A dual SHOX2:GFP; MYH6:mCherry knockin hESC reporter line for derivation of human SAN-like cells

Zaniar Ghazizadeh, Jiajun Zhu, Faranak Fattahi, ..., Geoffrey S. Pitt, Todd Evans, Shuibing Chen

xaniar@gmail.com (Z.G.)
tre2003@med.cornell.edu (T.E.)
shc2034@med.cornell.edu (S.C.)

Highlights
We developed a dual SHOX2:GFP; MYH6:mCherry knockin hESC reporter line

An efficient strategy was developed to generate and purify hESC-SAN cells

hESC-SAN cells display molecular and electrophysiological identities of nodal cells

Single-cell RNA sequencing analysis was applied to characterize the hESC-SAN cells
A dual SHOX2:GFP; MYH6:mCherry knockin hESC reporter line for derivation of human SAN-like cells

Zaniar Ghazizadeh,1,11,* Jiajun Zhu,1 Faranak Fattahi,2,3,4 Alice Tang,1 Xiaolu Sun,5 Sadaf Amin,1,12 Su-Yi Tsai,6 Mona Khalaj,1 Ting Zhou,2 Ryan M. Samuel,3,4 Tuo Zhang,7 Francis A. Ortega,8,9 Miriam Gordillo,1 Dorota Moroziewicz,10 The NYSCF Global Stem Cell Array® Team,10 Daniel Paull,10 Scott A. Noggle,10 Jenny Zhaoying Xiang,7 Lorenz Studer,2 David J. Christini,9 Geoffrey S. Pitt,5 Todd Evans,1,* and Shuibing Chen1,13,*

SUMMARY
The sinoatrial node (SAN) is the primary pacemaker of the heart. The human SAN is poorly understood due to limited primary tissue access and limitations in robust in vitro derivation methods. We developed a dual SHOX2:GFP; MYH6:mCherry knockin human embryonic stem cell (hESC) reporter line, which allows the identification and purification of SAN-like cells. Using this line, we performed several rounds of chemical screens and developed an efficient strategy to generate and purify hESC-derived SAN-like cells (hESC-SAN). The derived hESC-SAN cells display molecular and electrophysiological characteristics of bona fide nodal cells, which allowed exploration of their transcriptional profile at single-cell level. In sum, our dual reporter system facilitated an effective strategy for deriving human SAN-like cells, which can potentially be used for future disease modeling and drug discovery.

INTRODUCTION
Despite remarkable progress in our understanding of the transcriptional regulation of cardiogenesis and the blueprint that specifies the four-chambered mammalian heart, as well as many of the pathologic processes that lead to cardiac dysfunction, relatively less is known about the specialized network of cells that regulate cardiac rhythmicity, known as the cardiac conduction system (CCS). At the top of the CCS hierarchy is the sinoatrial node (SAN) comprised of SAN cells that generate the electrical impulses triggering each heartbeat. The SAN cells are unique based on their capacity to drive spontaneous electrical activity regulated by autonomic responsiveness. Loss of SAN cells and SAN fibrosis are the major causes of sinus node dysfunction (SND), and there are no cellular or pharmacological therapies to treat such disease. Thus, there is a need to develop strategies to treat patients with CCS defects. Artificial pacemakers and defibrillators can be effective in many cases, and over 200,000 devices are implanted each year in the US (Mond and Proclemer, 2011). However, these treatments can have low efficacy and also cause additional problems, such as pro-arrhythmia side effects, infection, and device malfunction (Holzmeister and Leclercq, 2011; Woods and Olgin, 2014). Devices placed into pediatric patients must be replaced as heart growth proceeds throughout life. Thus, there is a clear and compelling need to develop strategies to derive SAN cells for modeling CCS diseases and potentially for cell replacement therapy.

Human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs) provide, in theory, unlimited resources to derive functional human cells. Current progress in the directed differentiation to subtype-specific cardiomyocytes including SAN-like cells has been recently comprehensively reviewed (Zhao et al., 2020b). Some protocols have reported deriving SAN-like cells from hESC/iPSCs through stepwise directed differentiation (Prozeti et al., 2017b), inhibiting nodal signaling (Yechikov et al., 2020), modulating Wnt signaling (Liang et al., 2020; Ren et al., 2019), or through forced expression of TBX3 (Zhao et al., 2020a) or TBX18 (Gorabi et al., 2019), although most of these reports do not use selection, or use an NKO2.5 negative selection approach to enrich SAN-like cells. Owing to lack of a purification method, it is challenging to determine the efficiency, cellular identity, and function of the derived cells (Kapoor et al., 2020).
RESULTS

