head and the nearest SAN tail clusters (C2/C3 interface); 90% of SAN tail cells mapped to our SAN tail clusters (C0+C3); 82% of transitional cells mapped to our transitional clusters (C1+C4), 64% of sinus venosus cells mapped to our sinus venosus clusters (C5+C6+C7+C8), and 97% of proepicardial cells mapped our proepicardial cluster (C9) (Extended Data Fig 3a). Similarly, to test whether primary PCs from adult human hearts were also represented in our culture system, we projected cells from a recent single cell multi-omics study of human SAN tissue[29] into the SANCM UMAP space. In this analysis, 78% of human primary PCs (192/245) were assigned to SAN head or SAN tail, while an additional 18% were assigned to SAN transitional cell clusters. In contrast, only 9% of primary atrial cardiomyocytes were classified as SAN head or SAN tail, of which just 0.002% (18/9595) were assigned SAN head identity (Extended Data Fig 3b). Overall, the

concordance of our SAN head versus tail differential expression analysis with findings from primary human and mouse SAN cells, and with previous hiPSC-based SAN differentiation protocols strongly supports our conclusion that our SANCM differentiation yields specialized venous pole cardiomyocyte cell types that faithfully reflect distinct PC subtypes encountered *in-vivo*.

scATAC-seq Identifies Cell Type Specific Regulatory Regions in SANCMs.

The define cell subtype-specific epigenetic signatures, we performed single cell assay for transposase accessible chromatin with sequencing (scATAC-seq) on cells from 3 independent SANCM differentiations. UMAP projection identified 8 clusters (Fig. 4a and Extended Data Fig. 4a) which we annotated based on predicted expression of canonical SANCM marker genes: $ISL1^{+}/TBX18^{+}/SHOX2^{+}/NKX2.5^{LOW}$ SAN head (C2), $ISL1^{+}/SHOX2^{+}/NKX2.5^{+}/TBX18^{LOW}$ SAN tail (C1, C4, C6), $NKX2.5^{HIGH}/SHOX2^{LOW}/TBX18^{-}$ transitional cells (C0, C5), and TBX18+/NKX2-5- sinus venosus cells (C3) (Fig. 4b). We integrated our scATAC-seq and scRNA-seq datasets of SANCMs by assigning RNA-seq profiles to each scATAC-seq cell using canonical correlation analysis (CCA) and identified peaks highly correlated to RNA expression of proximal genes, which resulted in 51,563 peak-to-gene links (Fig. 4c). This analysis demonstrated concordance between the transcriptional and epigenetic status of each cell subtype.

The strong chromatin-mRNA correlation permitted us to establish distal enhancer-promoter links of up to several hundred kilobases, along with quantitative estimation of interaction strengths. Genomic tracks for accessibility, marker peaks and peak-to-gene links in different SANCM subtypes are shown for *SHOX2*, *ISL1*, *TBX3*, and *TBX18* (Fig. 4d-f and Extended Data Fig. 4b). Regions of differential accessibility between SAN subpopulations were observed at cis-regulatory elements that were reported to activate these genes (Fig. 4d-f and Extended Data Table 2) and numerous cluster-enriched ATAC-seq peaks were identified (Extended Data Fig. 4c). For example, the syntenic region to a mouse Isl1 enhancer previously characterized by our group demonstrated differential accessibility in PCs (SAN head and tail) as compared to more atrial-like transitional cells (Fig. 4d), analogous to ISL1 mRNA levels in these

subpopulations. In addition, our analysis identified other ISL1 enhancers linked to transcription that were previously validated in a zebrafish system [30]. Examination of the *SHOX2* and *TBX3* loci demonstrated linkage of previously defined SAN enhancers to transcription and simultaneously showed differential accessibility across the venous pole cardiomyocyte subtypes. Remarkably, the only peak-to-gene links over a 1-Mb window near *TBX3* identified an enhancer known to be functionally important in PCs. Comparison with a published dataset on a mixed SANCM population revealed that most of the ATAC-seq signal were derived from SAN head cells, with some contribution from SAN tail cells, suggesting that the upstream processes that regulate accessibility of the *TBX3* enhancer are critical for fine-patterning of the SAN (Fig. 4f). Finally, reclustering of the SAN head and SAN tail cells alone demonstrated expected patterns of predicted gene expression for canonical SAN transcriptional regulators, similar to what was observed for transcriptomic datasets (Extended Data Fig. 4d).

Identification of Transcriptional Pathways Important for SAN Subtype Differentiation.

To identify transcriptional pathways that drive cellular subtype-specific differentiation programs, we performed motif enrichment analysis of marker peaks for each cluster (Fig. 4g and Extended Data Table 3). This analysis resulted in association of canonical myocardial transcription factors MEF2, TBX5, and NKX2-5 with the more atrial-like transitional cells, whereas marker peaks for the main SAN Tail cluster (C1) demonstrated enrichment of ISL and GATA binding sites, as well as sites for the TALE-class transcription factors MEIS and TGIF. The SAN Head cluster (C2) exhibited the strongest enrichment for the NFI sites, followed by ISL1, while the sinus venosus cluster (C3) demonstrated strong enrichment for AP-1 sites. These findings were underscored by a direct comparison of motif enrichment between SAN Head and SAN Tail marker peaks, in which NFI sites were strongly enriched in SAN head whereas motifs for myogenic factors NKX2-5 and MEF2 were enriched in SAN tail (Fig. 4h and Extended Data Table 4). Motif enrichment analysis for the subset of ATAC-seq peaks that were positively linked to transcription via peak-to-gene analysis identified the top 20 positive transcriptional regulators across the dataset (Fig. 4i). This analysis identified known

regulators of SAN development ISL1, TBX5, and MEF2 but also identified NFI motifs, which exhibit cellular subtype-specific motif enrichment and differential gene expression (NFIA) in SAN head versus other venous pole cardiomyocytes.

Time Course scRNA-seq Analysis of hiPSC-Derived SANCM Recapitulates *In vivo* Developmental Transitions

Next, we performed scRNA-seq on cells from D3, D4, D6, D10, and D23 of SANCM differentiation and analyzed them with D34 SANCMs (Extended Data Fig. 5a). Unsupervised clustering and dimension reduction was carried out on data from all the cells, resulting in 37 clusters that were annotated based on their gene expression profiles and generally clustered together in UMAP space by differentiation timepoint (Fig. 5a and Extended Data Table 5).

We used pseudotime trajectory analysis using partition-based graph abstraction (PAGA), resulting in a similarity-based connectivity map (Extended Data Fig. 5b) that generated a model for the cellular lineage tree (Fig. 5b). The central SANCM differentiation pathway contained 11 cell types from day 3 to day 34. The earliest branch point occurred at day 3 when a subset of cardiac progenitors (C6) adopted a mesenchymal/fibroblast fate, while another population formed endoderm progenitors (C16).

The next major branch point occurred between day 6 and day 10, when TBX18^{+} progenitors (cluster 1) trifurcated into epicardial, PC, and sinus venosus lineages. By day 34, each of the PC subtypes (SAN head, SAN tail, and transitional cells) were resolvable as separate clusters in the UMAP. Visualizing each cluster along the main differentiation pathway using diffusion pseudotime analysis demonstrated that much of the differentiation towards distinct PC lineages occurred between D10 and D23 (Fig. 5c). In addition, between D23 and D34, when PC subtypes were resolved as separate clusters, PC progenitors (C28) and the differentiated PC subtypes (C0, C2, and C26) were relatively close in pseudotime. Indeed, the SAN head, tail and transitional subtypes emerged simultaneously in the hiPSC system from a common pool of progenitors rather than by differentiating sequentially, raising the possibility that PC subtypes are specified earlier than D23 in our differentiation process (Fig. 5d).

To test this possibility, we subclustered key populations along the SAN differentiation pathway, including C14 (*TBX18*$^{+}$ progenitors) and C28/C5 (SAN progenitors). This analysis revealed heterogeneity in expression of key venous pole myocardial transcription factors *TBX18*, *SHOX2*, *NKX2-5*, and *ISL1* at each of these key timepoints. Subclustering of C14, appearing at D6, already exhibited a discrete *TBX18*$^{+}$/*SHOX2*$^{+}$/*ISL1*$^{+}$ myocardial population that occupied a separate region of the UMAP compared to other TBX18 progenitors, suggesting that specification of PC myocardium occurred between D3 and D6 of the differentiation (Fig. 5e). Subclustering of C5, appearing at D10, revealed populations that resembled SAN head (*TBX18*$^{+}$/*ISL1*^{HIGH}/*SHOX2*^{HIGH}/*NKX2-5*$^{-}$), SAN tail (*TBX18*$^{-}$/*ISL1*^{LOW}/*SHOX2*^{HIGH}/*NKX2-5*^{LOW}) and transitional/atrial (*TBX18*$^{-}$/*NKX2.5*$^{+}$), suggesting that PC subtypes may also adopt their fates relatively early with progressive differentiation into discretely resolvable clusters by D34 (Fig. 5f). Importantly, this analysis suggests that some *TBX18*$^{+}$ myocardial progenitors downregulate *TBX18* expression during differentiation towards SAN tail or transitional phenotypes, as has been observed *in vivo*.

Pairwise comparison of D6 TBX18+ progenitors with epicardial, sinus venosus myocardium, and PC myocardium revealed that the majority of PC progenitors upregulated expression of pan-myocardial genes including *ACTC1*, *MYH6*, *RYR2*, *TTN*, *SLC8A1*, and *NKX2-5* (Fig. 5g). While expression increased compared to the progenitor state, only a small percentage of PC progenitors expressed *NPPA*, consistent with its general exclusion from differentiated SAN. The finding that most PC progenitors expressed *NKX2-5* agrees with the finding of transient activation of an *NKX2-5*-dependent myogenic program after PC specification in zebrafish [15]. By contrast, progenitors of the sinus venosus myocardium activated a hybrid smooth muscle/myocardial gene program that included *FN1*, *COL3AL1* as well as *SHOX2* and *PLN* (Extended Data Fig. 5c).

Comparison of D23 SAN progenitors with D34 PC subtypes revealed downregulation of *SHOX2* in more atrial-like transitional cells and upregulation of *SHOX2* in both SAN head and SAN tail, while only the SAN head upregulated *TBX18* and downregulated *NKX2-5*. In addition, both *BMP4* expression and *WNT2* expression were downregulated in the transitional cells (Fig. 5h). Gene ontology analysis of

upregulated genes in each of the SAN cardiomyocyte subtypes revealed several common biological processes related to cardiac and SAN development with each subtype enriched for genes associated with some unique GO terms (Extended Data Table 6)

Integration of Multi-Omics Data with Human Genetic Data for Gene Discovery and Annotation

Next, we tested whether putative regulatory regions in SANCM subtypes were significantly enriched for single nucleotide polymorphisms (SNPs) associated with heart rhythm-related traits. To this end, we generated Manhattan plots using summary statistics from previously published GWAS for resting heart rate and for atrial fibrillation but restricted the analysis to regions of open chromatin in SANCM clusters (C0, C1, C2, C4, and C6 in Fig. 4a) identified from our scATAC-seq dataset (Fig. 6a and Extended Data Fig. 6a,b). This restricted analysis almost completely recapitulated the known genetic associations identified in the source studies (Extended Data Fig. 6c,d and Extended Data Table 7), underscoring the relevance of the putative regulatory regions identified in our study to heritable genetic variation.

Next, we hypothesized that ATAC-seq marker peaks for different SANCM subtypes exhibit variable association strengths and effect sizes for different human traits. We therefore performed stratified linkage disequilibrium (LD) score regression to test whether trait heritability is enriched in accessible chromatin regions of each cell type using summary statistics from GWASs for 41 cardiac and non-cardiac traits (Fig. 6b). As expected, traits that are not directly related to cardiac physiology failed to show significant associations with regulatory regions of SANCM cells whereas multiple cardiac structural and functional traits showed differing degrees of association. Heritability of resting heart rate was significantly enriched within SNPs in the peaks of only SAN tail and transitional cells but not SAN head, while heart rate recovery after exercise (an indicator of cardiac autonomic responsiveness) was significantly associated with SAN head regions only. This strongly supports current models of SAN functional compartmentalization and highlight the concordance of our cluster assignments with *in vivo* functional classifications of these cells.

Next, we tested whether our integrative genomic analysis could facilitate novel gene discovery guided by colocalization of SNPs with regulatory regions and association of candidate regulatory regions with specific genes though integrated multi-omics analysis (incorporating scRNA-seq and peak-to-gene links). Because we only considered SNPs that overlapped with regions of differentially accessible chromatin, and we pruned SNPs in LD with one another, fewer SNPs were tested allowing for a higher P-value cut-off than is conventionally used for genome-wide analysis (5×10^{-8}). Employing this analysis for heart rate-associated SNPs uncovered 13 new loci from existing GWAS data that overlapped with SAN tail marker peaks, whereas a similar analysis of transitional cell marker peaks revealed 6 new loci from GWAS for atrial fibrillation (Fig. 6c). Of these newly identified loci, several were linked to genes expressed in one or more SAN clusters (Fig. 6d and Extended Data Fig. 6e).

For example, *FRMD3,* which encodes a cytoskeletal organizing protein linked to non-cardiac traits, was highly expressed in SAN head and tail clusters, less expressed in sinus venosus and transitional cell clusters, and undetected in non-myocytes. Closer examination of marker peak SNPs in relation to the scATAC-seq signal at this locus showed a differentially accessible ATAC-seq peak that colocalized with the 3'-untranslated region of the gene and was linked to the *FRDM3* promoter via a peak-to-gene link, supporting a role for this region in regulating expression of *FRDM3* specifically in PCs (Fig. 6e). Similarly, *TBX20,* which encodes for a transcription factor that regulates cardiomyocyte differentiation and transcription of many heart rhythm-related genes, had a transitional cell marker peak ~5 kilobases upstream of the promoter that colocalized with an atrial fibrillation-associated SNP (Extended Data Fig. 6f). Taken together, these results illustrate the potential of using our single cell multi-omics dataset to identify previously unknown genes associated with heritable traits, and to pinpoint cognate PC subtypes that may mediate their phenotypic effects.

DISCUSSION

In this work, we used single cell multi-omics of hiPSC-derived PCs to demonstrate that key aspects of *in vivo* SAN development and function can be modelled *in vitro*, including

differentiation of distinct PC subtypes and identification of genomic loci that influence heritability of heart rhythm traits and propensity to arrhythmias.

Large Scale Differentiation of PC Subtypes for Genomic and Functional Analyses

To generate PCs, we used a refinement of a growth factor-based approach that recapitulates key signaling pathways that are believed to regulate embryonic development, including biphasic Wnt signaling, retinoic acid, and inhibition of TGFβ and FGF signaling[11,24], yielding a cell population that compared favorably with other approaches, as evidenced by gene expression, electrophysiological behavior, and fidelity of epigenetic profile to *in vivo* counterparts. Notably, the minimal overlap of VCMs, ACMs, and SANCMs in the D34 scRNA-seq clustering demonstrates the superiority of directed differentiation over selection-based approaches.

Subclustering of our SANCMs allowed us to define DEGs that mark each PC subtype. These DEGs significantly overlapped with an scRNA-seq dataset derived from mouse neonatal SAN[31], indicating conservation of the gene regulatory networks controlling compartmentalization of the SAN. Finally, analysis of ~250 PCs dissected from the cranial SAN region of 5 adult human hearts also demonstrated agreement with our cluster assignments which strongly supports the fidelity of our SANCMs to *bona fide* human PCs.

Novel Transcriptional Pathways Implication in PC Development

Aggregating scATAC-seq data by cluster demonstrated differentially accessible chromatin regions associated with different PC subtypes and corroborated enhancers that were previously known for SAN transcription factors SHOX2, ISL1, TBX3, as well as unrecognized enhancers for TBX18. By demonstrating differential accessibility at these enhancers between SAN head, SAN tail, and transitional cells, our data suggest that the signaling upstream of these established SAN enhancers is involved not just in specification but also in fine patterning of the SAN. In addition, motif enrichment analysis demonstrated two novel transcriptional pathways. First, AP-1 motifs are highly overrepresented in peaks from sinus venosus myocardium, suggesting that this pathway could be critical for differentiation of this cell type from SAN myocardium. Of

note, sinus venosus myocardium exhibited a hybrid cardiomyocyte-smooth muscle gene program, so one possibility is that AP-1 may promote formation of TBX18+ progenitors and that persistent AP-1 activity could direct TBX18+ progenitors towards a sinus venosus fate. ATAC-seq data from mouse SAN revealed that AP-1 motifs are highly overrepresented in differentially accessible chromatin in neonatal mouse PCs as compared to atrial myocardium, suggesting an evolutionarily conserved role for AP-1 in regulating SAN gene expression[26].

In addition, we found that the NFI motif was highly overrepresented specifically in SAN head myocardium as compared to SAN tail myocardium and that *NFIA* and *NFIB* expression was enriched in SAN head cells. Further supporting a role for this novel pathway in SAN subtype differentiation, comparison of our scRNA-seq with a mouse dataset demonstrated that *NFIA* is differentially expressed in mouse SAN head versus SAN tail cells[31]. Although a recent study has shown that loss of NFIX transcription factor in mice results in altered adult SAN physiology without other effects on cardiac structure and function[32], NFI transcription factors have not been previously explored in SAN development. Along with ISL1, TBX, and MEF2, our integrative analysis supports a specific role for NFI as an activating transcriptional regulator in SAN Head versus San Tail. . Our *in vitro* system will permit further work on NFI-dependent mechanisms involved in SAN patterning.

hiPSC-Derived SAN Cells are Specified Shortly after Cardiomyogenesis

In our hiPSC system, spontaneous electrical activity was present at D10 and at faster rates in SANCM differentiations as early as D8, suggesting that key events in PC specification occurred shortly after cardiomyocytes are specified, analogous to the timing observed in embryos. In keeping with this result, clustering and pseudotime trajectory analysis of scRNA-seq data from D3, D6, D10, D23, and D34 revealed the likeliest differentiation pathway for PCs, and highlighted several precursor populations that demonstrate increasingly differentiated states until SAN Head, SAN Tail, and transitional cells form clearly resolved clusters by D34.

Additional clustering supported this model of early specification[15,33], as contrasting expression of *TBX18* and *NKX2-5* consistently emerged in UMAPs for each

of the clusters along the SANCM timeline, beginning with D6 *TBX18*$^{+}$ precursors. By D10, when robust beating was observed, the sinus venosus progenitors had already formed a separate cluster, while PC progenitors had separated into a *TBX18*$^{+}$ domain and a *NKX2-5*$^{+}$ domain. This pathway echoes observations from mouse embryos, in which subsets of cells from a broadly *Tbx18*$^{+}$ field at the venous pole downregulate *Tbx18*, upregulate *Nxk2-5*, and form the SAN Tail and transitional cells while persistent Tbx18 expression is required for formation of the SAN Head, as evidenced by the presence of a residual SAN Tail in *Tbx18*$^{-/-}$ mouse embryos with a severely hypoplastic SAN head[34].

ISL1 marks the entire posterior SHF and the SAN-forming field, and derivatives of *ISl1*$^{+}$ cells include cardiomyocytes, smooth muscle, epicardial cells, endocardial cells, and fibroblasts[8]. In SHF cardiomyocytes fated to form working myocardium, ISL1 maintains the proliferative progenitor state and then is downregulated as cardiomyocytes differentiate[35]. In contrast, ISL1 expression is maintained in cells fated to form PC myocardium and remains expressed in SAN cells in a gradient from the SAN head (highest expression) to SAN tail (lower expression) and transitional cells (minimal expression), findings that were also observed in SANCMs. A major question for future studies is whether, and if so, how, *ISL1* redistributes its genomic occupancy during the transition from cardiac progenitors to PCs. Integrative multi-omics with motif enrichment analysis strongly suggests that ISL1 is a cell type-specific positive transcriptional regulator that is most active in SAN head cells.

Integration of Omics Data with Human Genetic and Clinical Data for SNP Annotation and Novel Gene Discovery

Inherited variation in heart rhythm is associated with SNPs at numerous genetic loci that are presumed to influence traits through linkage to *cis*-regulatory elements that control gene expression at nearby loci. We found that differentially accessible chromatin in each of the SANCM subtypes significantly overlapped with SNPs associated with cardiac traits such as heart rate at rest (SAN tail and transitional cells) and heart rate recovery after exercise (SAN head). This finding is a striking confirmation of the hypothesized functional compartmentalization of the SAN, in which migration of the

leading PC along the cranial-caudal dimension is driven by changes in autonomic tone that affects SAN head cells disproportionately, whereas resting heart rate is driven by mechanisms of impulse generation and transmission that reflect the function of the SAN tail and transitional cells.

The strongest trait enrichment we observed was for atrial fibrillation. Clinically, there is a well- established link between sinus node function and the propensity to develop atrial fibrillation[36]. Because of the existence of common risk factors for both conditions, determining the causal basis for this association has been a long-standing question in the field with major clinical implications. More recent evidence indicates that heart rate determination and atrial fibrillation may have overlapping genetic architectures[37], although the ability to explore the mechanistic basis for these findings has been limited. Our hiPSC-based differentiation method coupled with multi-omics profiling offer a novel platform for mechanistic studies of genetic links between atrial fibrillation and sinus node dysfunction.