Generation of human SAN-like cells from hESCs

We sought to develop a protocol that maximizes the generation of pluripotent stem cell-derived SAN pacemaker lineages (PSC-SAN). Based on previous reports (Hoogaars et al., 2007; Ionta et al., 2015; McNally and Svensson, 2009), SAN progenitors are expected to be enriched for expression of key pacemaker transcription factors TBX3, TBX18, and SHOX2 (TTS). Previously, we engineered murine embryonic stem cell (ESC) lines for conditional expression of GATA4, GATA5, or GATA6, to efficiently generate diverse cardiac cell populations including, nodal, atrial, and ventricular cells (Turbendian et al., 2013). Among these cell lines, GATA5 expression resulted in the highest transcript levels of TTS (Figure S1A). We therefore performed an unbiased qRT-PCR-based chemical screen on GATA5-induced progenitors using annotated pathway modulating compounds to identify modulators that further enhance the expression of TTS and specification toward pacemaker lineage. Two compounds caused a minimum of 2-fold increase in transcript levels for at least two of the transcription factors (Figure S1B). Both were EGF pathway inhibitors (Figure S1C), suggesting an important role for the EGF pathway in nodal identity, consistent with previous mouse studies (Zhu et al., 2010). Given the association of the SAN with the right atrium, we also focused on the chemicals that increased the expression of atrial rather than ventricular cardiac lineage structural proteins. The screen identified curcubitacin, a STAT3 inhibitor, that enhances the expression levels of nodal and atrial structural marker Myh6 (Figure S1D).

These results were next translated into a hESC differentiation protocol, including two additional pathway modulators. Retinoic acid (RA) was included to shift regional identity of cardiac progenitors toward a more caudal phenotype, since pacemaker cells are thought to arise from a more caudal population of cardiac progenitors (Devalla et al., 2015). In addition, an FGF inhibitor was used to augment the expression of ISL1, a known regulator of the nodal program in early mouse development (Birket et al., 2015; Liang et al., 2015). Collectively, the full combination (condition 8) of 5 μM EGF inhibitor (Tyrphostin AG490), 0.1 μM STAT3 inhibitor (curcubitacin), 5 μM FGFR inhibitor (SU5402), and 1 μM RA significantly increased human TTS expression levels (Figures 1A, 1B, and S1E).

To quantify and purify SAN-like cells using this directed differentiation protocol, a dual knockin SHOX2:GFP; MYH6:mCherry reporter line was created using CRISPR/Cas9-based gene-targeting techniques (Figure S2A). Immunofluorescence and FACS analyses confirmed the success of the protocol for generating putative hPSC-SAN cells, as condition 8 maximized generation of SHOX2:GFP+ cells (Figures 1C, 1D, S2B, and S2C). The sorted GFP+ cells were enriched approximately 4000-fold for SHOX2 transcript levels compared to GFP− cells (Figure 1E), and most of the GFP+ cells co-express ISL1 and HCN4 (>95% of cells), proteins closely associated with the development and function of SAN cells (Figure S2D). To further enhance production of GFP+ cells, we considered that SAN development is also controlled epigenetically (Vedantham et al., 2013) and performed an unbiased screen using a library of known epigenetic modulatory compounds. Cells were differentiated using condition 8 with compounds added at day 7 (Figure 1F). After 14 days of differentiation, treatment with the HDAC inhibitor chidamide led to a significant increase in the percentage of GFP+ cells (≥50%). Using this directed differentiation protocol (Figure 1G), the GFP+ putative SAN-like cells could be maintained at this level for 30–40 days post differentiation, and without reducing the total cell number, demonstrated by both flow cytometry and immunostaining at day 30 (Figures 1H and S2E). We also differentiated the SHOX2:GFP; MYH6-mCherry reporter line using a standard ventricular cardiomyocyte differentiation protocol (Tsai et al., 2020), and very few SHOX2-GFP+ cells were detected (Figure S2F). Additionally, we performed staining for MYL2, a ventricular cardiomyocyte marker (Luo et al., 2021), and NR2F2, a marker expressed in both atrium and SAN (Burkhard and Bakkers, 2018; Hoffmann et al., 2021). 42.2 ± 7.8% of MYH6:mCherry+ cells are NR2F2+SHOX2-GFP+, suggesting that the protocol generates this level of atrial-like cells in the unpurified population. No MYL2 expression was detected using the current differentiation protocol (Figures S2G and S2H), suggesting that ventricular cardiomyocytes are not present. This result is expected since the protocol was designed to bias away from ventricular cardiomyocyte fate. Given the potential application of pluripotent stem cell-derived nodal cells in disease modeling, we set out to validate the platform in iPSCs. We differentiated two additional (non-reporter) iPSC lines using the same protocol. Owing to the lack of a
Figure 1. Generation of human SAN-like cells from hESCs

(A) Schematic representation of the differentiation protocols 1–8. The protocol involves three stages (S1, S2, and S3) during the first 9 days of directed differentiation.

(B–D) Quantitative PCR (B), live fluorescence images (C), and FACS plots (D) of the day 30 cells derived from the SHOX2:GFP;MYH6:mCherry H9 line using the corresponding differentiation protocols (n = at least 3 biological replicates for each condition). For quantitative PCR data, fold changes were normalized to protocol #1. p values calculated using ordinary one-way ANOVA with Fisher’s test (**p < 0.01, ***p < 0.001). Data are represented as mean ± SEM. The x axis in panel D is side scatter.

(E) Quantitative PCR analysis for SHOX2 transcriptional expression of GFP+ cells purified after sorting. n = 3 biological replicates. p value calculated by unpaired two-tailed Student’s t-test ***p < 0.001. Data are represented as mean ± SEM, and normalized to GFP-negative population.

(F) Schematic and results of the epigenetic library chemical screen.
commercially available reliable SHOX2 antibody, the differentiated cell population was stained for expression of ISL1 (Figure S2I), a marker that overlaps significantly with SHOX2:GFP in hESC-SAN cells (Figure S2C). The percentage of ISL1+ cells was comparable with the percentage of ISL1+ cells generated using the dual reporter hESC line, suggesting that this differentiation protocol can be applied for multiple hESC/iPSC lines.

**Transcriptome profiling of hESC-SAN cells**

To further validate the nodal identity of the SHOX2:GFP+ cells, the GFP+ versus mCherry+ (GFP/C0) populations were sorted at day 25 or day 40 of differentiation and compared by RNA-seq profiling. Owing to the lack of RNA-seq data profiling purified human SAN tissue, hESC-derived SAN cells were compared with profiles generated from murine SAN and right atrial (representing mostly atrial myocardium) tissues (GSE65658) (Vedantham et al., 2015). The data were analyzed by hierarchical clustering alongside RNA-seq databases generated using HCN4:GFP+ cells from dissected murine SAN tissue and GFP− cells from adjacent right atrial tissue. Whether the clustering analysis used whole transcriptomic data (Figure 2A) or selected datasets for genes encoding ion channels, structural proteins, metabolic proteins, or growth factors (Figure 2B), the GFP+ and mCherry+ cell profiles clustered with SAN and atrial cardiomyocyte profiles, respectively. The phenotypes were further confirmed by gene set enrichment analyses: the geneset representing genes with levels enhanced 10-fold in SAN vs atrial cells were highly enriched in GFP+ cells while those that have levels enhanced 10-fold in atrial vs SAN cells were highly enriched in mCherry+ cells (Figure 2C).

**Functional characterization of hESC-SAN cells**

We maintained differentiated cells for up to 60 days and evaluated gene expression patterns. At day 30, SHOX2:GFP+ cells stained positive by immuno-fluorescence for HCN4, ISL1, and TBX5 (Figure 3A). By day 60, there was increased expression levels for channels related to pacemaker activity, including Ca3.1, Ca1.3, and Cx30.2 (Figure 3A). Expression of these nodal markers is not increased in mCherry+; GFP− cardiomyocytes (Figure 3A). Transcript levels of pacemaker markers TBX3, TBX18, ISL1, and HCN4 were significantly higher in SHOX2:GFP+; MYH6:mCherry+ cells compared to SHOX2:GFP−; MYH6:mCherry+ cells. On the other hand, atrial cardiomyocyte marker Nppa was lower in SHOX2:GFP+; MYH6:mCherry+ cells compared to SHOX2:GFP−; MYH6:mCherry+ cells (Figure 3B). There was no significant difference in NKK2.5 expression levels comparing these two populations until day 45 (Figure 3B). This is in agreement with previous in vitro and in vivo studies showing that NKK2.5 and SHOX2 are co-expressed during the early stages of mouse development (Goodyer et al., 2019) and differentiating embryoid bodies (Ye et al., 2015). At day 45, the expression level of NKK2.5 in SHOX2:GFP+ cells is significantly lower than in MYH6:mCherry+ cells (Figure 3B), which is consistent with suppression of NKK2.5 by SHOX2 (Espinoza-Lewis et al., 2009). These results highlight the importance of development of strategies for direct isolation and characterization of SAN cells using nodal-specific markers.

A major function of the SAN is to regulate heart rhythm under the control of autonomic neurons (AN). ESC-derived cardiomyocytes show poor connectivity with ESC-derived AN (Oh et al., 2016). We probed the functionality of in-vitro-derived hPSC-SAN cells by assessing connectivity with AN (Figure 3C). For connectivity studies, an optogenetic reporter line was used to facilitate light-induced control of neural activity (Fattahi et al., 2016). AN were derived as previously described (Tchieu et al., 2017; Zeltner et al., 2016) from a hESC line expressing enhanced yellow fluorescent protein (eYFP)-tagged channelrhodopsin-2 (ChR2) under control of the human synapsin promoter. Co-culture of day 25 hPSC-SAN cells with these AN resulted in physical connection between SAN cells and AN, as shown by increased expression of synapsin, and junctional morphology indicated by electron microscopy (Figure 3D). Furthermore, the co-culture platform resulted in increased maturation of AN as shown by increased adrenaline levels in supernatant compared to AN co-cultured with hESC-derived...
ventricular-like cardiomyocytes (hESC-V cells) (Figures 3E and S3B). Light-activated contractions in SAN cells further validated the connectivity (Video S1).