Finally, in a separate genomic analysis, we focused on association signals that do not rise to genome-wide significance but are significant when multiple hypothesis testing is restricted to fewer regions representing cell-type specific differentially accessible chromatin. Although GWASs have generated numerous candidate genes and pathways relevant to heart rhythm, the sum of effect sizes from significant hits even from well-powered GWASs for heart rate [38] and atrial fibrillation [39] leave a significant degree of heritability unaccounted for, presumably because of the underlying genetic complexity of these traits. Identifying new associations with smaller effect sizes and assigning them to specific cell types, as we have done here, expands the list of known associations and can thereby augment our understanding of trait heritability. In addition to supporting deeper biological investigation into these pathways, these associations could eventually result in the creation of more accurate polygenic risk scores for clinically significant traits. We also note that safe and effective pharmacotherapies for sinus node dysfunction and atrial fibrillation are currently lacking, and that success of drug discovery is more likely for targets identified through human genetic associations [40]. Thus, the resource we provide here could expand the potential target list for translational efforts and provide a platform for mechanistic studies.

Conclusion

A single cell transcriptomic and epigenetic characterization of a novel *in vitro* platform for hiPSC-based differentiation of SAN cardiomyocytes demonstrates reproducibility, scalability, and fidelity to developmental and physiological pathways important for SAN development and function. These datasets constitute a rich resource on human SAN development and differentiation and the *in vitro* differentiation system could facilitate future applications including studies on SAN gene regulatory networks, SAN physiology, functional genomics of human heart rhythm, drug screening, and tissue engineering.

ONLINE METHODS

Cell line and maintenance of hiPSC

All differentiations were performed using a commercially available Human episomal iPSC line (ThermoFisher). Culture of hiPSCs was carried out on hESC-Qualified Matrigel (Corning) coated dishes with mTeSR Plus medium (STEMCELL Technologies) and was refreshed every day. Cells were passaged at 80-90% confluence (~every 4 days) using ReLeSR (STEMCELL Technologies) according to manufacturer's instructions.

Differentiation of hiPSCs

Monolayer differentiations of cardiomyocytes were carried out using a modified GiWi strategy (Lian et al 2013). hiPSCs were grown on Matrigel to 90% confluency then harvested as single cell suspension using TrypLE (ThermoFisher) and resuspended in mTeSR Plus containing 10µM ROCK inhibitor Y-27632 (Reprocell). Cells were counted using a Vi-CELL XR Cell Viability Analyzer (Beckman Coulter) and re-plated onto Matrigel coated plates at 15,625-18,230 cells/cm^2 in mTeSR Plus containing 10µM ROCK inhibitor Y-27632 (day -3 of differentiation). Cell plating density was re-optimized every 10 passages or upon a new thaw. At day -2 media was changed to mTeSR Plus. For ventricular cardiomyocytes, at day 0 when cells were 80-90% confluent media was

changed to Cardiac Differentiation Media (CDM; RPMI1640 with 25mM HEPES and Glutamax (ThermoFisher) and 1x B27 minus insulin (ThermoFisher)) containing 8µM CHIR99021 (Sigma). On day 2 of differentiation cells were changed to CDM containing 5µM IWP2 (Sigma). At day 4, media was changed to CDM until day 6 when media was switched to Cardiac Maintenance Media (CMM; RPMI1640 with 25mM HEPES and Glutamax (ThermoFisher) and 1x B27 supplement (ThermoFisher). Media was changed every 2 days until day 20. On day 20 cells were split 1:4 in preparation for metabolic selection. Cells were washed with TrypLE for 5 minutes at room temperature then digested into single cell suspension using prewarmed Accumax (Sigma) for 20-30 minutes at 37°C. Digestion was quenched using CMM containing 20% Knock-out serum replacement (ThermoFisher) and 10µM ROCK inhibitor Y-27632 and cells collected by centrifugation at 200xg for 5 minutes. Cells were then resuspended in CMM containing 20% Knock-out serum replacement and 10µM ROCK inhibitor Y-27632 and plated onto new Matrigel coated dishes at 25% the original confluence (1:4 split). 24 hours later (day 21) media was changed to CMM. At day 23 media was changed to Cardiomyocyte Selection Media (CSM; RPMI1640 with Glutamine and No glucose (ThermoFisher), 5mM Sodium DL-lactate solution (Sigma) and 0.5mg/ml Human Serum Albumin (Sigma)). CSM was changed every 2 days until day 31 when cells were switched back to CMM. Cells were maintained in CMM (media changed every 2 days) until cells were either frozen (day 34) or used for analysis.

For atrial cardiomyocytes, the same protocol was followed except for the following addition. On day 3 of differentiation, 1µM Retinoic Acid (RA; Sigma) was spiked into the CDM containing IWP2. For SAN cardiomyocytes the ventricular protocol was following with the following modifications. At day 3, 2.5ng/ml BMP4 (R&D Systems), 5.4µM SB431542 (Tocris) and 0.5µM RA were spiked into the CDM containing IWP2. On day 4, media was replaced with CDM containing 240nM PD173074 (Tocris).

Quantitative real-time RT-PCR

Total RNA was extracted from cells using RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Purified RNA was subjected to reverse transcription as

described in the SuperScript IV VILO with ezDNase enzyme kit (ThermoFisher). Following cDNA synthesis, expression level of the genes was quantified by quantitative Real-Time PCR (qPCR) using TaqMan Fast Advanced Master Mix (ThermoFisher) on a QuantStudio 7 Flex Real-Time PCR system (ThermoFisher). The quantification was analysed using the C_T (threshold cycle) values. The expression of the target genes were normalized to *GAPDH.* Expression was then calculated using the $2^{-\Delta\Delta Ct}$ method and graphed relative to ventricular cardiomyocytes.

Immunofluorescence staining

Cells were cultured in CellCarrier-96 Ultra Microplates (Perkin Elmer) coated with 50µg/ml fibronectin (Sigma) and 0.1% Gelatin (STEMCELL Technologies). Cells were washed with PBS then fixed with 4% PFA for 20 min at room temperature and washed again with PBS. Cells were then blocked and permeabilized with 0.1% Triton X-100 and 5% donkey serum in PBS for 1 hour at room temperature on a rotator. Primary antibodies diluted in blocking/permeabilization buffer were added and incubated at 4°C overnight on a shaker at 3000 rpm. Cells were washed 3 times for 3 minutes in PBS at room temperature on a shaker at 3000 rpm. Secondary antibodies diluted in PBS with 5% donkey serum were added and incubated at room temperature for 1 hour shaking at 3000 rpm then followed by 2x3 minute washes in PBS with shaking. The cell nuclei were counterstained with Hoesht 33342 (1:2000; Thermo Fisher) for 10 minutes with shaking and then washed 2x3 minutes with shaking in PBS. Cells were imaged with an Opera Phenix High Content Screening System (Perkin Elmer) using Harmony software and analyzed using Columbus software (Perkin Elmer).

The following primary antibodies were used: mouse anti-Cardiac Troponin T (ThermoFisher; 1:200), rabbit anti-Cardiac Troponin T (Abcam; 1:200), mouse anti-ISL1/2 (DSHB; 1:200), rabbit anti-NKX2.5 (Cell Signaling; 1:200), rabbit anti-HCN4 (Alomone; 1:200) and mouse anti-SHOX2 (Abcam; 1:200). The following secondary antibodies were used: donkey anti-mouse Alexa Fluor 488 (ThermoFisher; 2µg/ml) and donkey anti-rabbit Alexa Fluor 555 (ThermoFisher; 2µg/ml).

Imaging of Voltage-Sensitive Dye

Voltage imaging of VCMs and SANCMs was performed by adding FluoVolt membrane potential dye (ThermoFisher) to 96 well-plates seeded with 100k cells/well. Briefly, FluoVolt was mixed 1:20 with PowerLoad concentrate, and diluted 100x with FluoroBrite DMEM (ThermoFisher). 100x water-soluble Probenecid (ThermoFisher) was then added to achieve a final concentration of 1x. After washing with FluoroBrite DMEM, cells were incubated for 5 min at RT with the dye solution. The dye was then removed and replaced with FluoroBrite DMEM with 1x Probenecid. The cell plate was then placed inside the culture chamber of an iMX4 microscope (Molecular Devices) and equilibrated for 5 minutes. Time lapse images were then acquired for 5 seconds per site at 100 fps, using a 10x objective. For Zatebradine measurements, cells were prepared as described above then treated with 0, 0.1, 0.3, 1 or 3 µM Zatebradine (Tocris), incubated for 5 minutes then imaged. The images (5 fields per well, 4 wells per treatment) were then analyzed using a custom Matlab script that extracted individual voltage traces for each cell in the field and calculated electrophysiological parameters including beating rate, amplitude, and action potential duration at 20%, 50% and 90% repolarization. All parameters were averaged and compared between groups using a student's T-test after each distribution passed a normality test.

scRNA-seq data analysis

Initial processing of scRNA-seq data was done with the Cell Ranger Pipeline (https://support.10xgenomics.com/single-cell-geneexpression/software/pipelines/latest/what-is-cell-ranger, v.6.0.0) by first running cellranger mkfastq to demultiplex the bcl files and then running cellranger count with 10x Genomics' pre-built Cell Ranger reference GRCh38-2020-A_build. After running Cell Ranger, the filtered_feature_bc_matrix produced by Cell Ranger was read into R (v.4.0.4) with the Seurat (v.4.3.0) function Read10X [41]. scDblFinder (v.1.2.0) was run for each sample for detecting doublets [42]. Cells classified as doublets were not enriched in any clusters identified in the downstream clustering analysis and thus retained. Cells from all samples were merged into a single Seurat object and analyzed separately for each dataset.

For the analysis of cells of day 34 datasets (ACM, VCM, and SANCM cultures), cell barcodes were filtered based on the number of genes per cell (between 1,000 and 9,000), unique molecular identifiers (UMIs) per cell (more than 1500), and percentage of mitochondrial reads per cell (less than 50%). The filtered cell number used for subsequent analysis was 20,712 cells for the day 34 SANCM culture, 21,676 cells for the day 34 ACM culture, 21,934 cells for the day 34 VCM culture, 102,496 cells for the SANCM time-course day 3 to day 34 cultures, respectively.

For the analysis of SANCM time-course datasets (day 3 to day 34 SANCM cultures), the following criteria was applied: number of genes per cell between 2,000 and 10,000, UMIs per cell (between 5,000 and 80,000), percentage of mitochondrial reads per cell (less than 10% for day3, day 4, and day 6 datasets; less than 15% for day 10 and day 23 datasets; less than 40% for day 34). Read counts were normalized using Seurat's sctransform method. Cell-cycle scoring was calculated with the CellCycleScoring algorithm from Seurat using cell-cycle-related genes and cell cycle effects were regressed out (vars.to.regress = c("S.Score", "G2M.Score")). The 5000 most variable genes were identified for subsequent analysis.

For the analysis of day 34 datasets and SANCM time-course datasets, principal component analysis (PCA) was performed using the RunPCA function and the top 40 principal components (PCs) were selected. UMAP was computed using the RunUMAP function and cells were clustered using a k-nearest neighbour graph and the Louvain algorithm with the FindNeighbors and FindClusters functions. A resolution of 0.6 and 1.0 was used for day 34 datasets and time-course datasets, respectively. Expression of canonical marker genes and enriched pathways in the marker genes (described below) in the resulting clusters were then used to label clusters.

For the sub-clustering analysis of day 34 SANCM, ACM, and VCM cultures, replicates were integrated using the Seurat integration procedure. When running FindIntegrationAnchors and IntegrateData, the normalization.method parameter was set to the "SCT" using 30 dimensions and 5000 features. For dimensionality reduction, the integrated expression matrix was used to compute PCs by RunPCA. The top 40 PCs were selected for UMAP visualizations and clustering analysis. The cells were then clustered using Seurat's FindNeighbors and FindClusters with a resolution of 0.5 for the

day 34 SANCM dataset, 0.4 for the day 34 VCM dataset, and 0.3 for the day 34 ACM dataset, respectively. The clusters were manually annotated based on canonical markers and enriched pathways.

For the sub-clustering analysis of the day 34 SANCM dataset, cells were first separated into 20 clusters as described above. Eleven clusters were removed due to enrichment of cell cycle genes (3 clusters), low gene count (4 clusters), high percentage of mitochondrial reads (2 clusters), enrichment of genes in cell stress related pathways (1 cluster), or without consistent expression of canonical markers of any cell type (3 clusters). Cells in the retained clusters were normalized and integrated again using the same pipeline. Dimensionality reduction was recomputed, and clustering analysis was performed with a resolution of 0.8.

For the sub-clustering analysis of SAN head and tail cells, the three clusters identified as SAN head and tail cells (cluster 2, 0, and 3 in Fig. 3C) were subset and analyzed separately. Normalization was performed by Seurat's sctransform method with the 3,000 most variable genes and regressing out cell cycle effects, and batch effect was removed using Harmony (v. 0.1.0)[43].Cluster identification and UMAP reduction were conducted as above, except that a resolution of 0.1 was used for the clustering. For the sub-clustering of clusters 28, 5, and 14 in the SANCM time course datasets (Fig. 5A), the same pipeline was used except that a resolution of 0.3, 0.2, and 1.0 were used for these clusters, respectively.

For each dataset, the FindAllMarkers and FindMarkers functions were performed after normalizing the expression counts with Seurat's "LogNormalize" method to identify the marker genes for each cluster (min.pct = 0.25, logfc.threshold = 0.25) and intercluster differentially expressed genes (min.pct = 0.25, Padj < 0.05, logfc.threshold = 0.5 for SANCM time-course datasets and 0.25 for day 34 datasets) via the Wilcoxon rank sum test. Pathway analysis of marker genes and differentially expressed genes was conducted with MetaScape (v3.5) [44].

Trajectory Analysis

PAGA [45] implemented in scanpy (v.1.9.1[46]) was used to compute the connectivites between clusters (sc.tl.paga) and compute the PAGA graph after

denoising with diffusion map. Cell differentiation lineages were constructed manually based on the similarities between clusters, timepoints and annotation of the clusters. For each lineage, diffusion map and diffusion pseudotime were computed using the R package destiny (v.3.2.0)[47].

Integration of publicly available scRNA-seq datasets

One preprocessed scRNA-seq datasets was downloaded from the Gene Expression Omnibus (GEO) (accession code GSE189782[19]) and processed using Seurat as described in the studies. Processed snRNA-seq data of human SAN from donors with normal conduction parameters [PMID: 37438528] was downloaded from the Heart Cell Atlas (https://www.heartcellatlas.org). We day performed reference mapping to project the cells in the two public datasets onto the UMAP structure of our day 34 SANCM scRNA-seq data according to Seurat's "Mapping and annotating query datasets" tutorial (https://satijalab.org/seurat/articles/integration_mapping.html). Specifically, the FindTransferAnchors function was used to find a set of anchors between the reference and query object with the top 40 PCs and "SCT" normalization. These anchors were then passed into the MapQuery function to project the query data onto the UMAP embedding of the reference. day

scATAC-seq data analysis

Initial processing of the day 34 SANCM scATAC-seq data was performed using Cell Ranger ATAC (https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/what-is-cell-ranger-atac, v.2.0.0) by first running cellranger-atac mkfastq to demultiplex the bcl files into FASTQ files and then running cellranger-atac count to generate scATAC fragments files with 10x Genomics' pre-built Cell Ranger reference GRCh38-2020-A_build and default parameters.

These fragments files were loaded into R (v.4.0.4) using the createArrowFiles function in ArchR (v.1.0.1) [48]. Quality control metrics were computed for each cell, and cells with TSS enrichments less than 4 and number of unique fragments less than 1,000 were filtered out for all samples. The number of cells retained was 38,744 (ranging from 10,565 to 15,823). Doublets were detected by AMULET (v.1.1) [49] for each sample. Cells

classified as doublets were retained due to lack of enrichment in any clusters identified in the subsequent clustering analysis.

After generating arrow files for each sample, an ArchR project containing all samples was created. As part of construction of the arrow files, a tile matrix consisting of reads in 500-bp tiles was constructed. This tile matrix was used as input to compute an iterative latent semantic indexing (LSI) dimensionality reduction using the addIterativeLSI function in ArchR with a total of 2 iterations, a clustering resolution of 0.2 following the first iteration, 50,000 variable features, 30 dimensions and sampling 10,000 cells. Harmony (v. 0.1.0) [43] was applied to the dataset to correct for the batch effects. Using the Harmony-corrected dimension reduction, clustering was performed using ArchR's addClusters with a resolution of 0.45. We next ran addUMAP on the Harmony-corrected dimensionality reduction with 30 nearest neighbors and a minimum distance of 0.5. Gene activity scores computed by ArchR base on the accessibility within and around a given gene body were examined for the known marker genes. Clusters with less than 34 cells consisted mainly of cells from only a subset of the samples and were removed from the subsequent analysis. One cluster containing about 3000 cells that could not be confidently assigned to one cell type and probably constitute poor quality nuclei was also removed from further analysis.

For sub-clustering of SAN head and tail cells, scATAC-seq clusters identified as SAN head and tail cells were subset and used to recompute iterative LSI dimensionality reduction as described above. Batch correction, cluster identification, and UMAP reduction were also performed as above, except that a resolution of 0.2 was used for the clustering.

Integration of scATAC-seq and scRNA-seq data

day 34 SANCM scRNA-seq gene expression counts were integrated with scATAC-seq data with addGeneIntegrationMatrix in ArchR as a constrained integration, grouping GC B cell clusters, other B cell clusters, and non–B cell clusters together during addGeneIntegrationMatrix to the following groups: SAN head, SAN tail, SAN transitional, ACM-like transitional, Sinus Venosus, proepicardial and neuronal cells. The output of the integration step resulted in snATAC-seq cells having both a chromatin

accessibility and a gene expression profile. Accessibility gene scores and transferred RNA expression counts were imputed with addImputeWeights.

Cell-type-specific peak and TF motif enrichment

Pseudobulk group coverages of cell type clusters were calculated with the 'addGroupCoverages' function and used for peak calling using the 'addReproduciblePeakSet' function in ArchR. A background peak set controlling for total accessibility and GC content was generated using addBgdPeaks for TF enrichment analyses. Cell type–specific marker peaks and differentially accessible peaks between clusters were identified with 'getMarkerFeatures' with the Wilcoxon test (maxCells = 10,000, FDR ≤ 0.05, and Log2FC ≥ 0.5 for marker peaks, and maxCells =6,000, FDR ≤ 0.01, and |Log2FC| ≥ 1 for differential accessible peaks). The enriched motifs for marker peaks of each cluster and differential accessible peaks were predicted using the 'addMotifAnnotations' function in ArchR based on the cisbp motif set. Chromvar was run with addDeviationsMatrix to calculate enrichment of chromatin accessibility at different TF motif sequences in single cells. To identify correlations between the gene expression and TF activity, the GeneIntegrationMatrix and the Chromvar deviations (MotifMatrix) were correlated using correlateMatrices. A correlation of greater than 0.5 and an adjusted p-value less than 0.01 were used to determine whether TF expression and activity were positively correlated. "Peak-to-gene links" were calculated using correlations between peak accessibility and integrated scRNA-seq expression data using addPeak2GeneLinks with the parameter 'maxDist = 1.5e6' and getPeak2GeneLinks in ArchR. We further visualized the peak to gene links using the 'plotBrowserTrack' function in ArchR.

Stratified linkage disequilibrium (LD) score regression

We used the LDSC package (v1.0.1) [50] to estimate genome-wide GWAS enrichment for a variety of phenotypes within the cell type-specific scATAC-seq peaks. We first downloaded GWAS summary statistics for: atrial fibrillation, resting heart rate, heart rate increase in response to exercise, heart rate response to recovery post

exercise (20 sec) (the GWAS catalog https://www.ebi.ac.uk/gwas/; accessions codes GCST006414, GCST007609, GCST005845, and GCST005847); height (https://rpubs.com/leaknielsen/1027689); diastolic blood pressure, systolic blood pressure, years of education, smoking status, auto immune traits, asthma, dermatologic diseases, high cholesterol, type 2 diabetes, age at menarche, BMI (https://alkesgroup.broadinstitute.org/UKBB/); schizophrenia, major depression disorder, Alzheimer's disease (https://pgc.unc.edu/for-researchers/download-results/; pubMedIDlink 25056061, 29700475, 30617256); coronary artery disease, QT prolongation, LV stroke vol. indexed, sscending aorta diameter index, LV max. area index, LV min. area index, pulmonary artery:aorta ratio, RA fractional area change, RA max. area index, RA min. area index, RV end diastolic vol. index, RV ejection fraction, RV end systolic vol. index, P-wave duration, PR interval, QRS duration, QTc interval, RR interval, heart failure, LV ejection fraction (https://personal.broadinstitute.org/ryank/; Aragam_2022_CARDIoGRAM_CAD_GWAS.zip, Nauffal_2022_QT_GWAS_SAIGE.zip, Pirruccello_2022_UKBB_HeartStructures.zip, Choi2021_TOPMed_freeze6.zip, HERMES_Jan2019_HeartFailure_summary_data.txt.zip, MRI_lvef_filtered.zip); IBD (https://www.dropbox.com/s/ttuc6s7tv26voq3/iibdgc-trans-ancestry-filtered-summary-stats.tgz?dl=0); total cholesterol (http://csg.sph.umich.edu/abecasis/public/lipids2013/jointGwasMc_TC.txt.gz). We then used the provided munge_sumstats.py script to convert these GWAS summary statistics to a format compatible with LDSC.

Using all the ATAC-seq peaks called within each cluster, we generated the annotation-specific partitioned LD score files for each cluster separately according to the "Partitioned LD scores" section of the LD score estimation tutorial (https://github.com/bulik/ldsc/wiki/LD-Score-Estimation-Tutorial). We downloaded the list of 1,154,611 candidate cis-regulatory elements (cCREs) for 222 distinct cell types from the Human Enhancer Atlas (http://catlas.org/humanenhancer/#!/)[51] and used them as background annotations. We lifted over peak coordinates from hg38 to hg19 for both day 34 SANCM ATAC-seq peaks and background peaks using CrossMap (v.0.6.3) [52] before computing annotation-specific LD scores. We then used these hg19 bed files to make annotation files for each cell type. For each trait, we used LD score regression to

estimate coefficient p value for each cell type relative to the background annotations according to the cell-type-specific analysis tutorial (https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses) and used the Benjamini-Hochberg correction for multiple testing.