One key functional property of a nodal cell is its characteristic action potential with fast spontaneous firing rates and slow maximum upstroke velocities (<30 V/s). SHOX2:GFP+ cells generated from the differentiation protocol at day 30 were sorted and patch-clamping experiments showed 8 out of 10 demonstrated the characteristic electrophysiologic phenotype of nodal cells, distinct from hPSC-V cells (Figure 3F). In addition to the spontaneous firing of hESC-SAN cells, we used rate-modifying drugs and explored their effect on these cells. Ivabradine, a specific HCN4 channel modulator (Postea and Biel, 2011), has been shown to slow SAN activity. A Ca2+-flux assay was used to record the baseline calcium activity of hESC-SAN or hESC-V cells, which were then treated with 1 μM ivabradine and recordings taken after 5 min. While no significant difference was observed in
**Figure 3. Functional characterization of hESC-SAN cells**

(A) Immunofluorescent co-staining of GFP with pacemaker-associated transcription factors (ISL1, TBX5, and NKX2.5) and ion channels (HCN4, Cav3.1, Cav1.3, and Cx30.2) in cells differentiated to day 30 or day 60. Scale bar = 200 µm.

(B) qRT-PCR analysis using samples from GFP+ (G+) or mCherry+ (Ch+) cells at day 25 or day 45. n = 3 biological replicates. Fold change was normalized to undifferentiated hESCs. p values calculated by one-way ANOVA with Fisher’s test (*p < 0.05, **p < 0.01, ***p < 0.001). Data are represented as mean ± SEM.

(C) Schematic for co-culture of hESC-derived ANs expressing eYFP-channelrhodopsin-2 with either hESC-SAN or hESC-V.

(D) Representative immunofluorescence (left panel) and electron microscopy images (right panel) of hESC-AN and hESC-SAN after 12 weeks of co-culture. Scale bar: 100 µm (left panel) and 1 µm (right panel).

(E) Basal release of catecholamine was analyzed using an ELISA kit for AN alone, AN + hESC-SAN and AN + hESC-V conditions after 8 weeks of co-culture. n = 3 biological replicates. p values calculated by one-way ANOVA (**p < 0.01). Data are represented as mean ± SEM.

(F) Representative image of averaged action potentials recorded by perforated patch clamp from individual hESC-SAN (left) and hESC-V (right) cells and distribution of dV/dtmax, APD30, and APD90 recorded used day 30 cells. p values calculated by one-way ANOVA used to compare means (*p < 0.05). Data are represented as mean ± SEM.

(G) Beating rate of hESC-SAN and hESC-V cells. p values calculated by unpaired two-tailed Student’s t-test ****p < 0.0001. Data are represented as mean ± SEM.

(H) Spontaneous firing of hESC-SAN cells. Note that for these experiments, cells were replated as single cells on cover slips and they were maintained in physiological buffers (rather than their optimal cell culture media). Cells are manipulated for up to an hour at room temperature, during which time the beating rate decreases substantially (compared to conditions in G).

(I) Isc current measurements in hESC-SAN or hESC-V cells by patch clamp. Left: Example traces of hESC-SAN cells; Middle: Averaged I-V curves; Right: Example traces of hESC-V cells. (n = 5 biological replicates).

(J) Quantitative analysis of day 30 hESC-SAN function in response to 1 µM ivabradine. Fold change in calcium transient amplitude (ΔF/F0) of hESC-SAN or hESC-V cells before and after treatment with ivabradine (n = 22 biological replicates for hESC-SAN and n = 17 biological replicates for hESC-V). Unpaired two-tailed Student’s t-test was used to compare means (*p < 0.05). Each dot represents data measured from independent cells.

(K) Quantitative analysis of day 30 hESC-SAN function in response to 1 µM isoproterenol and 5 µM carbachol. p values calculated by unpaired two-tailed Student’s t-test **p < 0.01, ***p < 0.001. Data are represented as mean ± SEM.

See also Figure S3A and Video S1.

hESC-V cell cultures, a significant reduction was observed in Ca2+-flux activity in the hESC-SAN cells (Figure 3J). A Ca2+-flux assay was also used to monitor the response of hESC-SAN cells to 1 µM isoproterenol or 5 µM carbachol treatment (Figure 3K). Carbachol treatment had negative chronotropic and isoproterenol had positive inotropic effects on hESC-SAN cells. Therefore, based on multiple criteria, including gene expression profiles, AN connectivity, electrophysiological characteristics, and drug sensitivity, the hESC-SAN cells recapitulate expected features of human nodal pacemaker cells.