Integration of GWAS and scATAC-seq data

GWAS results for resting heart rate, heart rate response to recovery post exercise, and atrial fibrillation were used to integrate with scATAC-seq data. First, summary statistics for resting heart rate (GCST007609), and heart rate response to recovery post exercise (GCST005847), and atrial fibrillation (accessions codes GCST006414 [53] and GCST006061[54]) were downloaded from the GWAS catalog. All the SNPs were then assigned to their nearest genes based on their coordinates. Bedtools (v.2.30.0)[55] was used to intersect the SNP coordinates with marker peaks of SAN tail for resting heart rate, SAN head for heart rate response to recovery post exercise, and SAN transitional cells for atrial fibrillation respectively and also the union of all the peaks identified in these SAN cells. To calculate the customized genome-wide significance threshold for the SNPs within the ATAC peaks, the SNPclip tool of LDlink (https://ldlink.nih.gov/?tab=home) [56] was used to prune the list of SNPs in the marker peaks of each cell type by linkage disequilibrium with the CEU population (R^2 0.1 and MAF 0.01) and the GRCh38 genome build. The p-value threshold was then calculated through dividing 0.05 by the number of SNPs left after pruning (0.05/6522 = 7.67e-06 for resting heart rate in SAN tail, 0.05/10818 = 4.62e-06 for heart rate response to recovery post exercise in SAN head, and 0.05/13351 = 3.75e-06 for atrial fibrillation in SAN transitional cells). To identify the 'novel' genes associated with the three traits, the SNPs in the marker peaks with p-value between the traditional GWAS cutoff 5e-08 and the customized p-value threshold were selected, and a list of their nearest genes was obtained, for each trait separately. Only the genes which were not assigned with a genome-wide significant SNP (p-value < 5e-08), whose assigned SNPs were not in high LD (r^2 > 0.8) with a genome-wide significant SNP, and which were not associated with the traits in the GWAS catalog were considered as novel findings.

Acknowledgements. The authors wish to thank members of the Vedantham Lab for helpful discussions.

Sources of Funding. This work was supported by a grant from the NIH (DP2HL152425 to V.V.) and a philanthropic gift from the Ryan Family Trust.

REFERENCES

1 Mangoni, M. E. & Nargeot, J. Genesis and regulation of the heart automaticity. *Physiol Rev* **88**, 919-982, doi:88/3/919 [pii]

10.1152/physrev.00018.2007 (2008).

2 Bleeker, W. K., Mackaay, A. J., Masson-Pevet, M., Bouman, L. N. & Becker, A. E. Functional and morphological organization of the rabbit sinus node. *Circ Res* **46**, 11-22 (1980).

3 Liu, J., Dobrzynski, H., Yanni, J., Boyett, M. R. & Lei, M. Organisation of the mouse sinoatrial node: structure and expression of HCN channels. *Cardiovasc Res* **73**, 729-738, doi:10.1016/j.cardiores.2006.11.016 (2007).

4 Unudurthi, S. D., Wolf, R. M. & Hund, T. J. Role of sinoatrial node architecture in maintaining a balanced source-sink relationship and synchronous cardiac pacemaking. *Front Physiol* **5**, 446, doi:10.3389/fphys.2014.00446 (2014).

5 Boyett, M. R., Honjo, H. & Kodama, I. The sinoatrial node, a heterogeneous pacemaker structure. *Cardiovasc Res* **47**, 658-687, doi:10.1016/s0008-6363(00)00135-8 (2000).

6 Mommersteeg, M. T. *et al.* The sinus venosus progenitors separate and diversify from the first and second heart fields early in development. *Cardiovasc Res* **87**, 92-101, doi:10.1093/cvr/cvq033 (2010).

7 Liang, X. *et al.* HCN4 dynamically marks the first heart field and conduction system precursors. *Circ Res* **113**, 399-407, doi:CIRCRESAHA.113.301588 [pii]

10.1161/CIRCRESAHA.113.301588 (2013).

8 Sun, Y. *et al.* Islet 1 is expressed in distinct cardiovascular lineages, including pacemaker and coronary vascular cells. *Dev Biol* **304**, 286-296, doi:10.1016/j.ydbio.2006.12.048 (2007).

9 Devalla, H. D. *et al.* Atrial-like cardiomyocytes from human pluripotent stem cells are a robust preclinical model for assessing atrial-selective pharmacology. *EMBO Mol Med* **7**, 394-410, doi:10.15252/emmm.201404757 (2015).

10 Funakoshi, S. *et al.* Generation of mature compact ventricular cardiomyocytes from human pluripotent stem cells. *Nat Commun* **12**, 3155, doi:10.1038/s41467-021-23329-z (2021).

11 Protze, S. I. *et al.* Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker. *Nat Biotechnol* **35**, 56-68, doi:10.1038/nbt.3745 (2017).

12 Chou, P. C. *et al.* Fibroblasts Drive Metabolic Reprogramming in Pacemaker Cardiomyocytes. *Circ Res* **131**, 6-20, doi:10.1161/CIRCRESAHA.121.320301 (2022).
13 Hou, X. *et al.* Chemically defined and small molecules-based generation of sinoatrial node-like cells. *Stem Cell Res Ther* **13**, 158, doi:10.1186/s13287-022-02834-y (2022).
14 Wakimizu, T. *et al.* SHOX2 refines the identification of human sinoatrial nodal cell population in the in vitro cardiac differentiation. *Regen Ther* **21**, 239-249, doi:10.1016/j.reth.2022.07.012 (2022).
15 Ren, J. *et al.* Canonical Wnt5b Signaling Directs Outlying Nkx2.5+ Mesoderm into Pacemaker Cardiomyocytes. *Dev Cell* **50**, 729-743 e725, doi:10.1016/j.devcel.2019.07.014 (2019).
16 Jung, J. J. *et al.* Programming and isolation of highly pure physiologically and pharmacologically functional sinus-nodal bodies from pluripotent stem cells. *Stem Cell Reports* **2**, 592-605, doi:10.1016/j.stemcr.2014.03.006

S2213-6711(14)00085-X [pii] (2014).
17 Yan, Y., Liu, F., Dang, X., Zhou, R. & Liao, B. TBX3 induces biased differentiation of human induced pluripotent stem cells into cardiac pacemaker-like cells. *Gene Expr Patterns* **40**, 119184, doi:10.1016/j.gep.2021.119184 (2021).
18 Gorabi, A. M. *et al.* TBX18 transcription factor overexpression in human-induced pluripotent stem cells increases their differentiation into pacemaker-like cells. *J Cell Physiol* **234**, 1534-1546, doi:10.1002/jcp.27018 (2019).
19 Wiesinger, A. *et al.* A single cell transcriptional roadmap of human pacemaker cell differentiation. *Elife* **11**, doi:10.7554/eLife.76781 (2022).
20 Lee, J. H., Protze, S. I., Laksman, Z., Backx, P. H. & Keller, G. M. Human Pluripotent Stem Cell-Derived Atrial and Ventricular Cardiomyocytes Develop from Distinct Mesoderm Populations. *Cell Stem Cell* **21**, 179-194 e174, doi:10.1016/j.stem.2017.07.003 (2017).
21 Cyganek, L. *et al.* Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes. *JCI Insight* **3**, doi:10.1172/jci.insight.99941 (2018).
22 Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nat Protoc* **8**, 162-175, doi:10.1038/nprot.2012.150 (2013).
23 Tohyama, S. *et al.* Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell* **12**, 127-137, doi:10.1016/j.stem.2012.09.013 (2013).
24 Wiesinger, A., Boink, G. J. J., Christoffels, V. M. & Devalla, H. D. Retinoic acid signaling in heart development: Application in the differentiation of cardiovascular lineages from human pluripotent stem cells. *Stem Cell Reports* **16**, 2589-2606, doi:10.1016/j.stemcr.2021.09.010 (2021).
25 Vedantham, V., Galang, G., Evangelista, M., Deo, R. C. & Srivastava, D. RNA Sequencing of Mouse Sinoatrial Node Reveals an Upstream Regulatory Role for Islet-1 in Cardiac Pacemaker Cells. *Circulation research* **116**, 797-803, doi:10.1161/CIRCRESAHA.116.305913 (2015).

26 Galang, G. *et al.* ATAC-Seq Reveals an Isl1 Enhancer That Regulates Sinoatrial Node Development and Function. *Circ Res* **127**, 1502-1518, doi:10.1161/CIRCRESAHA.120.317145 (2020).
27 van Eif, V. W. W. *et al.* Transcriptome analysis of mouse and human sinoatrial node cells reveals a conserved genetic program. *Development* **146**, doi:10.1242/dev.173161 (2019).
28 Wiese, C. *et al.* Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by Tbx18 and Tbx3. *Circulation Research* **104**, 388-397, doi:10.1161/CIRCRESAHA.108.187062 (2009).
29 Kanemaru, K. *et al.* Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801-810, doi:10.1038/s41586-023-06311-1 (2023).
30 van Eif, V. W. W. *et al.* Genome-Wide Analysis Identifies an Essential Human TBX3 Pacemaker Enhancer. *Circ Res* **127**, 1522-1535, doi:10.1161/CIRCRESAHA.120.317054 (2020).
31 Goodyer, W. R. *et al.* Transcriptomic Profiling of the Developing Cardiac Conduction System at Single-Cell Resolution. *Circulation Research* **125**, 379-397, doi:10.1161/CIRCRESAHA.118.314578 (2019).
32 Landi, S. *et al.* Lack of the transcription factor Nfix causes tachycardia in mice sinus node and rats neonatal cardiomyocytes. *Acta Physiol (Oxf)*, e13981, doi:10.1111/apha.13981 (2023).
33 Bressan, M., Liu, G. & Mikawa, T. Early mesodermal cues assign avian cardiac pacemaker fate potential in a tertiary heart field. *Science* **340**, 744-748, doi:science.1232877 [pii]

10.1126/science.1232877 (2013).
34 Wiese, C. *et al.* Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by Tbx18 and Tbx3. *Circ Res* **104**, 388-397, doi:10.1161/CIRCRESAHA.108.187062 (2009).
35 Cai, C. L. *et al.* Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell* **5**, 877-889, doi:S1534580703003630 [pii] (2003).
36 John, R. M. & Kumar, S. Sinus Node and Atrial Arrhythmias. *Circulation* **133**, 1892-1900, doi:10.1161/CIRCULATIONAHA.116.018011 (2016).
37 Siland, J. E. *et al.* Resting heart rate and incident atrial fibrillation: A stratified Mendelian randomization in the AFGen consortium. *PLoS One* **17**, e0268768, doi:10.1371/journal.pone.0268768 (2022).
38 Xhaard, C. *et al.* Heritability of a resting heart rate in a 20-year follow-up family cohort with GWAS data: Insights from the STANISLAS cohort. *Eur J Prev Cardiol* **28**, 1334-1341, doi:10.1177/2047487319890763 (2021).
39 Roselli, C., Rienstra, M. & Ellinor, P. T. Genetics of Atrial Fibrillation in 2020: GWAS, Genome Sequencing, Polygenic Risk, and Beyond. *Circ Res* **127**, 21-33, doi:10.1161/CIRCRESAHA.120.316575 (2020).
40 Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-860, doi:10.1038/ng.3314 (2015).
41 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

42 Germain, P. L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979, doi:10.12688/f1000research.73600.2 (2021).
43 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).
44 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).
45 Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59, doi:10.1186/s13059-019-1663-x (2019).
46 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).
47 Angerer, P. *et al.* destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241-1243, doi:10.1093/bioinformatics/btv715 (2016).
48 Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411, doi:10.1038/s41588-021-00790-6 (2021).
49 Thibodeau, A. *et al.* AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol* **22**, 252, doi:10.1186/s13059-021-02469-x (2021).
50 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
51 Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001 e5919, doi:10.1016/j.cell.2021.10.024 (2021).
52 Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006-1007, doi:10.1093/bioinformatics/btt730 (2014).
53 Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet* **50**, 1234-1239, doi:10.1038/s41588-018-0171-3 (2018).
54 Roselli, C. *et al.* Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet* **50**, 1225-1233, doi:10.1038/s41588-018-0133-9 (2018).
55 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
56 Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557, doi:10.1093/bioinformatics/btv402 (2015).

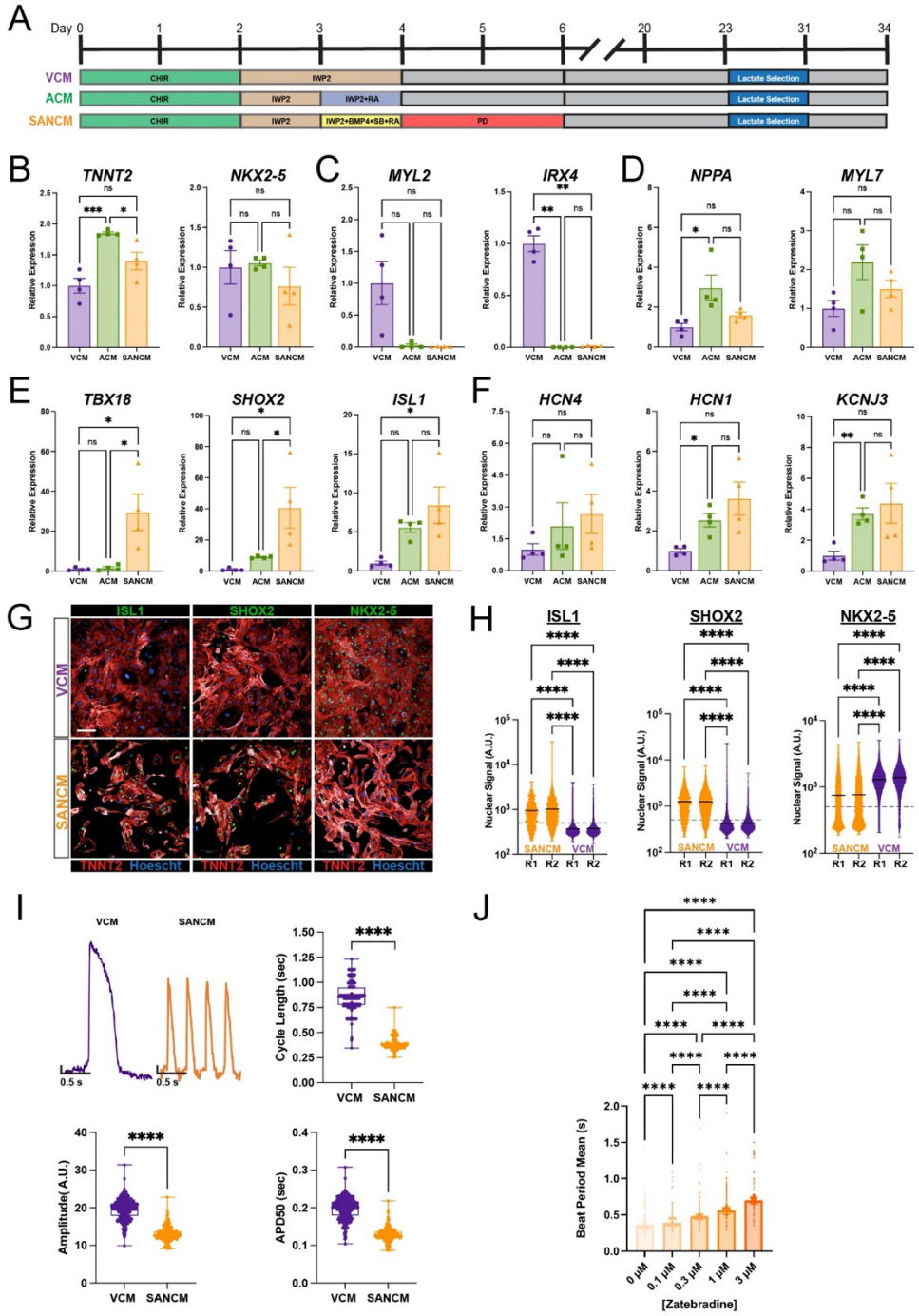


Figure 1: Generation and characterization of hiPSC derived Cardiomyocyte-like subtypes. (A) Schematic showing differentiation protocols for generation of hiPSC derived Cardiomyocyte-like subtypes. (B-F) qPCR for (B) pan-cardiac, (C) ventricular, (D) atrial, (E) pacemaker and (F) ion channel related genes in hiPSC derived Cardiomyocyte subtypes. (G) Immunocytochemistry of ISL1, SHOX2 and NKX2-5 expression (green) in hiPSC derived VCMs and SANCMs stained with TNNT2 (red) and Hoescht (blue). (H) Quantification of nuclear signal intensities from G in two replicates. Dotted lines indicate threshold cutoffs for quantification. (I) Representative traces, cycle length, amplitude and ADP50 measurements recorded from VCMs and SANCMs using FluoVolt. (J) Dose response of SANCMs treated

with the HCN channel blocker Zatebradine. Significance was determined using one-way ANOVA with multiple comparisons tests. * $p<0.05$ ** $p<0.01$ *** $p<0.001$ **** $p<0.0001$

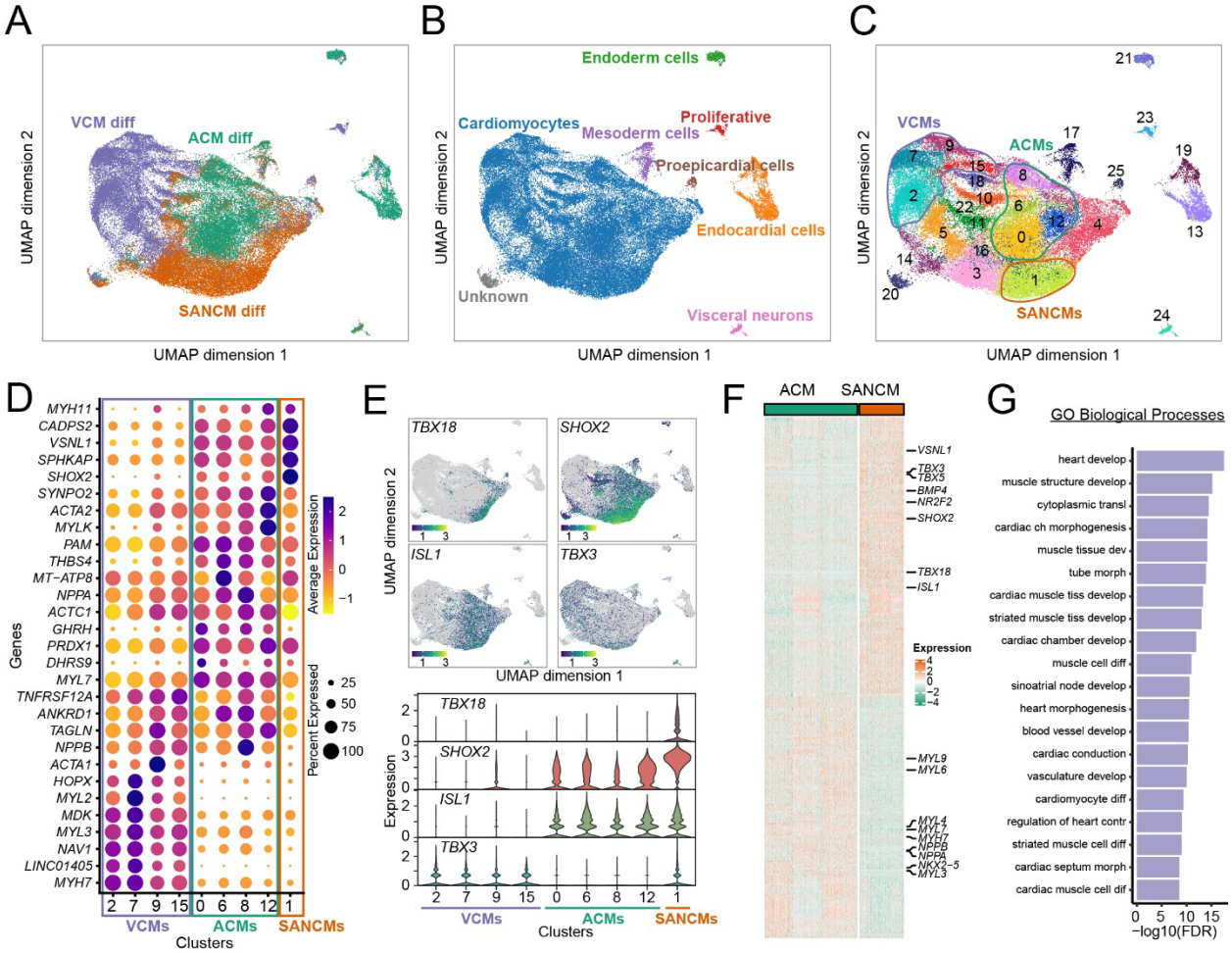


Figure 2: scRNA-seq of day 34 hiPSC derived VCMs, ACMs and SANCMs. (A) UMAP visualization of transcriptionally distinct cell populations from different CM differentiations. (B) UMAP embedding showing annotated cell clusters. (C) Unsupervised clustering of cells in the UMAP space. Cells of clusters 2, 7, 9, and 15 represent VCMs, 0, 6, 8, and 12 are ACMs and cluster 1 are SANCMs. (D) Dot plot showing the top 10 marker genes for VCM, ACM and SANCMs clusters in (C). (E) Feature plots (top) and violin plots (bottom) showing expression of selected canonical genes used to annotate SANCMs. (F) Heatmap of differentially expressed genes between ACMs and SANCMs. (G) Bar plot showing the top 20 enriched pathways in the upregulated DEGs in SANCMs vs ACMs.

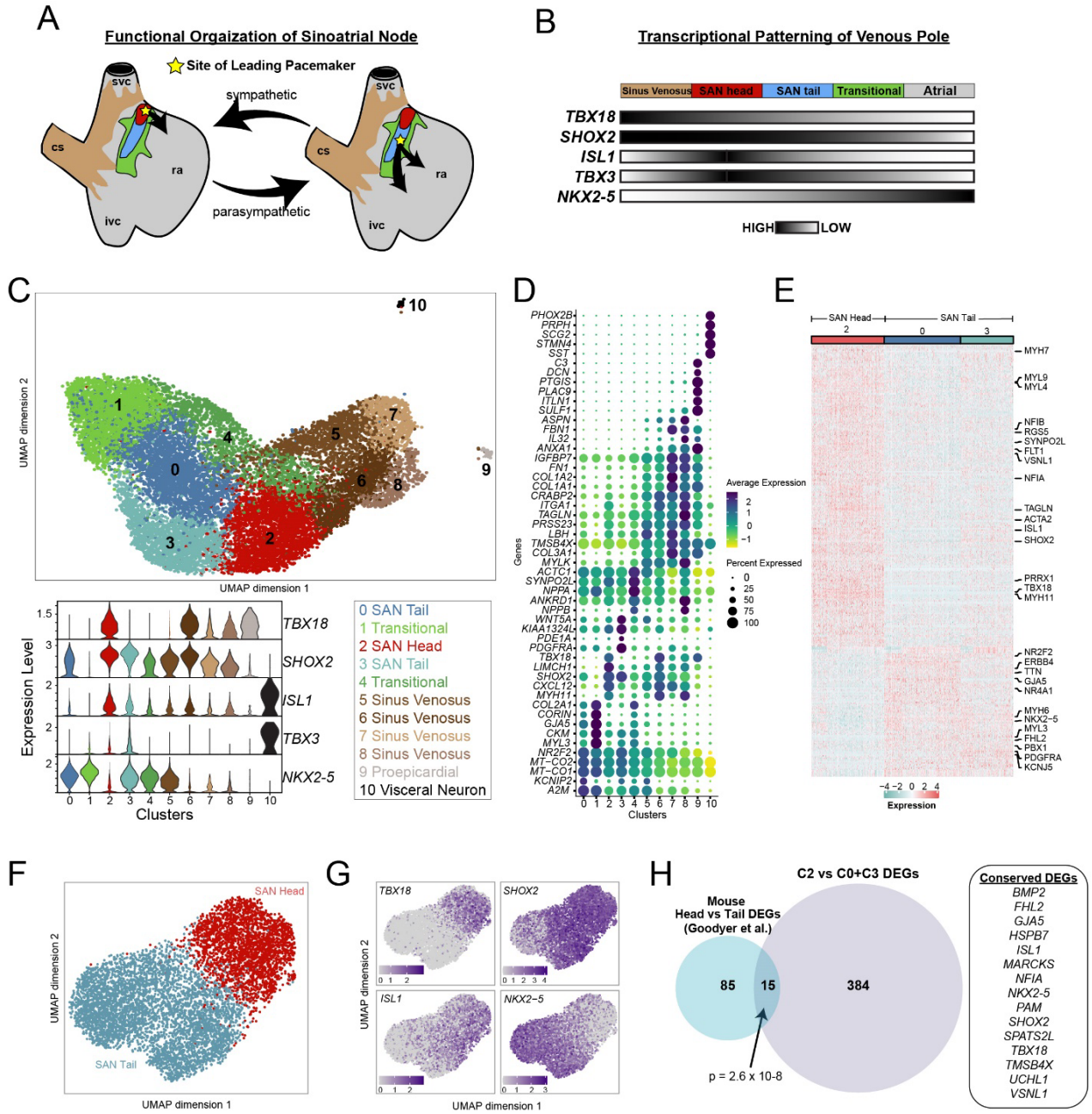


Figure 3: Sub-cluster analysis of day 34 hiPSC derived SANCMs. (A) Illustration showing the anatomical and functional organization of the Sinoatrial Node (SAN). svc; superior vena cava, cs; coronary sinus, ivc; inferior vena cava, ra; right atrium. (B) Diagram depicting the gradient expression of key transcription factors in the patterning of the Venous Pole. (C) UMAP visualization (top) and violin plots showing canonical SANCM marker gene expression in the SANCM clusters (bottom left), and annotation of cell types present (bottom right) in the SANCM differentiation. (D) Dot plot of top 5 marker genes for each cluster in (C). (E) Heatmap of differentially expressed genes between SAN Head cells (cluster 2) and SAN Tail cells (clusters 0 and 3). (F) UMAP of sub-clustered SAN Head cells (cluster 2) and SAN Tail cells (clusters 0 and 3) only. (G) Canonical marker gene expression of sub-clustered SAN Head and SAN Tail cells from (F). (H) Venn diagram showing the overlap of conserved differentially expressed genes (DEGs) from Human iPSC derived SAN Head vs Tail cells (our data) and Mouse *in vivo* isolated SAN Head vs Tail cells (Goodyear et al.).

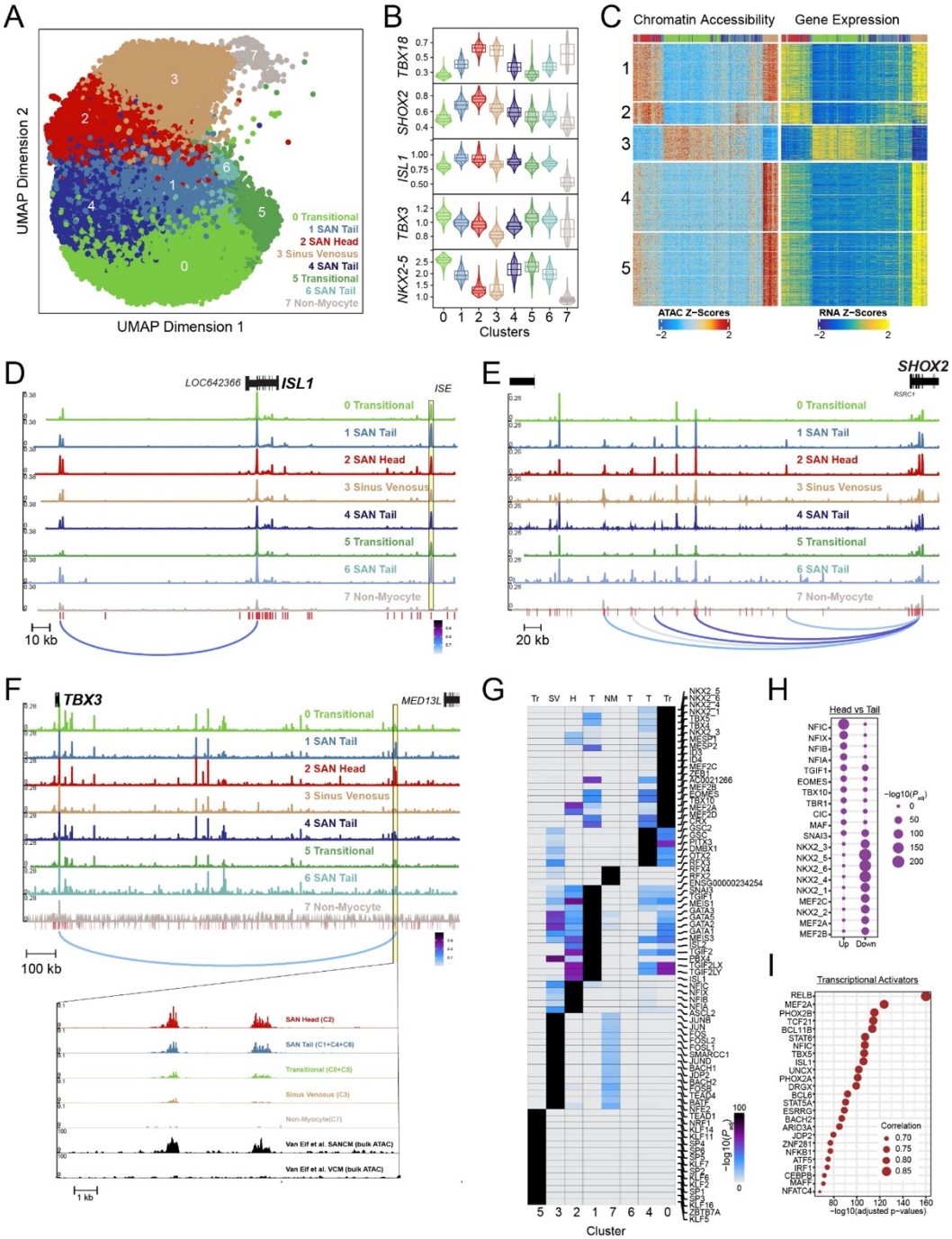


Figure 4: scATAC-seq analysis of day 34 hiPSC derived SANCMs. (A) UMAP projection and cluster annotation of cell types present. (B) Predicted expression of canonical SANCM marker genes in each cluster from scATAC-seq data. (C) Heatmap summary of ArchR peak-to-gene linkages between scATAC and scRNA-seq data of SANCMs (correlation≥0.45). Z-scores for peak accessibility and RNA expression are shown in the left and right heatmaps, respectively. (D-F) Genomic tracks for accessibility and predicated peak-to-gene links around (D) ISL1, (E) SHOX2 and (F) TBX3 loci for all cell types annotated in (A) recapitulate known key regulatory elements in the SAN. (G) Hypergeometric enrichment of top 20 TF binding motifs in marker peaks of each cluster. (H) Top 10 upregulated and downregulated transcription factor binding motifs in SAN Head cells (cluster 2) versus SAN Tail cells (clusters 1,4 and 6). (I) Top 25 predicated positive regulatory transcription factors whose gene expression is positively correlated to changes in the accessibility of their corresponding motif.

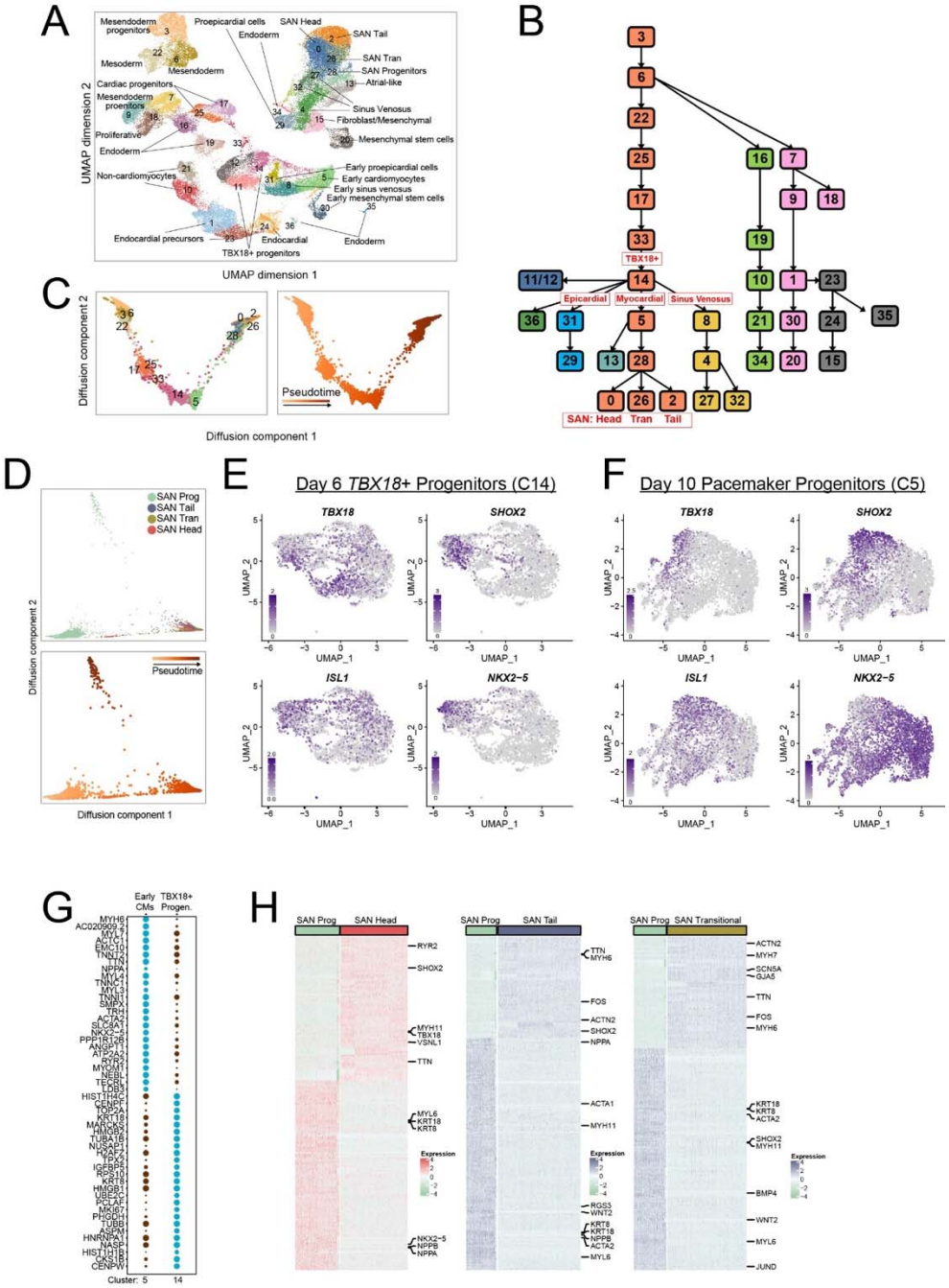


Figure 5: Time course scRNA-seq analysis of hiPSC derived SANCMs. (A) UMAP plot of unsupervised clustering with annotated cell types. (B) Schematic of predicted differentiation lineages inferred from PAGA analysis. (C) Diffusion map of clusters in the SAN cell lineage from day 3, 6, 10, 23, and 34 of differentiation with cells colored by clusters (left) and pseudotime (right). (D) Diffusion map of clusters in the SAN cell lineage from later stages of differentiation (day 23 and 34) with cells colored by clusters (top) and pseudotime (bottom). (E) UMAPs showing canonical SAN marker genes after sub-clustering of TBX18+ progenitors (cluster 14 of fig 5A) from day 6 of differentiation. (F) UMAPs showing canonical SAN marker genes after sub-clustering of Pacemaker progenitors (cluster 5 of fig 5A) from day 10 of differentiation. (G) Dot plot of top 25 upregulated and downregulated DEGs in Early CMs (cluster 5) vs TBX18+ Progenitors (cluster 14). (H) Heatmaps of DEGs from SAN Progenitors (cluster 28) vs SAN Head (clusters 2 of fig 3C; left), SAN Tail (cluster 0 and 3 of fig 3C, center), and SAN Transitional (clusters 1 and 4 of fig 3C, right), respectively.

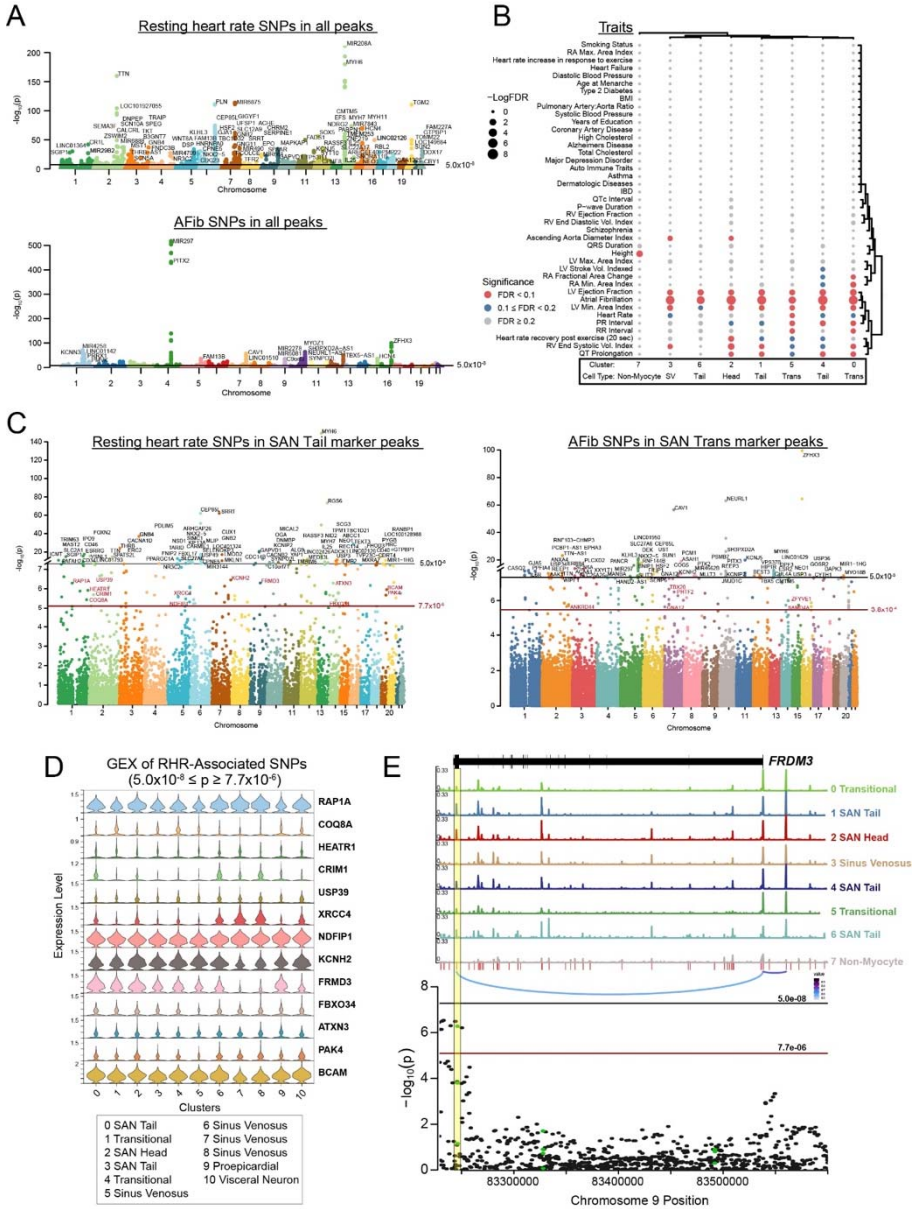
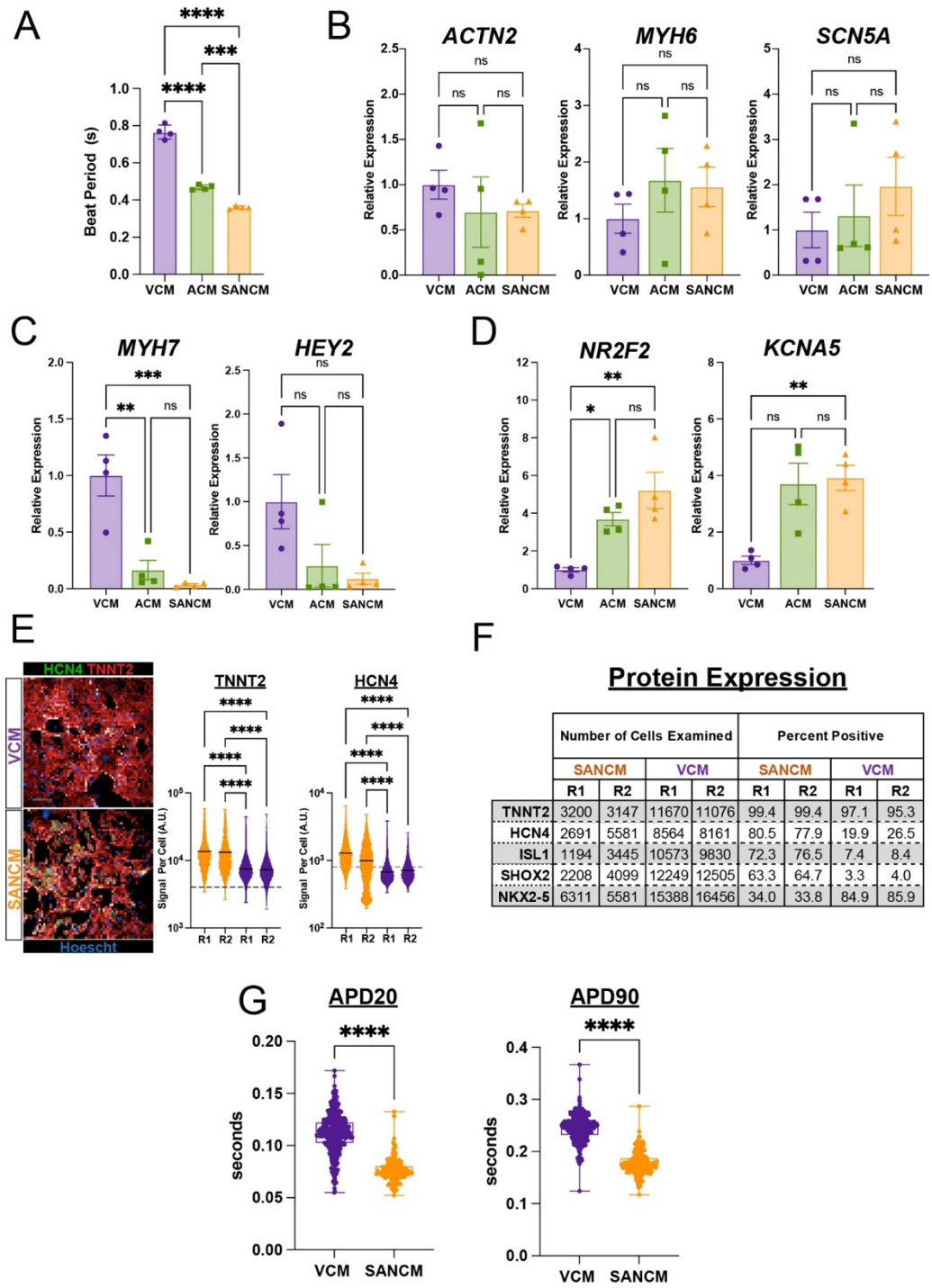