**Single-cell RNA sequencing analysis of differentiated hESC-SAN and hESC-V cells**

The hPSC-SAN cells were analyzed in a single-cell RNAseq experiment to explore heterogeneity of the cell population and identify lineage-restricted transcriptome features. For this purpose, SHOX2-GFP+ cells were sorted from the pacemaker differentiation culture, and MYH6:mCherry+ cells were sorted following cardiomyocyte directed differentiation to generate a ventricular-like population (hESC-V). After pooled sequencing, tSNE plots (Figure 4A) clustered cells with similar gene expression patterns. Cluster analysis using global gene expression values classified the cells into three groups including high-SHOX2, no-SHOX2, and a smaller cluster of mixed-SHOX2 cells. The mixed-SHOX2 population contains both GFP+ and mCherry+ cells (Figure S3C). Because cardiac markers are not highly expressed in the mixed-SHOX2 population, these cells might represent non-cardiac SHOX2+ cells (Figure 4B). We therefore focused on high-SHOX2 and no-SHOX2 cells for further analysis. Principal component analysis (PCA, Figure 4C) confirmed that high-SHOX2 and no-SHOX2 cells are two distinct populations (Figure 4C).

High-SHOX2 cells significantly transcribed structural transcripts encoding structural proteins TPM1, MYL7, MYH6, MYL6, ACTA2, DSTN, ACTC1, NEXN, NEBL, MYH11, and MYLK (Figure 4D). The role in SAN physiology for some of these proteins including MYLK, MYH11, and ACTA2 is unknown, although they were also reported to be present in the murine SA node (Vedantham et al., 2015). Notable structural protein genes expressed in the no-SHOX2 cell population include TTN, TPM1, MYL7, MYH6, and ACTC1 (Figure 4D). The high-SHOX2 cell population was relatively homogenous in expression of transcripts for the key ion channel HCN4 and transcriptional regulator ISL1, while no-SHOX2 cells showed heterogeneity in the expression levels of these two genes (Figure 4E). PITX2c, a gene responsible for suppression of the SAN program in atrial cells (Wang et al., 2010), was highly expressed in the no-SHOX2 cells while expressed at very low relative levels in the high-SHOX2 cells. The presence of IRX4 transcripts in the no-SHOX2 cells (and absence in the high-SHOX2 cells) confirms the ventricular identity of the former (Figure 4F). The high-SHOX2 cell population had lower but detectable NKX2.5 expression levels, and was also associated with relatively higher...
levels of HEY1 and ID2 transcripts, consistent with an SA nodal identity rather than an AV node (Hoogaars et al., 2007; Horsthuis et al., 2009; McNally and Svensson, 2009). Consistent with previous studies, CACNA1D (encoding CaV1.3), which is necessary for the pace-making activity of nodal cells (Torrente et al., 2016), was relatively enriched in high-SHOX2 cells. Among contactin (CNTN) isoforms, CNTN4 transcripts were relatively more abundant in the high-SHOX2 cell population, compared to variable expression levels of CNTN1.

Consistent with the heat maps, jitter plots show that the high-SHOX2 population showed relatively homogeneous expression of HCN4, ISL1, TBX5, and TBX3 genes compared to TBX18 (Figure 4G). Previous murine developmental studies suggested that the TBX18 expression pattern might determine the identity of different parts of the SA node, including head or tail (Wiese et al., 2009). The single-cell sequencing data indicate that the in vitro generated hPSC-SAN cells recapitulate a spectrum of TBX18 expression patterns, and therefore may represent multiple segments of the SA node. Another key calcium regulator identified in high-SHOX2 cells was VSNL1 (Figure 4H). Given the importance of SAN action potential in regulation of rate and rhythm, we further focused on whether this gene is expressed in adult tissue. Single-cell analysis from adult tissue biopsies proved that VSNL1 is co-expressed with HCN4 and CACNA1D in the adult tissue (Figures S3C and S3D). Previously, a number of transcription factors or calcium channels, such as SMOC2, BMP2, GFRA2, and NOTCH3, were proposed to be enriched in the murine SA node (van Eif et al., 2019). However, none of these genes were co-expressed with HCN4 in adult heart tissue single-cell analysis, suggesting the need for further studies to elucidate their role in human nodal physiology (Figure S3D).

DISCUSSION

This study describes an efficient strategy to derive and purify SAN-like cells from hPSCs. The ability to generate human SAN-like cells in vitro can facilitate study of genetic and iatrogenic arrhythmias and may eventually allow scaled production and transplantation as biological pacemakers. Previous efforts focused on in vivo direct conversion or an NKX2.5 negative selection approach to obtain nodal-like cells. However, due to lack of a purification method, the studies were limited in characterization and validation of nodal identity for the cells at cellular and molecular levels (Kapoor et al., 2013; Protze et al., 2017a). Our reporter strategy enabled direct visualization and standardized in vitro assays for developmental studies, purification, and drug discovery (Birket et al., 2015; Cingolani et al., 2017; Kapoor et al., 2013). We demonstrate functionality of the in-vitro-derived cells through several methods including global gene expression analysis, cell-type-specific expression profiling, and electrophysiological mapping. While the ESC-derived SAN-like cell profiles clustered with murine SAN tissue profiles rather than atrial myocardium profiles, it will be important in future experiments to compare directly to human tissues. Notably these cells were able to make functional connectivity with autonomic neurons, enabling studying synaptic activity and neurotransmitter release. Furthermore, the ability to directly purify these cells led to single-cell analysis of an enriched population of SAN cells. This showed a spectral expression of TBX18 and TBX3, similar to the developing nodal cells in addition to identification of previously unknown genes involved in nodal cell physiology. Therefore, the precise isolation and expansion of these cells can pave the way for better understanding of cell and non-cell autonomous processes involved in nodal tissue biology in regulating rate and rhythm in conjunction with the autonomic nervous system. The human SAN-like cells derived here can be potentially used for cardiac nodal disease modeling and drug screens.