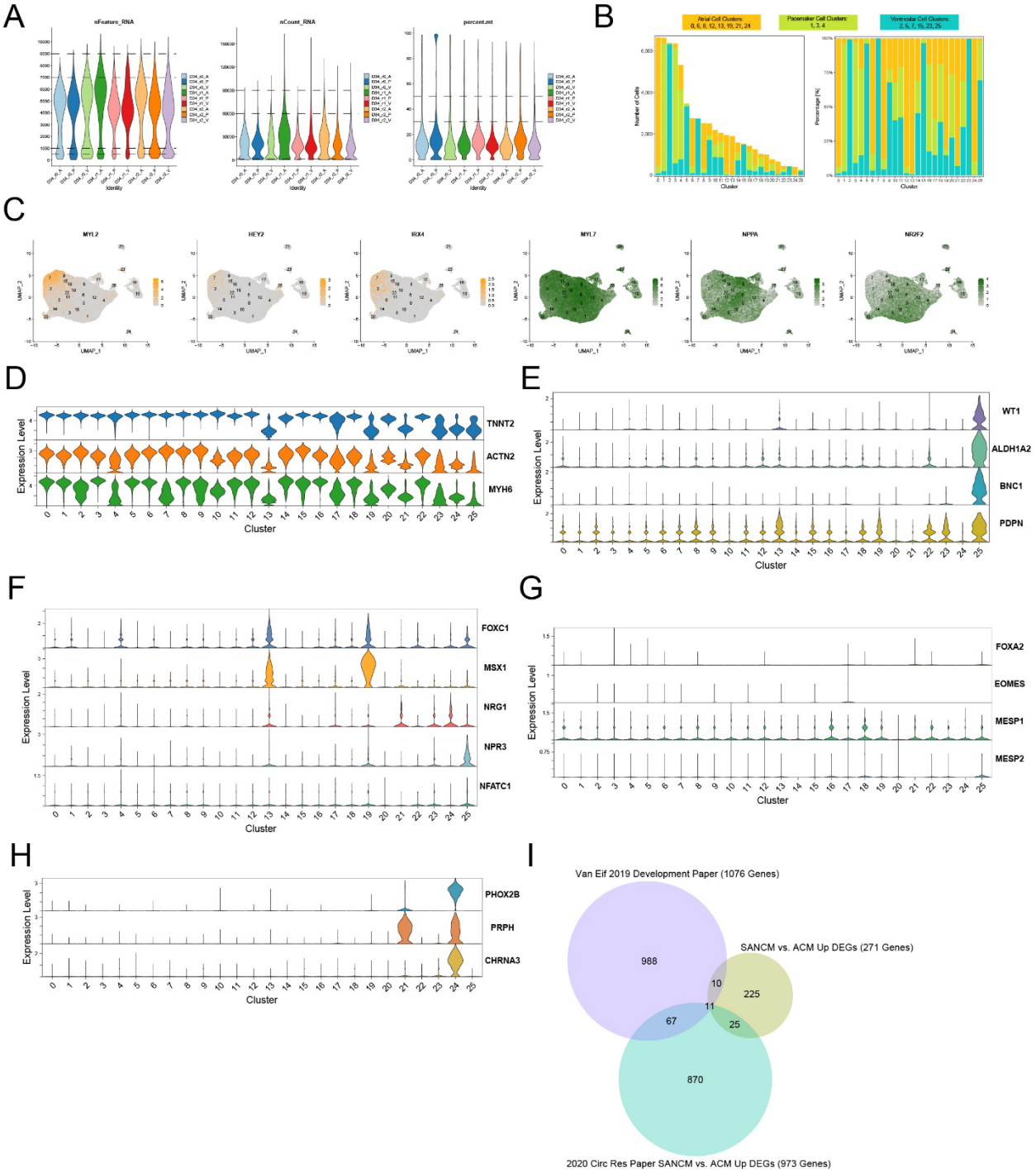
Figure 6: Integrative analysis of scATAC-seq data and GWAS. (A) Manhattan plots showing association of resting heart rate (top) and atrial fibrillation (bottom) SNPs ascertained from the UK Biobank cohort within all ATAC-seq peaks from our study. (B) Dot plot showing association of SAN cell types with complex traits and diseases. Cell-type-stratified linkage disequilibrium score regression (LDSC) analysis was performed using GWAS summary statistics for 41 phenotypes. All ATAC peaks identified from each fetal and adult cell type by human enhancer atlas (http://catlas.org/humanenhancer) were used as background for analysis. P-values were corrected using the Benjamini-Hochberg procedure for multiple tests. FDRs of LDSC coefficient are displayed. (C) Manhattan plots showing association of resting heart rate SNPs within SAN Tail marker peaks (left) and atrial fibrillation SNPs within SAN Transitional marker peaks (right) ascertained from the UK Biobank cohort. Red line indicates significance threshold of Bonferroni adjusted P value of 0.05 after pruning the SNPs with LD (R^2 > 0.1). Red text indicates novel gene associated SNPs identified in our analyses. (D) Violin plots of novel gene expression in all scRNA-seq clusters identified from SNPs associated with resting heart rate in SAN Tail marker peaks using adjusted p-values from our study. (E) Genomic tracks for accessibility and predicated peak-to-gene links around the FRDM3 locus for all cell types annotated in Fig. 4 illustrate the potential of using our single cell multi-omics dataset to identify previously unknown genes associated with heritable traits.



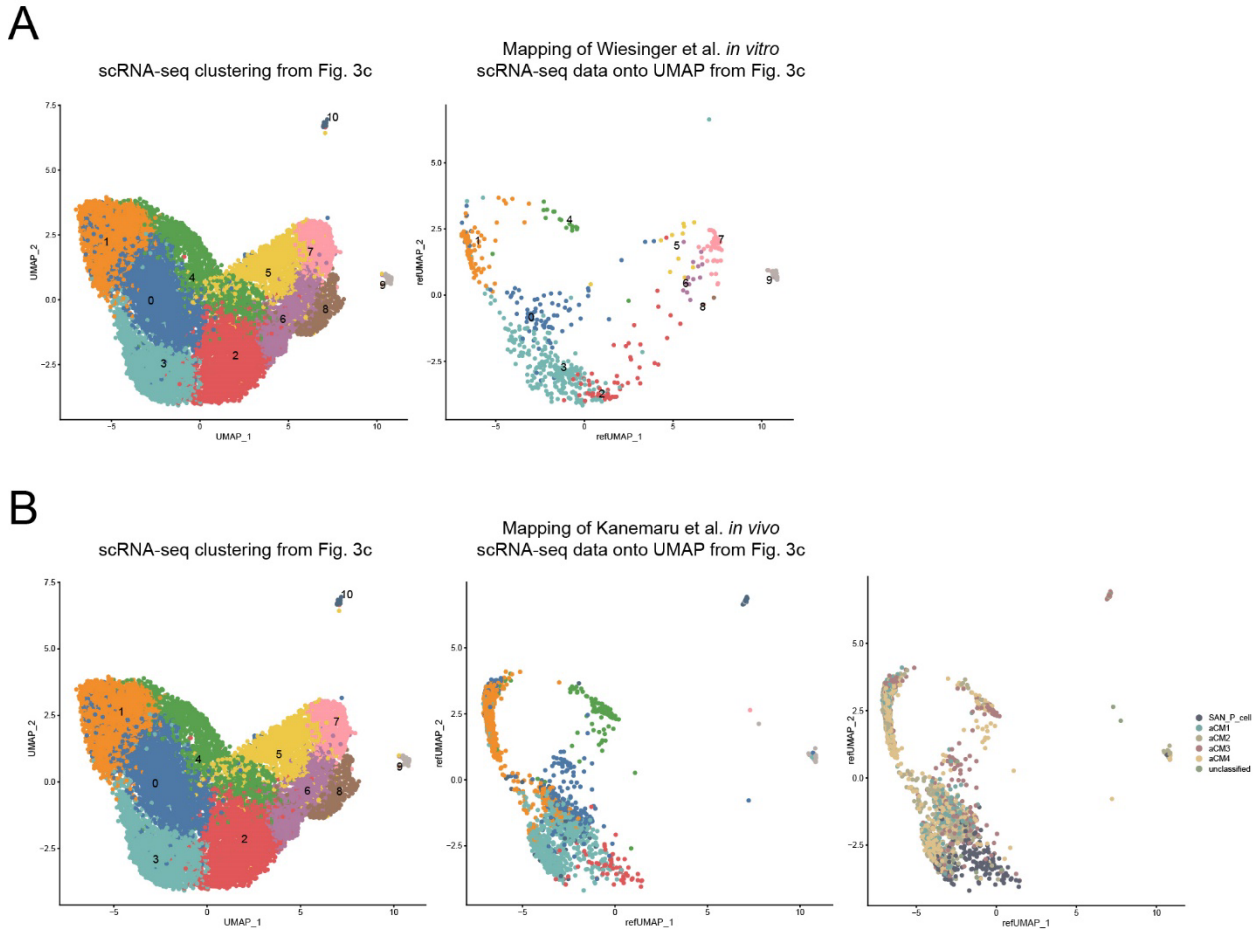
Protein Expression

	Number of Cells Examined				Percent Positive			
	SANCM		VCM		SANCM		VCM	
	R1	R2	R1	R2	R1	R2	R1	R2
TNNT2	3200	3147	11670	11076	99.4	99.4	97.1	95.3
HCN4	2691	5581	8564	8161	80.5	77.9	19.9	26.5
ISL1	1194	3445	10573	9830	72.3	76.5	7.4	8.4
SHOX2	2208	4099	12249	12505	63.3	64.7	3.3	4.0
NKX2-5	6311	5581	15388	16456	34.0	33.8	84.9	85.9

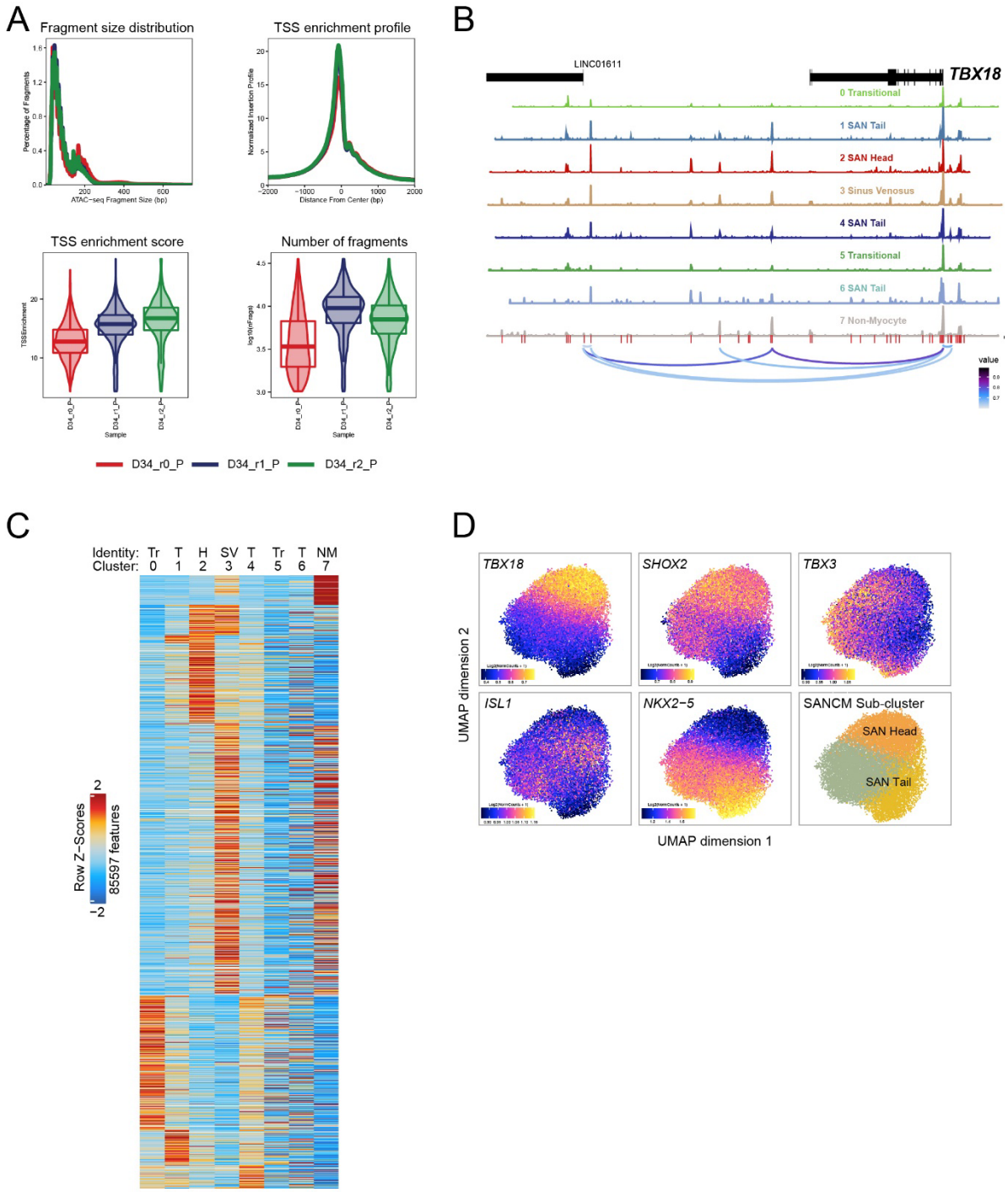
Extended Data Figure 1: Characterization of hiPSC derived Cardiomyocyte-like subtypes. (A) Beat period as measured by MEA in VCMs, ACMs and SANCMs. (B-D) qPCR for extended (B) pan-cardiac, (C) ventricular and (D) atrial genes in hiPSC derived Cardiomyocyte subtypes. (E) Immunocytochemistry of HCN4 (green) and TNNT2 (red) expression in hiPSC derived VCMs and SANCMs counterstained with Hoescht (blue) and quantification of signal intensities in two replicates. Dotted lines indicate threshold cutoffs for quantification. (F) Table quantifying percent of VCMs and SANCMs expressing TNNT2, HCN4, ISL1, SHOX2 and NKX2-5 from two replicates of immunocytochemistry experiments. (G) ADP20 and ADP90 measurements recorded from VCMs and SANCMs using FluoVolt. Significance was determined using one-way ANOVA with multiple comparisons tests. * $p<0.05$ ** $p<0.01$ *** $p<0.001$ **** $p<0.0001$



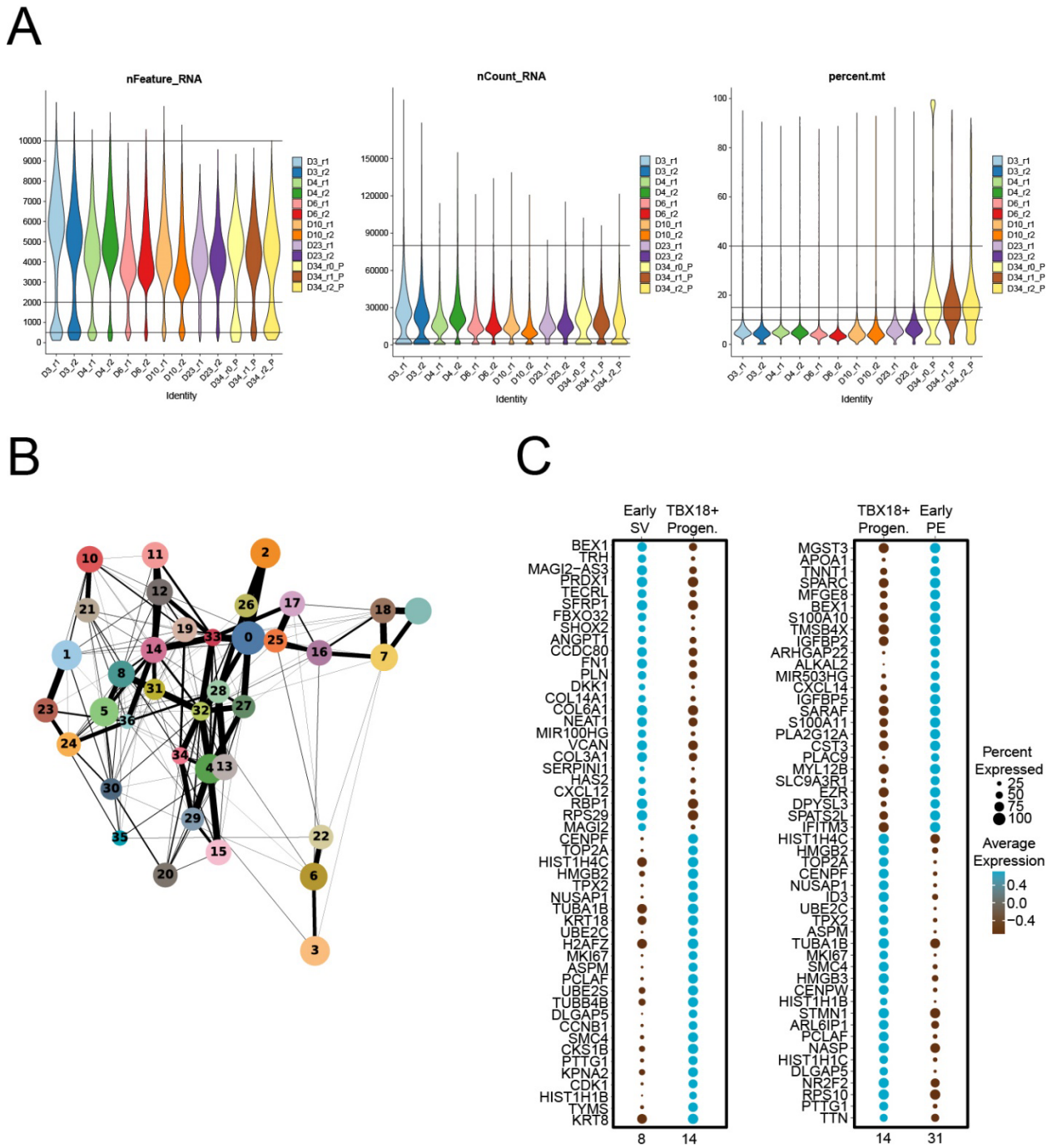
Extended Data Figure 2: scRNA-seq of day 34 hiPSC derived VCMs, ACMs and SANCMs. (A) Summary of scRNA-seq data metrics. (B) Cluster assignments of Fig. 2c based on cell type composition of each cluster. (C) Feature plots showing expression of selected canonical genes used to annotate VCMs and ACMs in Fig. 2c. (D-H) Violin plots showing expression of canonical (D) pan-myocardial, (E) proepicardial, (F) endocardial, (G) primitive mesendoderm, and (H) visceral neuron markers in all clusters in Fig. 2c. (I) Venn diagram showing the comparison of differentially expressed genes in SANCMs vs ACMs with DEGs identified in previously published neonatal mouse and human fetal SAN vs Right atrium datasets.



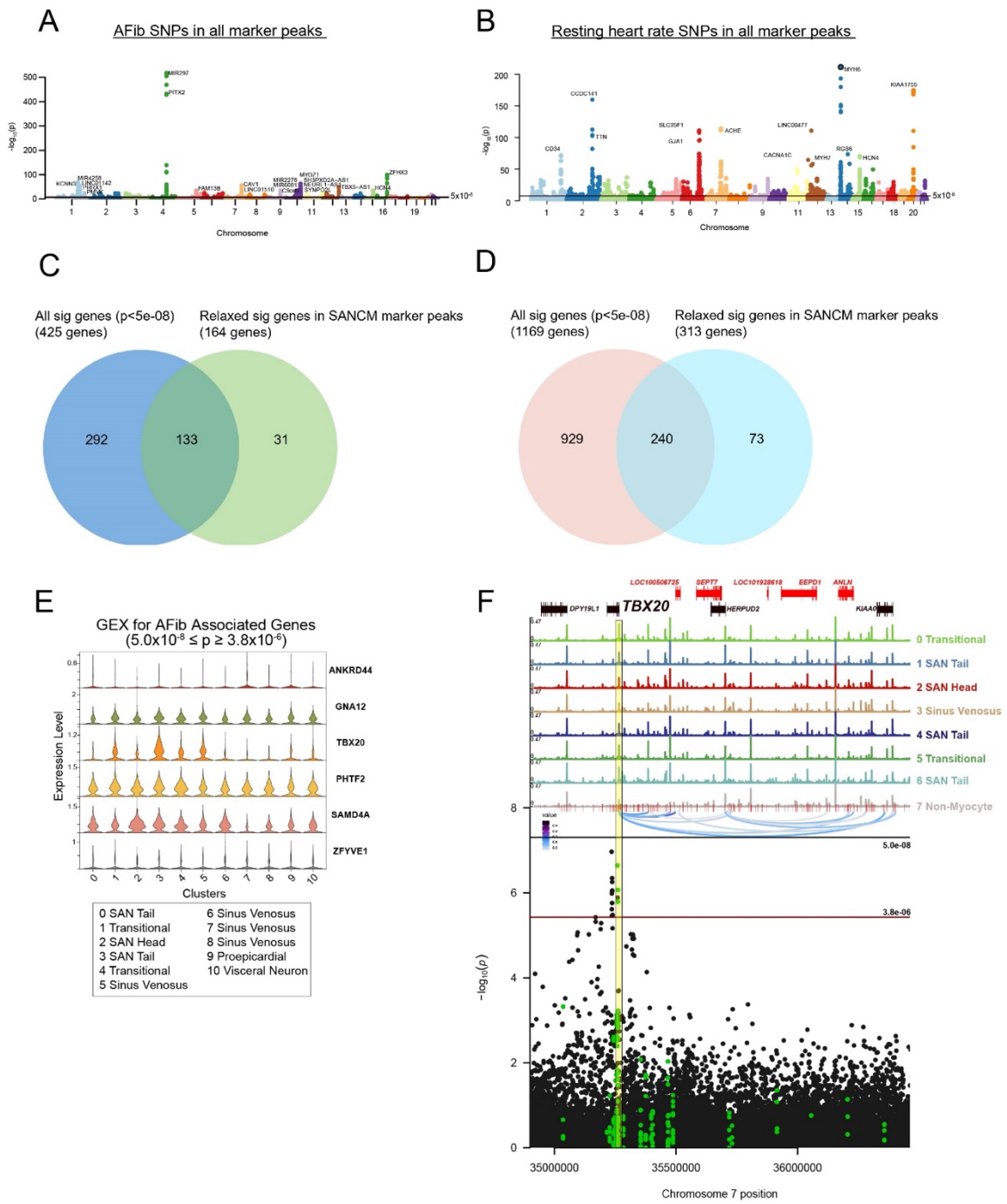
Extended Data Figure 3: Sub-cluster comparison of hiPSC derived SANCMs to other *in vitro* and *in vivo* derived scRNA-seq studies. (A) Mapping of cells from *in vitro* scRNA-seq data from Wiesinger et al. [ref 19] into UMAP scRNA-seq cluster data generated in our study. (B) Mapping of cells from *in vivo* scRNA-seq data from Kanemaru et al. [ref 29] into UMAP scRNA-seq cluster data generated in our study.



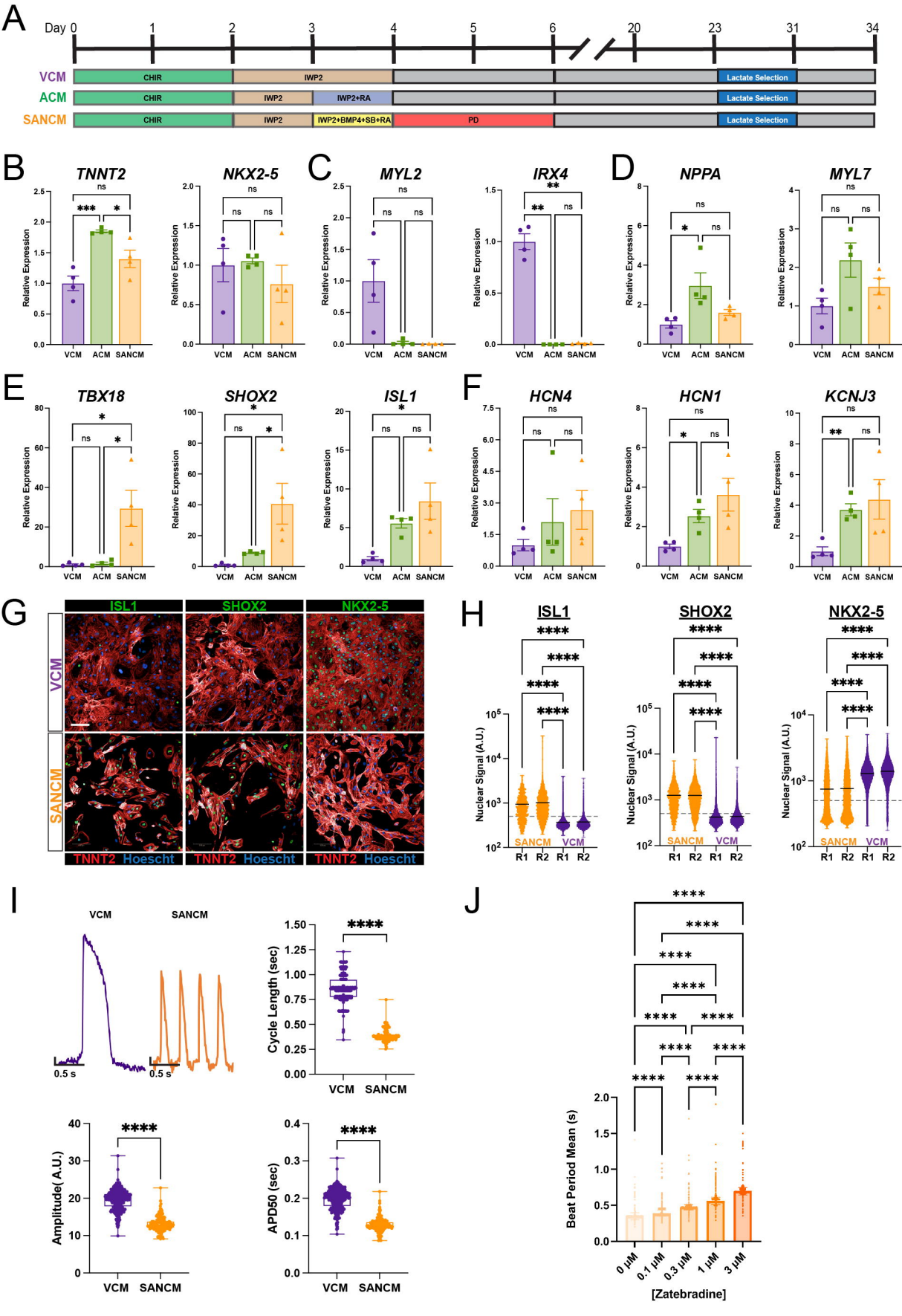
Extended Data Figure 4: scATAC-seq analysis of day 34 hiPSC derived SANCMs. (A) Summary of scATAC-seq data metrics. (B) Genomic tracks for accessibility and predicated peak-to-gene links around the TBX18 locus for all cell types annotated in Fig. 4a. (C) Heatmap of marker peaks for each cluster in Fig. 4a. (D) Feature plots showing predicted expression of SANCM canonical genes in sub-clustering analysis of SAN Head and SAN Tail only clusters from Fig. 4a.

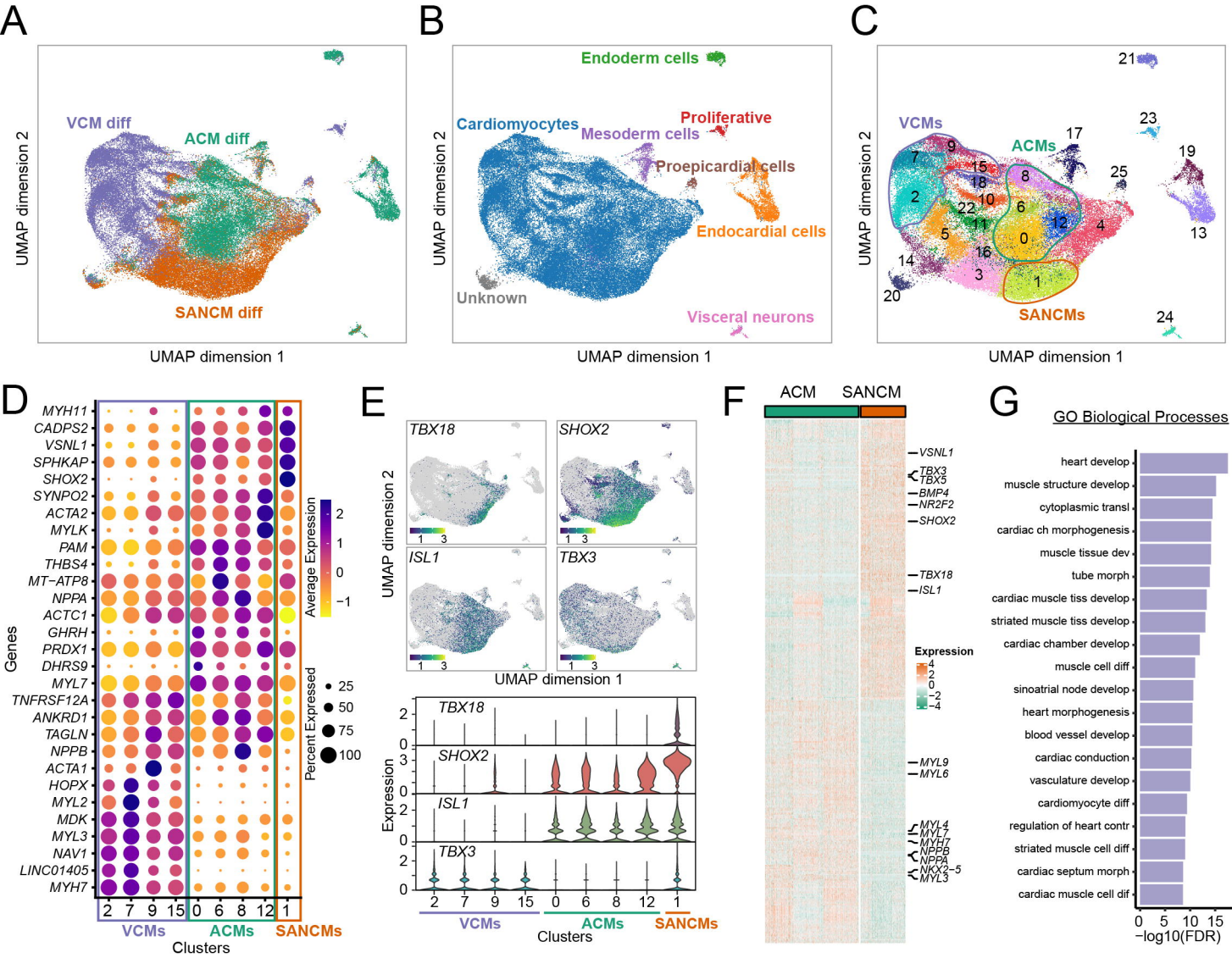


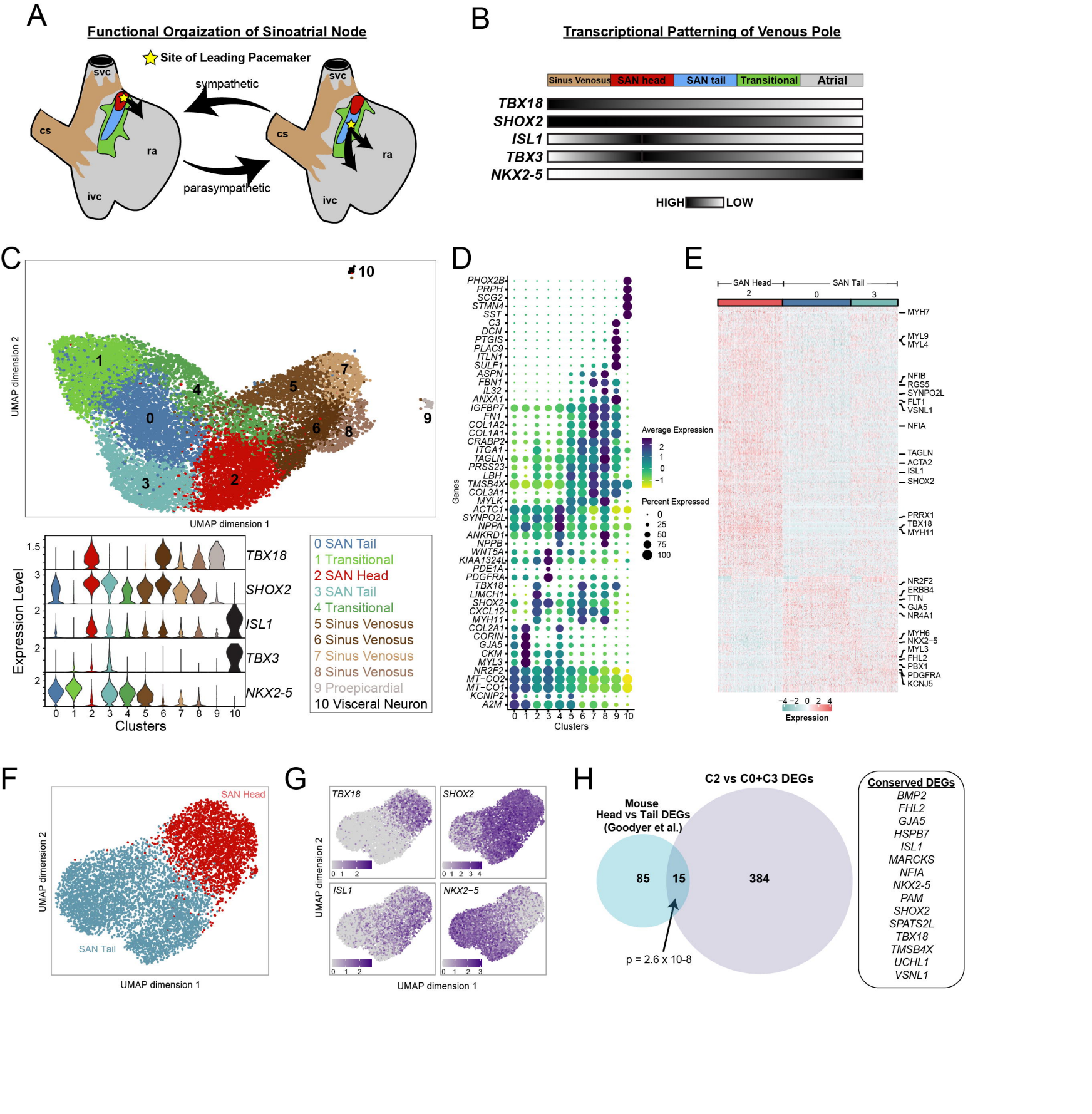
Extended Data Figure 5: Time course scRNA-seq analysis of hiPSC derived SANCMs. (A) Summary of time course scATAC-seq data metrics. (B) Similarity-based connectivity map generated by PAGA analysis. Each cluster is represented as a node and the weighted edges between nodes represent their connectivity. (C) Dot plot of top 25 upregulated and downregulated DEGs in Early Sinus Venosus (SV) (cluster 8) vs TBX18+ Progenitors (cluster 14) (left) and dot plot of top 25 upregulated and downregulated DEGs in Early Proepicardial (PE) (cluster 31) vs TBX18+ Progenitors (cluster 14) (right).

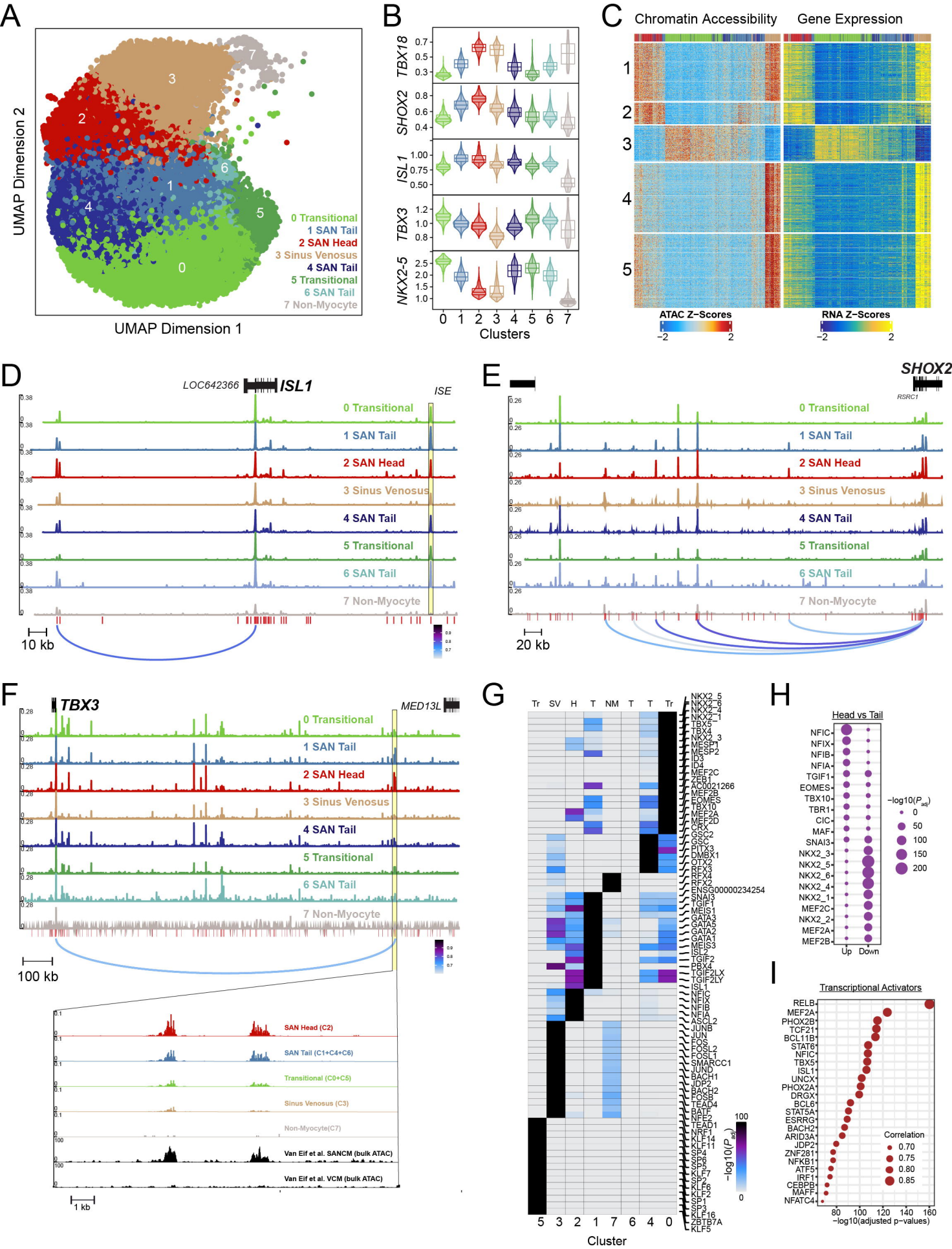


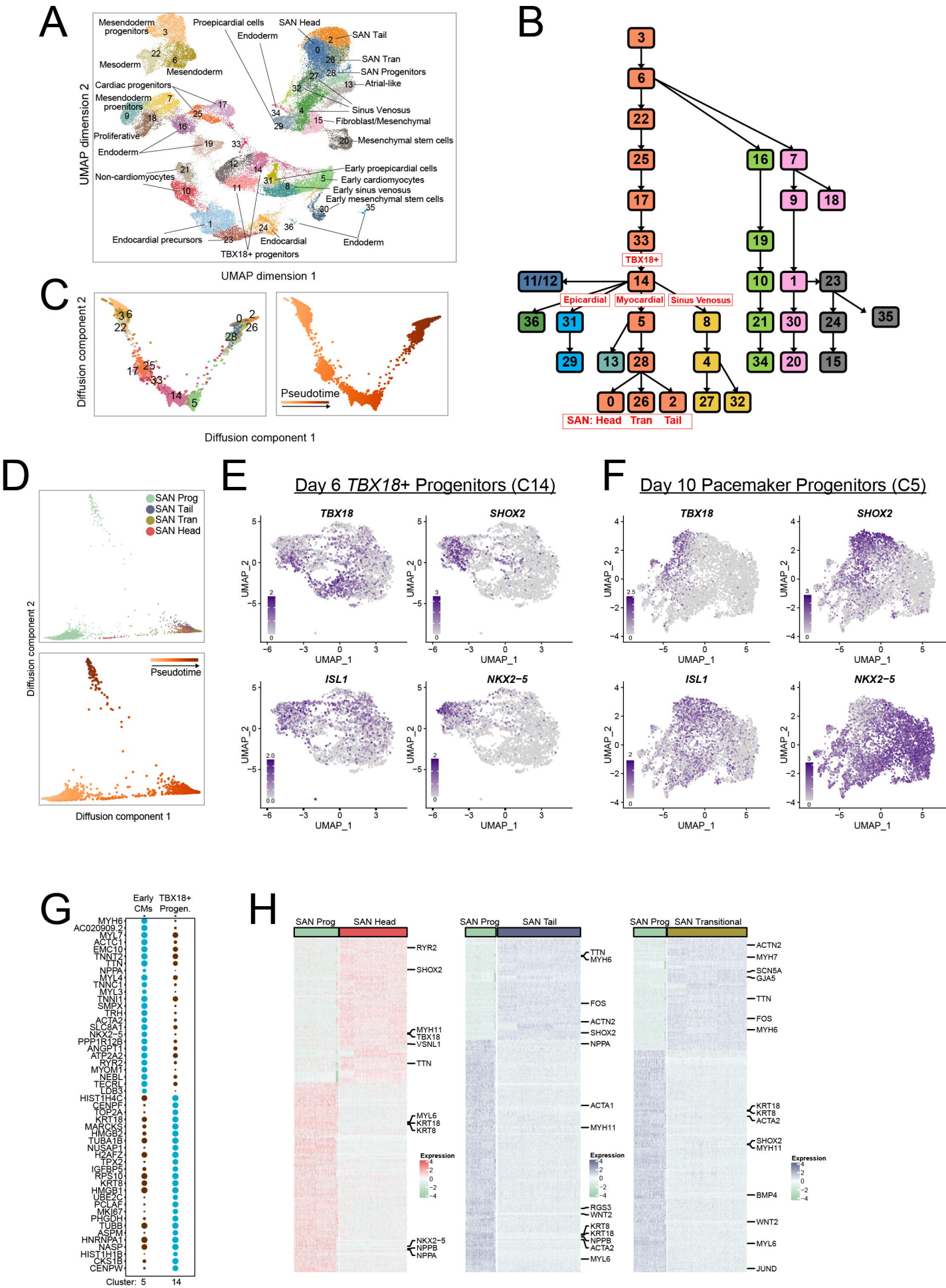
Extended Data Figure 6: Integrative analysis of scATAC-seq data and GWAS. (A-B) Manhattan plots showing association of SNPs within ATAC-seq marker peaks in (A) all SAN cell types with atrial fibrillation and (B) all SAN cell types with resting heart rate ascertained from the UK Biobank cohort. (C-D) Venn diagrams depicting the number of overlapping genome-wide significant genes in (C) atrial fibrillation GWAS studies and (D) heart rate GWAS studies with marker peaks of SANCM clusters in this study. (E) Violin plots of novel gene expression in all scRNA-seq clusters identified from SNPs associated with atrial fibrillation in SAN Transitional marker peaks using adjusted p-values from our study. (F) Genomic tracks for accessibility and predicated peak-to-gene links around the TBX20 locus for all cell types annotated in Fig. 4 illustrate the potential of using our single cell multi-omics dataset to identify previously unknown genes associated with heritable traits.