Limitations of the study

Although scRNA-seq analysis of hESC-derived SAN-like cells identified some previously unknown markers for SAN, additional work is needed to validate the expression and determine the biological function of the...
identified markers. While the cell profiles did not identify specific SAN-like cell subsets, the broad range of expression for some key transcription factors such as TBX18 might indicate heterogeneity, and it will be of interest to determine if this relates to endogenous patterning of the natural SAN. The dual reporter line will provide a powerful tool to study human SAN cells. However, additional purification and maturation strategies will be needed to enrich the SAN-like cells from non-reporter lines and explore potential use in regenerative therapy.

STAR METHODS
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
- METHOD DETAILS
  - Sinoatrial and ventricular myocyte induction
  - FACS and immuno-fluorescence
  - Gene expression analysis
  - Generation of the SHOX2:GFP; MYH6:mCherry dual hESC reporter line
  - Bulk RNA sequencing
  - Single cell RNA sequencing
  - Adult heart single cell analysis
  - Calcium imaging
  - Cellular electrophysiology and characterization
  - Chemical screen
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION
Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104153.

ACKNOWLEDGMENTS
This work was supported by the American Heart Association (18CSA34080171, S.C., T.E.), Department of Surgery, Weill Cornell Medicine (T.E., S.C.), Tri-Institutional Stem Cell Institute (2021-026, S.C.), NIDDK (R01DK130454, R01DK119667-02S1, S.C.), NHLBI (R01HL151190, R01HL160089, G.S.P.). S.C is supported as Irma Hirschl Trust Research Award Scholars. T.E. is supported by an Outstanding Investigator Award (R35 HL135778). We thank Dr. Friederike Dundar for her help on single-cell RNA-seq analysis.

AUTHOR CONTRIBUTIONS
S.C., T.E., and L.S. conceived and designed the experiments. Z.G., J.Z., F.F., A.T., S.A., S.T., M.K., T.Z., and M.G. performed experiments. X.S., F.A.O., D.C., and G.S.P. performed electrophysiology experiments. R.M.S., T.Z., and J.Z.X. performed the bioinformatics analyses.

DECLARATION OF INTERESTS
T.E. and S.C. are founding owners of OncoBeat, LLC. A patent has been filled for “Compositions and methods for generation of sinoatrial node-like cells and their use in drug discovery”.

Received: September 20, 2021
Revised: January 25, 2022
Accepted: March 22, 2022
Published: April 15, 2022
REFERENCES

Birke, M.J., Ribeiro, M.C., Verkerk, A.O., Ward, D.L., Leteigunho, A., den Hout, S.C., Orfora, V.V., Devalla, H.D., Schwach, V., Bellin, M., et al. (2015). Expansion and patterning of cardiovascular progenitors derived from human pluripotent stem cells. Nat. Biotechnol. 33, 970–979.

Burkhard, S.B., and Bakkers, J. (2018). Spatially resolved RNA-seq-sequencing of the embryonic heart identifies a role for Wnt/β-catenin signaling in autonomc control of heart rate. Elife 7, e31515.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. https://doi.org/10.1038/nbt.4096.

Cingolani, E., Goldhaber, J.I., and Marban, E. (2017). Next-generation pacemakers: from small Cingolani, E., Goldhaber, J.I., and Marban, E. (2009). Shox2 is essential for the resolution. Circ. Res.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

Devalia, H.D., Schwach, V., Ford, J.W., Milnes, J.T., El-Hauou, S., Jackson, C., Gkatzis, K., Elliott, D.A., Chirikian, O., Choi, S., Goodyer, W.R., Beyersdorf, B.M., Paik, D.T., Tian, L., Li, G., Buikema, J.W., Zeltner, N., Mica, Y., Zimmer, B., Kishinevsky, S., Zeltner, N., Chen, S., and Studer, L. (2011). The 11th World Society of Arrhythmia’s project. Pacing Clin. Electrophysiol.: Real-Time eXperiment Interface (RTXI). PLoS Comput. Biol. 13, e1005430.

Postea, O., and Bie, M. (2011). Exploring HCN channels as novel drug targets. Nat. Rev. Drug Discov. 10, 903–914.

Protei, S., Li, J., Nussinovitch, U., Ohana, L., Backx, P.H., Gepstein, L., and Keller, G.M. (2017a). Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker. Nat. Biotechnol. 35, 56–68.