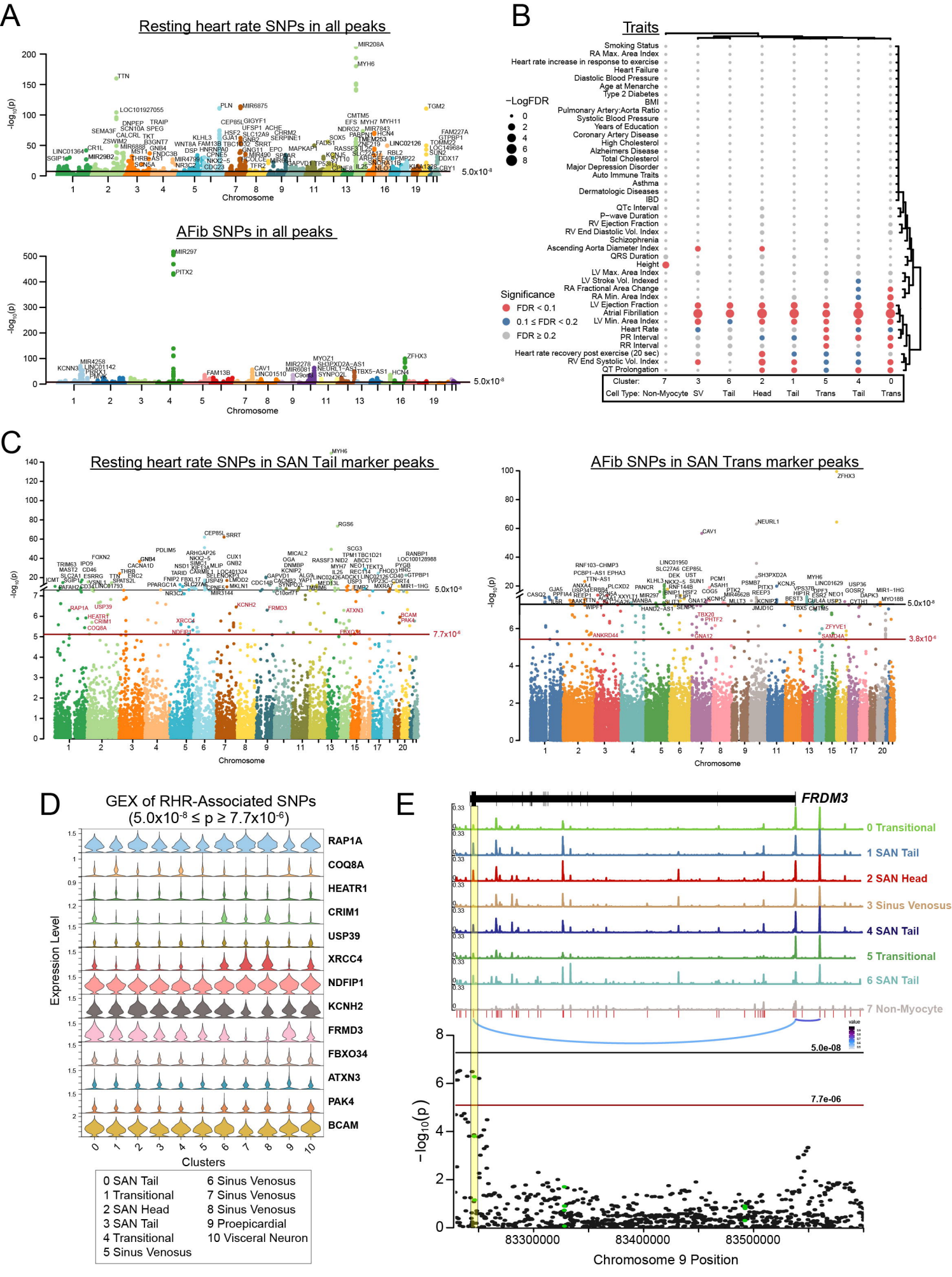

A
Day 0
1
2
3
4
5
6
20
23
31
34
VCM
CHIR
IWP2
Lactate Selection
ACM
CHIR
IWP2
IWP2+RA
Lactate Selection
SANCM
CHIR
IWP2
IWP2+BMP4+SB+RA
PD
Lactate Selection
B
TNNT2
NKX2-5
C
MYL2
IRX4
D
NPPA
MYL7
E
TBX18
SHOX2
ISL1
F
HCN4
HCN1
KCNJ3
Relative Expression
VCM
ACM
SANCM
G
ISL1
SHOX2
NKX2-5
VCM
SANCM
TNNT2 Hoescht
H
ISL1
SHOX2
NKX2-5
Nuclear Signal (A.U.)
SANCM
VCM
R1
R2
I
VCM
SANCM
0.5 s
Cycle Length (sec)
Amplitude(A.U.)
APD50 (sec)
J
Beat Period Mean (s)
0 μM
0.1 μM
0.3 μM
1 μM
3 μM
[Zatebradine]

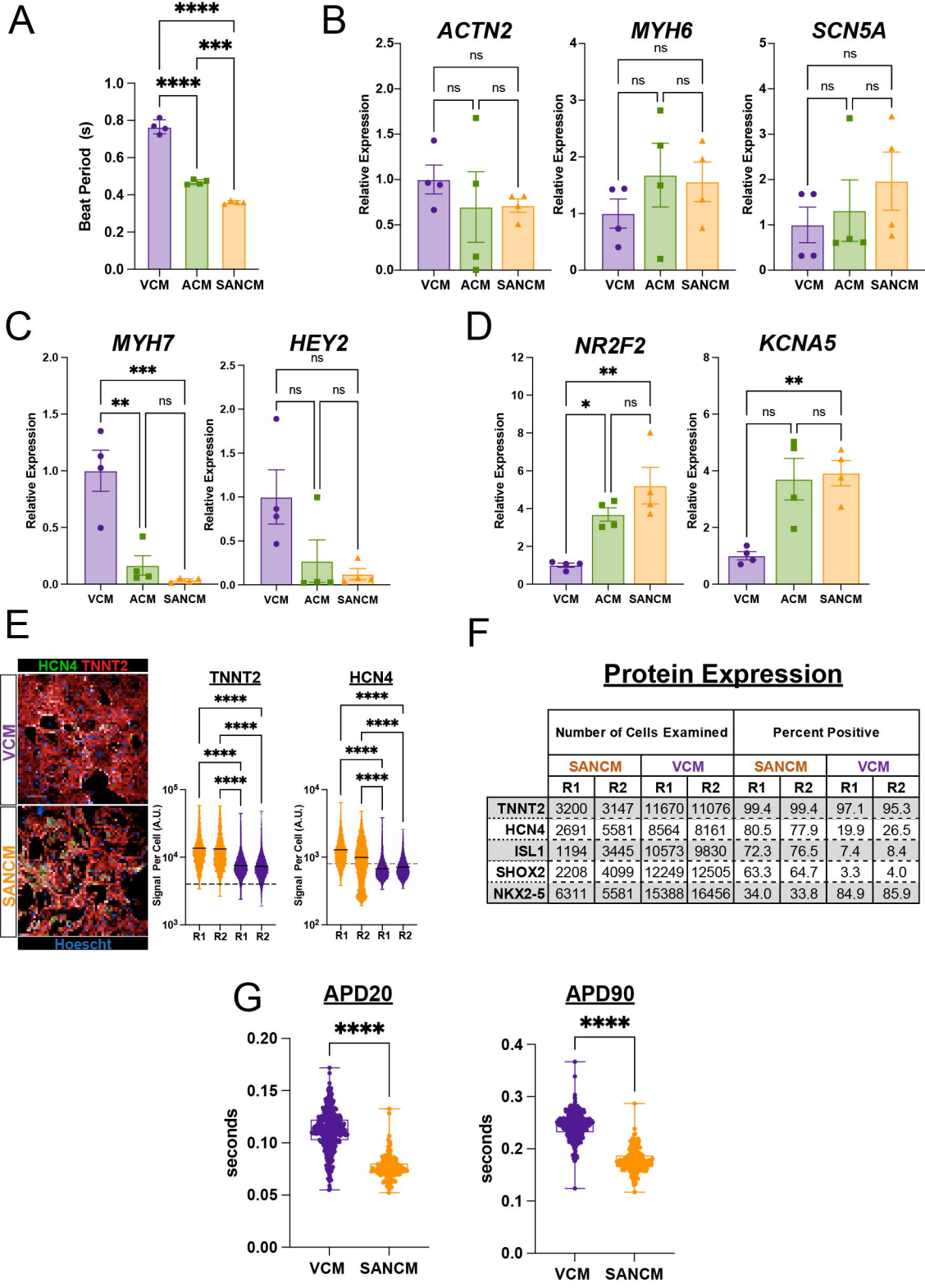

A
VCM diff
ACM diff
SANCM diff
UMAP dimension 2
UMAP dimension 1
B
Endoderm cells
Cardiomyocytes
Mesoderm cells
Proliferative
Proepicardial cells
Endocardial cells
Unknown
Visceral neurons
UMAP dimension 2
UMAP dimension 1
C
VCMs
ACMs
SANCMs
UMAP dimension 2
UMAP dimension 1
D
MYH11
CADPS2
VSNL1
SPHKAP
SHOX2
SYNPO2
ACTA2
MYLK
PAM
THBS4
MT−ATP8
NPPA
ACTC1
GHRH
PRDX1
DHRS9
MYL7
TNFRSF12A
ANKRD1
TAGLN
NPPB
ACTA1
HOPX
MYL2
MDK
MYL3
NAV1
LINC01405
MYH7
Genes
Average Expression
Percent Expressed
VCMs
ACMs
SANCMs
Clusters
E
TBX18
SHOX2
ISL1
TBX3
UMAP dimension 2
UMAP dimension 1
Expression
VCMs
ACMs
SANCMs
Clusters
F
ACM
SANCM
VSNL1
TBX3
TBX5
BMP4
NR2F2
SHOX2
TBX18
ISL1
MYL9
MYL6
MYL4
MYL7
MYH7
NPPB
NPPA
NKX2−5
MYL3
Expression
G
GO Biological Processes
heart develop
muscle structure develop
cytoplasmic transl
cardiac ch morphogenesis
muscle tissue dev
tube morph
cardiac muscle tiss develop
striated muscle tiss develop
cardiac chamber develop
muscle cell diff
sinoatrial node develop
heart morphogenesis
blood vessel develop
cardiac conduction
vasculature develop
cardiomyocyte diff
regulation of heart contr
striated muscle cell diff
cardiac septum morph
cardiac muscle cell dif
−log10(FDR)


A
Functional Orgaization of Sinoatrial Node
Site of Leading Pacemaker
sympathetic
parasympathetic
svc
cs
ra
ivc
B
Transcriptional Patterning of Venous Pole
Sinus Venosus
SAN head
SAN tail
Transitional
Atrial
TBX18
SHOX2
ISL1
TBX3
NKX2-5
HIGH
LOW
C
UMAP dimension 1
UMAP dimension 2
Expression Level
Clusters
TBX18
SHOX2
ISL1
TBX3
NKX2-5
0 SAN Tail
1 Transitional
2 SAN Head
3 SAN Tail
4 Transitional
5 Sinus Venosus
6 Sinus Venosus
7 Sinus Venosus
8 Sinus Venosus
9 Proepicardial
10 Visceral Neuron
D
Genes
Clusters
Average Expression
Percent Expressed
E
SAN Head
SAN Tail
Expression
F
SAN Head
SAN Tail
UMAP dimension 1
UMAP dimension 2
G
TBX18
SHOX2
ISL1
NKX2-5
UMAP dimension 1
UMAP dimension 2
H
C2 vs C0+C3 DEGs
Mouse
Head vs Tail DEGs
(Goodyer et al.)
85
15
384
p = 2.6 x 10-8
Conserved DEGs
BMP2
FHL2
GJA5
HSPB7
ISL1
MARCKS
NFIA
NKX2-5
PAM
SHOX2
SPATS2L
TBX18
TMSB4X
UCHL1
VSNL1


A
UMAP Dimension 2
UMAP Dimension 1
0 Transitional
1 SAN Tail
2 SAN Head
3 Sinus Venosus
4 SAN Tail
5 Transitional
6 SAN Tail
7 Non-Myocyte
B
TBX18
SHOX2
ISL1
TBX3
NKX2-5
Clusters
C
Chromatin Accessibility
Gene Expression
ATAC Z-Scores
RNA Z-Scores
D
LOC642366
ISL1
ISE
10 kb
E
SHOX2
RSRC1
20 kb
F
TBX3
MED13L
100 kb
SAN Head (C2)
SAN Tail (C1+C4+C6)
Transitional (C0+C5)
Sinus Venosus (C3)
Non-Myocyte(C7)
Van Eif et al. SANCM (bulk ATAC)
Van Eif et al. VCM (bulk ATAC)
1 kb
G
Cluster
H
Head vs Tail
Up
Down
I
Transcriptional Activators
Correlation
−log10(adjusted p−values)

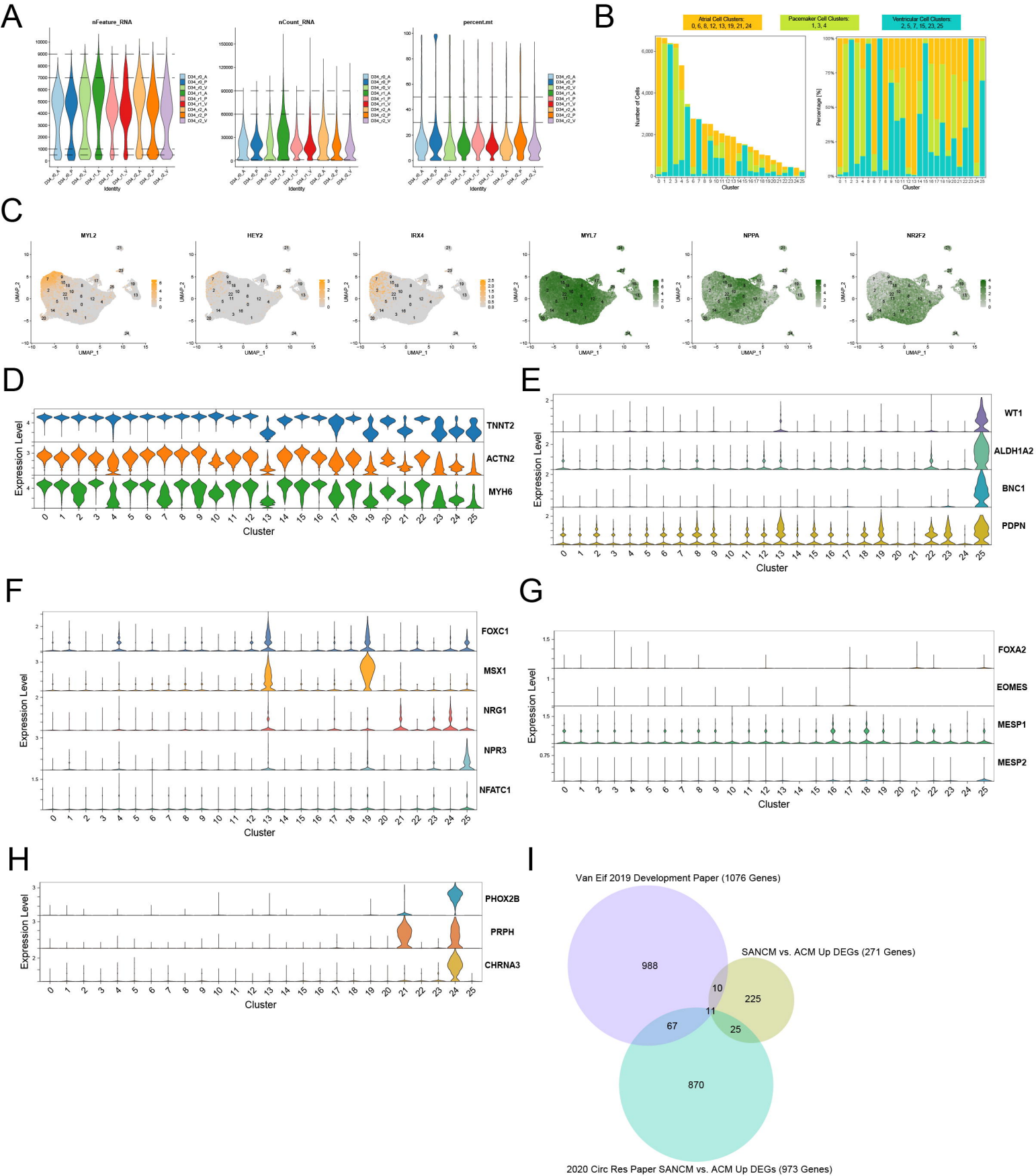

A
Mesendoderm progenitors
Proepicardial cells
SAN Head
Endoderm
SAN Tail
Mesoderm
Mesendoderm
SAN Tran
SAN Progenitors
Atrial-like
Cardiac progenitors
Mesendoderm proenitors
Sinus Venosus
Fibroblast/Mesenchymal
Proliferative
Endoderm
Mesenchymal stem cells
Early proepicardial cells
Non-cardiomyocytes
Early cardiomyocytes
Early sinus venosus
Early mesenchymal stem cells
Endocardial precursors
Endocardial
Endoderm
TBX18+ progenitors
UMAP dimension 2
UMAP dimension 1
B
TBX18+
Epicardial
Myocardial
Sinus Venosus
SAN: Head Tran Tail
C
Diffusion component 2
Pseudotime
Diffusion component 1
D
SAN Prog
SAN Tail
SAN Tran
SAN Head
Pseudotime
Diffusion component 2
Diffusion component 1
E
Day 6 TBX18+ Progenitors (C14)
TBX18
SHOX2
ISL1
NKX2−5
UMAP_1
UMAP_2
F
Day 10 Pacemaker Progenitors (C5)
TBX18
SHOX2
ISL1
NKX2−5
UMAP_1
UMAP_2
G
Early CMs
TBX18+ Progen.
MYH6
AC020909.2
MYL7
ACTC1
EMC10
TNNT2
TTN
NPPA
MYL4
TNNC1
MYL3
TNNI1
SMPX
TRH
ACTA2
SLC8A1
NKX2−5
PPP1R12B
ANGPT1
ATP2A2
RYR2
MYOM1
NEBL
TECRL
LDB3
HIST1H4C
CENPF
TOP2A
KRT18
MARCKS
HMGB2
TUBA1B
NUSAP1
H2AFZ
TPX2
IGFBP5
RPS10
KRT8
HMGB1
UBE2C
PCLAF
MKI67
PHGDH
TUBB
ASPM
HNRNPA1
NASP
HIST1H1B
CKS1B
CENPW
Cluster: 5 14
H
SAN Prog
SAN Head
RYR2
SHOX2
MYH11
TBX18
VSNL1
TTN
MYL6
KRT18
KRT8
NKX2−5
NPPB
NPPA
Expression
SAN Prog
SAN Tail
TTN
MYH6
FOS
ACTN2
SHOX2
NPPA
ACTA1
MYH11
RGS5
WNT2
KRT8
KRT18
NPPB
ACTA2
MYL6
Expression
SAN Prog
SAN Transitional
ACTN2
MYH7
SCN5A
GJA5
TTN
FOS
MYH6
KRT18
KRT8
ACTA2
SHOX2
MYH11
BMP4
WNT2
MYL6
JUND
Expression


A
Resting heart rate SNPs in all peaks
AFib SNPs in all peaks
Chromosome
B
Traits
Smoking Status
RA Max. Area Index
Heart rate increase in response to exercise
Heart Failure
Diastolic Blood Pressure
Age at Menarche
Type 2 Diabetes
BMI
Pulmonary Artery:Aorta Ratio
Systolic Blood Pressure
Years of Education
Coronary Artery Disease
High Cholesterol
Alzheimers Disease
Total Cholesterol
Major Depression Disorder
Auto Immune Traits
Asthma
Dermatologic Diseases
IBD
QTc Interval
P-wave Duration
RV Ejection Fraction
RV End Diastolic Vol. Index
Schizophrenia
Ascending Aorta Diameter Index
QRS Duration
Height
LV Max. Area Index
LV Stroke Vol. Indexed
RA Fractional Area Change
RA Min. Area Index
LV Ejection Fraction
Atrial Fibrillation
LV Min. Area Index
Heart Rate
PR Interval
RR Interval
Heart rate recovery post exercise (20 sec)
RV End Systolic Vol. Index
QT Prolongation
−LogFDR
Significance
FDR < 0.1
0.1 ≤ FDR < 0.2
FDR ≥ 0.2
Cluster: 7 3 6 2 1 5 4 0
Cell Type: Non-Myocyte SV Tail Head Tail Trans Tail Trans
C
Resting heart rate SNPs in SAN Tail marker peaks
AFib SNPs in SAN Trans marker peaks
Chromosome
D
GEX of RHR-Associated SNPs
(5.0x10-8 ≤ p ≥ 7.7x10-6)
RAP1A
COQ8A
HEATR1
CRIM1
USP39
XRCC4
NDFIP1
KCNH2
FRMD3
FBXO34
ATXN3
PAK4
BCAM
Expression Level
Clusters
0 SAN Tail
1 Transitional
2 SAN Head
3 SAN Tail
4 Transitional
5 Sinus Venosus
6 Sinus Venosus
7 Sinus Venosus
8 Sinus Venosus
9 Proepicardial
10 Visceral Neuron
E
FRDM3
0 Transitional
1 SAN Tail
2 SAN Head
3 Sinus Venosus
4 SAN Tail
5 Transitional
6 SAN Tail
7 Non-Myocyte
5.0e-08
7.7e-06
−log10(p)
Chromosome 9 Position