Protei, S., Li, J., Nussinovitch, U., Ohana, L., Backx, P.H., Gepstein, L., and Keller, G.M. (2017b). Sinoatrial node cardiomyocytes derived from human pluripotent cells function as a biological pacemaker. Nat. Biotechnol. 35, 56–68.

Ren, J., Han, P., Ma, X., Farah, E.N., Bloomekat, J., Zeng, X., Zhang, R., Swin, M.M., Witty, A.D., Knight, H.G., et al. (2019). Canonical Wnt5b signaling directly opens Nkx2.5+ mesoderm into pacemaker cardiomyocytes. Dev. Cell 50, 729–743.e5.

Tchieu, J., Zimmer, B., Fattahi, F., Amin, S., Zeltner, N., Chen, S., and Studer, L. (2017). A modular platform for differentiation of human PSCs into all major ectodermal lineages. Cell Stem Cell 21, 399–410.e7.

Torrente, A.G., Mesirca, P., Neco, P., Rizetto, R., Dubel, S., Barrere, C., Sinnegger-Brauns, M., Stierssing, J., Richardson, S., Nargeot, J., et al. (2016). L-type Cav1.3 channels regulate synadine receptor-dependent Ca2+ release during sinoatrial node pacemaker activity. Cardiovasc. Res. 109, 451–461.

Tsai, S.Y., Ghazizadeh, Z., Wang, H.J., Amin, S., Ortega, F.A., Badieyan, Z.S., Hsu, Z.T., Gordillo, M., Kumar, R., Christini, D.J., et al. (2020). A human embryonic stem cell reporter line for monitoring chemical-induced cardiotoxicity. Cardiovasc. Res. 116, 658–670.

Tucker, N.R., Chaffin, M., Fleming, S.J., Hall, A.W., Parsons, V.A., Bedi, K.C., Jr., Akkad, A.D., Herndon, C.N., Arduini, A., Papangeli, I., et al. (2020). Transcriptional and cellular diversity of the human heart. Circulation 142, 466–482.

Turbendian, H.K., Gordillo, M., Tsai, S.Y., Lu, J., Kang, G., Liu, T.C., Tang, A., Liu, S., Fishman, G.I., and Evans, T. (2013). GATA factors efficiently direct cardiac fate from embryonic stem cells. Development 140, 1639–1644.

van Eif, W.Y.W., Stefanovic, S., van Duijbroaden, K., Bakker, M., Wakker, V., de Gier-v de Vries, C., Zaffr, S., Verkerk, A.O., Boukens, B.J., and Christoffels, V.M. (2019). Transcriptome analysis of mouse and human sinoatrial node cells reveals a conserved genetic program. Development 146, dev173161.

Vedantham, V., Evangelista, M., Huang, Y., and Srivastava, D. (2013). Spatiotemporal regulation of an Hcn4 enhancer defines a role for Meis2 and
HDACs in cardiac electrical patterning. Dev. Biol. 373, 149–162.

Vedantham, V., Galang, G., Evangelista, M., Deo, R.C., and Srivastava, D. (2015). RNA sequencing of mouse sinoatrial node reveals an upstream regulatory role for Islet-1 in cardiac pacemaker cells. Circ. Res. 116, 797–803.

Wang, J., Klysik, E., Sood, S., Johnson, R.L., Wehrens, X.H., and Martin, J.F. (2010). Pitx2 prevents susceptibility to atrial arrhythmias by inhibiting left-sided pacemaker specification. Proc. Natl. Acad. Sci. U.S.A. 107, 9753–9758.

Wiese, C., Grieskamp, T., Airik, R., Mommersteeg, M.T., Gardiwal, A., de Gier-de Vries, C., Schuster-Gossler, K., Moorman, A.F., Kapert, A., and Christoffels, V.M. (2009). Formation of the sinus node head and differentiation of sinus node myocardium are independently regulated by Tbx18 and Tbx3. Circ. Res. 104, 388–397.

Woods, C.E., and Olgin, J. (2014). Atrial fibrillation therapy now and in the future: drugs, biologicals, and ablation. Circ. Res. 114, 1532–1546.

Ye, W., Wang, J., Song, Y., Yu, D., Sun, C., Liu, C., Chen, F., Zhang, Y., Wang, F., Harvey, R.P., et al. (2015). A common Shox2-Nkx2-5 antagonistic mechanism primes the pacemaker cell fate in the pulmonary vein myocardium and sinoatrial node. Development 142, 2521–2532.

Yechikov, S., Kao, H.K.J., Chang, C.W., Pretto, D., Zhang, X.D., Sun, Y.H., Smithers, R., Sirish, P., Nolta, J.A., Chan, J.W., et al. (2020). NODAL inhibition promotes differentiation of pacemaker-like cardiomyocytes from human induced pluripotent stem cells (hiPSCs) increases their differentiation into cardiac pacemaker-like cells. Biomed. Pharmacother. 130, 110612.

Zhao, M.T., Shao, N.Y., and Garg, V. (2020b). Subtype-specific cardiomyocytes for precision medicine: where are we now? Stem Cells 38, 822–833.