Protein Expression

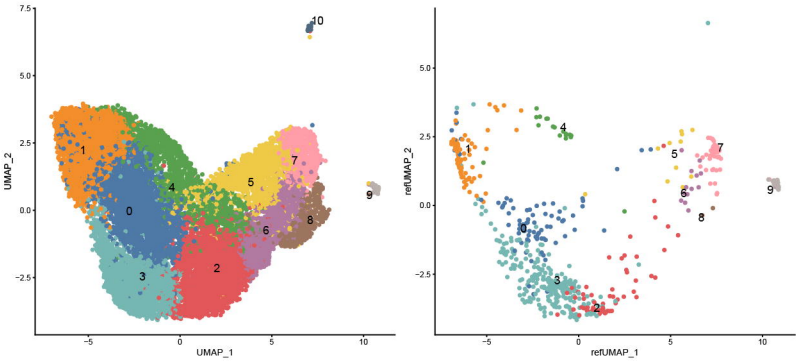
	Number of Cells Examined				Percent Positive			
	SANCM		VCM		SANCM		VCM	
	R1	R2	R1	R2	R1	R2	R1	R2
TNNT2	3200	3147	11670	11076	99.4	99.4	97.1	95.3
HCN4	2691	5581	8564	8161	80.5	77.9	19.9	26.5
ISL1	1194	3445	10573	9830	72.3	76.5	7.4	8.4
SHOX2	2208	4099	12249	12505	63.3	64.7	3.3	4.0
NKX2-5	6311	5581	15388	16456	34.0	33.8	84.9	85.9


A
nFeature_RNA
nCount_RNA
percent.mt
Identity
B
Atrial Cell Clusters: 0, 6, 8, 12, 13, 19, 21, 24
Pacemaker Cell Clusters: 1, 3, 4
Ventricular Cell Clusters: 2, 5, 7, 15, 23, 25
Number of Cells
Percentage [%]
Cluster
C
MYL2
HEY2
IRX4
MYL7
NPPA
NR2F2
UMAP_1
UMAP_2
D
TNNT2
ACTN2
MYH6
Expression Level
Cluster
E
WT1
ALDH1A2
BNC1
PDPN
Expression Level
Cluster
F
FOXC1
MSX1
NRG1
NPR3
NFATC1
Expression Level
Cluster
G
FOXA2
EOMES
MESP1
MESP2
Expression Level
Cluster
H
PHOX2B
PRPH
CHRNA3
Expression Level
Cluster
I
Van Eif 2019 Development Paper (1076 Genes)
SANCM vs. ACM Up DEGs (271 Genes)
2020 Circ Res Paper SANCM vs. ACM Up DEGs (973 Genes)
988
10
225
11
67
25
870

A

scRNA-seq clustering from Fig. 3c

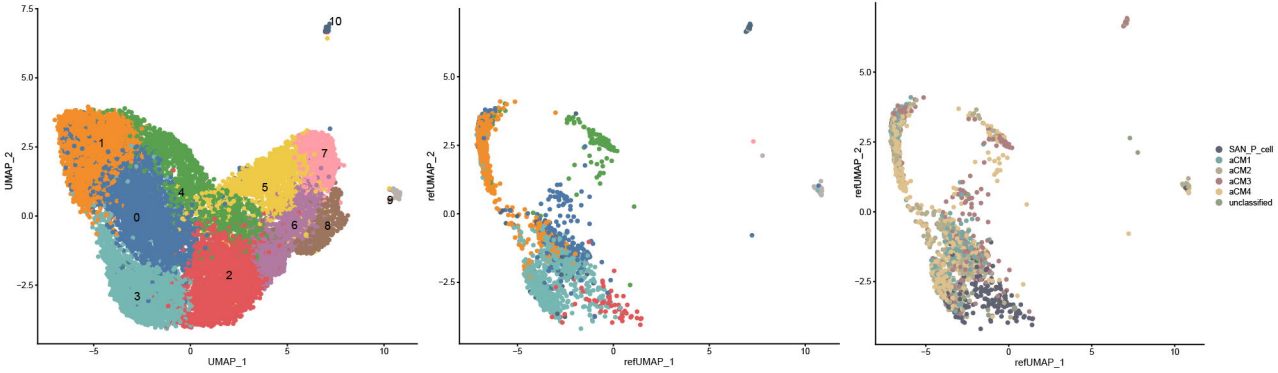
Mapping of Wiesinger et al. *in vitro* scRNA-seq data onto UMAP from Fig. 3c

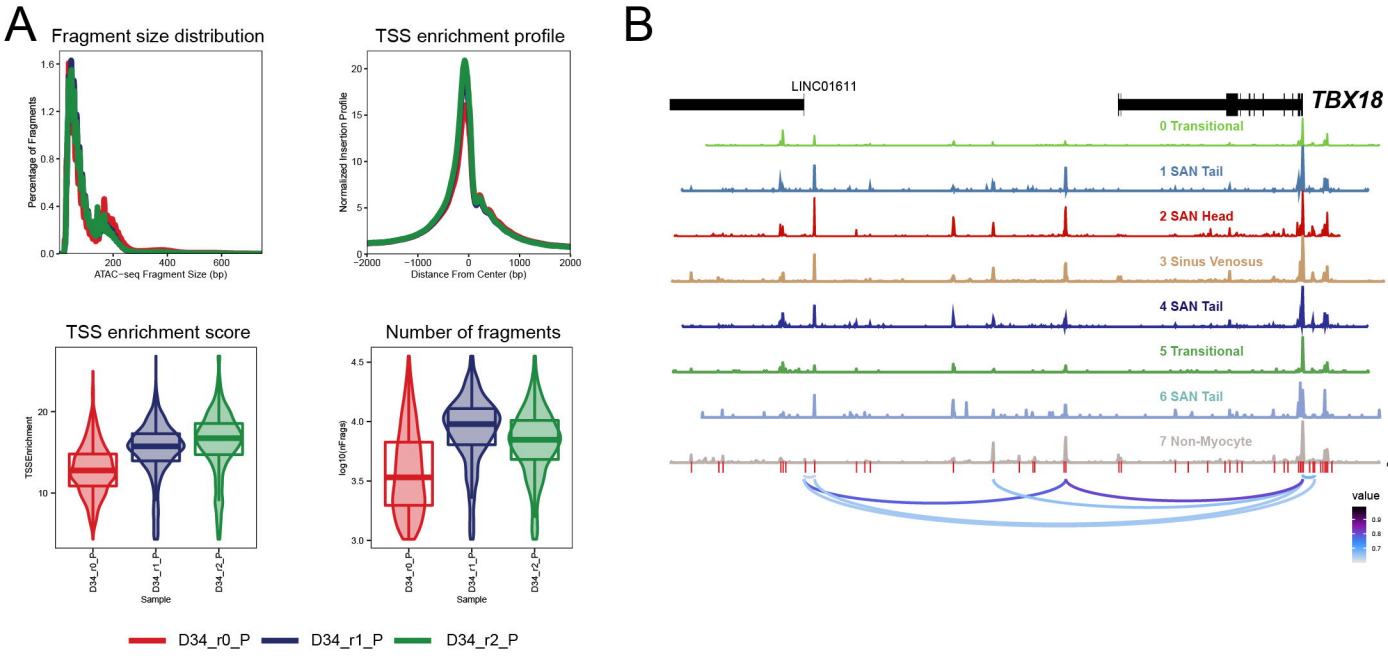


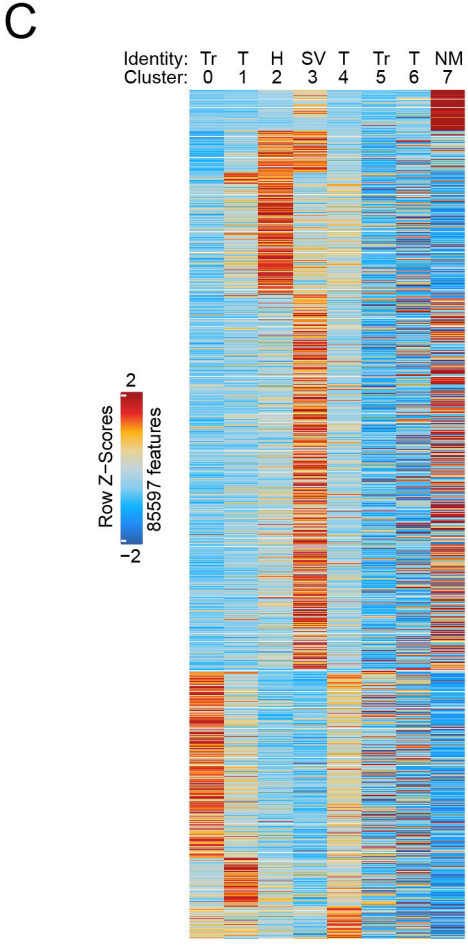
B

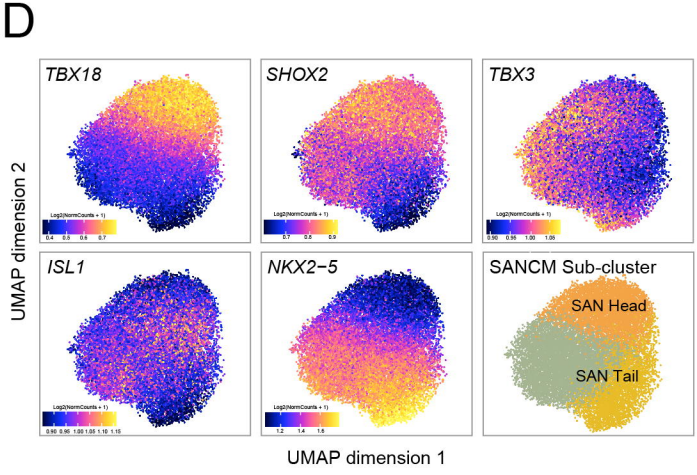
scRNA-seq clustering from Fig. 3c

Mapping of Kanemaru et al. *in vivo* scRNA-seq data onto UMAP from Fig. 3c

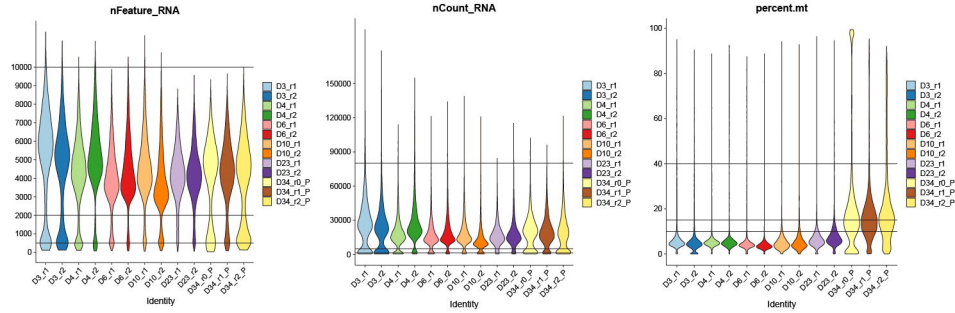



A
Fragment size distribution
Percentage of Fragments
ATAC−seq Fragment Size (bp)
TSS enrichment profile
Normalized Insertion Profile
Distance From Center (bp)
TSS enrichment score
TSSEnrichment
Number of fragments
log10(nFrags)
Sample
D34_r0_P
D34_r1_P
D34_r2_P
B
LINC01611
TBX18
0 Transitional
1 SAN Tail
2 SAN Head
3 Sinus Venosus
4 SAN Tail
5 Transitional
6 SAN Tail
7 Non-Myocyte
value

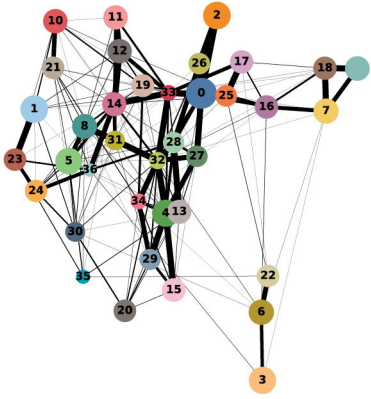

C
Identity: Tr T H SV T Tr T NM
Cluster: 0 1 2 3 4 5 6 7
Row Z−Scores
85597 features
2
−2


D
TBX18
SHOX2
TBX3
ISL1
NKX2−5
SANCM Sub-cluster
SAN Head
SAN Tail
UMAP dimension 2
UMAP dimension 1

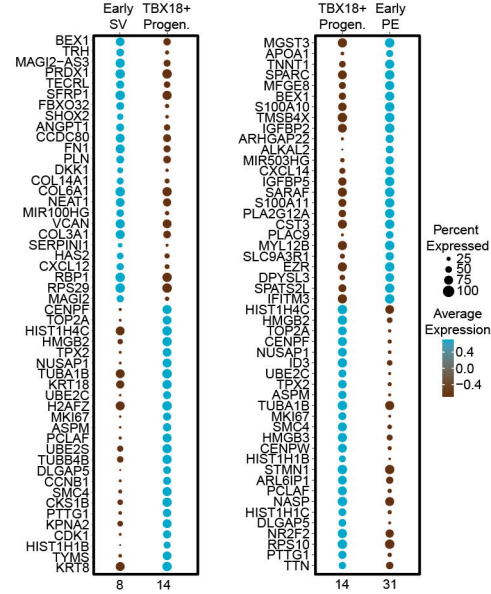
A

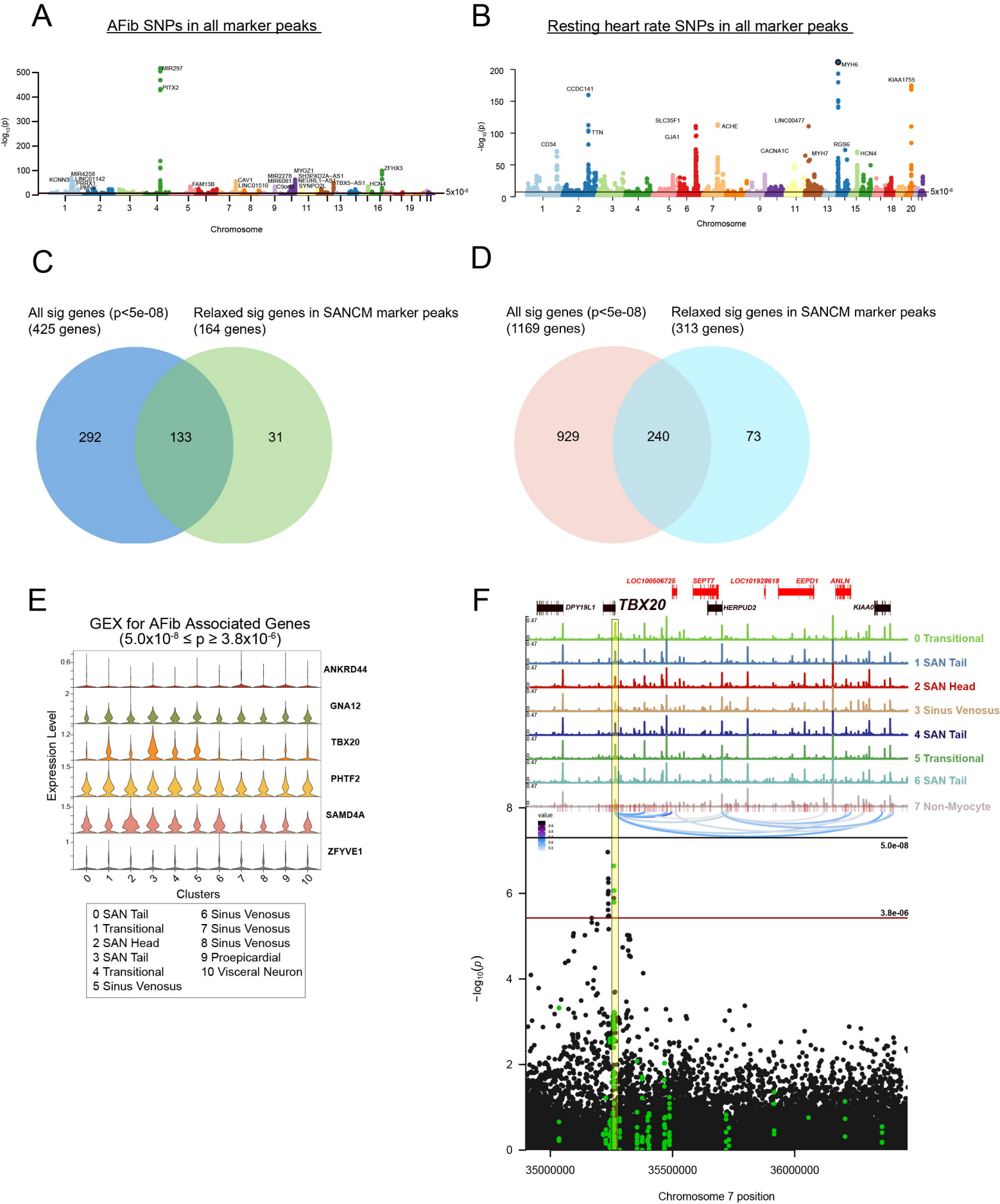

nFeature_RNA
nCount_RNA
percent.mt
Identity
D3_r1
D3_r2
D4_r1
D4_r2
D6_r1
D6_r2
D10_r1
D10_r2
D23_r1
D23_r2
D34_r0_P
D34_r1_P
D34_r2_P

B



C


Early SV
TBX18+ Progen.
TBX18+ Progen.
Early PE
BEX1
TRH
MAGI2-AS3
PRDX1
TECR1
SFRP1
FBXO32
SHOX2
ANGPT1
CCDC80
FN1
PLN
DKK1
COL14A1
COL6A1
NEAT1
MIR100HG
VCAN
COL3A1
SERPINI1
HAS2
CXCL12
RBP1
RPS29
MAGI2
CENPF
TOP2A
HIST1H4C
HMGB2
TPX2
NUSAP1
TUBA1B
KRT18
UBE2C
H2AFZ
MKI67
ASPM
PCLAF
UBE2S
TUBB4B
DLGAP5
CCNB1
SMC4
CKS1B
PTTG1
KPNA2
CDK1
HIST1H1B
TYMS
KRT8
8
14
MGST3
APOA1
TNNT1
SPARC
MFGE8
BEX1
S100A10
TMSB4X
IGFBP9
ARHGAP22
ALKAL2
MIR503HG
CXCL14
IGFBP5
SARAF
S100A11
PLA2G12A
CST3
PLAC9
MYL12B
SLC9A3R1
EZR
DPYSL3
SPATS2L
IFITM3
HIST1H4C
HMGB2
TOP2A
CENPF
NUSAP1
D3
UBE2C
TPX2
ASPM
TUBA1B
MKI67
SMC4
HMGB3
CENPW
HIST1H1B
STMN1
ARL6IP1
PCLAF
NASP
HIST1H1C
DLGAP5
NR2F2
RPS10
PTTG1
TTN
14
31
Percent Expressed
25
50
75
100
Average Expression
0.4
0.0
−0.4


A
AFib SNPs in all marker peaks
-log10(p)
Chromosome
5x10-8
MIR297
PITX2
KCNN3
MIR4258
LINC01142
PRRX1
FAM13B
CAV1
LINC01510
MIR2278
MIR6081
C9orf3
MYOZ1
SH3PXD2A-AS1
NEURL1-AS1
SYNPO2L
TBX5-AS1
HCN4
ZFHX3
B
Resting heart rate SNPs in all marker peaks
-log10(p)
Chromosome
5x10-8
MYH6
KIAA1755
CCDC141
TTN
CD34
SLC35F1
GJA1
ACHE
LINC00477
CACNA1C
MYH7
RGS6
HCN4
C
All sig genes (p<5e-08)
(425 genes)
Relaxed sig genes in SANCM marker peaks
(164 genes)
292
133
31
D
All sig genes (p<5e-08)
(1169 genes)
Relaxed sig genes in SANCM marker peaks
(313 genes)
929
240
73
E
GEX for AFib Associated Genes
(5.0x10-8 ≤ p ≥ 3.8x10-6)
Expression Level
ANKRD44
GNA12
TBX20
PHTF2
SAMD4A
ZFYVE1
Clusters
0 SAN Tail
1 Transitional
2 SAN Head
3 SAN Tail
4 Transitional
5 Sinus Venosus
6 Sinus Venosus
7 Sinus Venosus
8 Sinus Venosus
9 Proepicardial
10 Visceral Neuron
F
LOC100506725
SEPT7
LOC101928618
EEPD1
ANLN
DPY19L1
TBX20
HERPUD2
KIAA0
0 Transitional
1 SAN Tail
2 SAN Head
3 Sinus Venosus
4 SAN Tail
5 Transitional
6 SAN Tail
7 Non-Myocyte
value
5.0e-08
3.8e-06
−log10(p)
35000000
35500000
36000000
Chromosome 7 position