Zhu, W.Z., Xie, M., Moyer, K.W., Gold, J.D., Askari, B., and Laflamme, M.A. (2010). Neuregulin/ErbB signaling regulates cardiac subtype specification in differentiating human embryonic stem cells. Circ. Res. 107, 776–786.
### STAR METHODS

#### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Antibodies**      |        |            |
| ISL1                | DSHB   | #39.4D5 RRID:AB_2314683 |
| GFP                 | Abcam  | # ab13970 RRID:AB_300798 |
| HCN4                | Abcam  | # ab85023 RRID:AB_2120039 |
| TBX5                | Ptglab | # 13178-1-AP RRID:AB_10638435 |
| NKX2.5              | Ptglab | # 13921-1-AP RRID:AB_1851479 |
| Cav3.1              | Alomone| # ACC-021 RRID:AB_2039779 |
| Cav1.3              | Alomone| # ACC-005 RRID:AB_2039775 |
| Cx30.2              | Atlas antibodies | # HPA015024 RRID:AB_1847132 |
| Synapsin            | Enzo   | # ADI-905-782-100 RRID:AB_2039598 |
| Cardiac Troponin    | R&D Systems | #ab137346 |
| Cleaved Caspase-3   | Cell Signaling 9602S | # NB110-2546 RRID:AB_792426 |
| Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 | Thermo Fisher Scientific | #A-21202 RRID:AB_141607 |
| Alexa Fluor 488 AffiniPure Donkey Anti-Guinea Pig IgG (H+L) | Jackson Immunoresearch Labs | #706-545-148 RRID:AB_2340472 |
| Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594 | Thermo Fisher Scientific | #A-21203 RRID:AB_2535789 |
| Donkey anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 594 conjugate | Thermo Fisher Scientific | #A-21207 RRID:AB_141637 |
| Donkey anti-Rabbit IgG (H+L) Secondary Antibody, Alexa Fluor 647 conjugate | Thermo Fisher Scientific | #A-31573 RRID:AB_2536183 |
| Donkey anti-Mouse IgG (H+L) Secondary Antibody, Alexa Fluor 647 | Thermo Fisher Scientific | #A-31571 RRID:AB_162542 |
| Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 | Thermo Fisher Scientific | #A-21447 RRID:AB_2535864 |
| Donkey anti-Chicken IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 | Jackson Immunoresearch Labs | #703-545-155 RRID:AB_2340375 |
| Donkey anti-Sheep IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647 | Thermo Fisher Scientific | #A-21448 RRID:AB_10374882 |
| DAPI                | Santa Cruz | #sc-3598 |
| **Chemicals, peptides, and recombinant proteins** | | |
| Y-27632             | MedchemExpress | #HY-10583 |
| ChIR99021           | Stem-RD | #13122 |
| Cucurbitacin        | Sigma Aldrich | #C4493 |
| WAS402              | Tocris | #3300 |
| Retinoic acid       | Sigma Aldrich | #R2625-500MG |
| AG 490              | Sigma Aldrich | #T3434 |

(Continued on next page)
Continued

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Chidamide           | Selleckchem | #S8567     |
| Doxorubicin Hydrochloride | Sigma Aldrich | #D1515   |
| Physcion            | Sigma Aldrich | #17797     |
| XAV939              | Tocris | #3748      |
| IWP                 | Tocris | #3533      |
| IWP2                | R&D Systems | #3533/50   |
| Alrestatin          | Selleckchem | SS803      |
| Epalrestat          | Sigma Aldrich | SMLO527    |
| Cayman Epigenetic Chemical library | Cayman | #11076    |
| Prestwick Drug Collection | Prestwick |            |
| Recombinant Human BMP-4 Protein | R & D Systems | #314-BP    |
| Activin A           | R&D Systems | #338-AC-500/CF |

Deposited data

| Deposited data | DATABASE | IDENTIFIER |
|----------------|----------|------------|
| RNA-seq        | GEO database | GSE118085  |
| scRNA-seq      | GEO database | GSE119682  |

Experimental models: Cell lines

| Experimental models | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Human: H9 (WA-09) hESC line | WiCell Research Institute | NIHhESC-10-0062 |
| Human: iPSC line #5 | New York Stem Cell Foundation | N/A |
| Human: iPSC line #33 | New York Stem Cell Foundation | N/A |

Oligonucleotides

| Oligonucleotides | SOURCE | IDENTIFIER |
|------------------|--------|------------|
| Primers for hNPPA, hTBX3, hTBX18, hTBX5, hNIX2.5, hISL1, hHCN4, hSHOX2 | This paper | N/A |
| Primers for SHOX2-GFP genotyping, see Table S2 | This paper | N/A |

Software and algorithms

| Software and algorithms | SOURCE | IDENTIFIER |
|-------------------------|--------|------------|
| Rstudio                 | Rstudio | https://rstudio.com |
| Seurat R package v3.1.4 | Butler et al., 2018 | https://satijalab.org/seurat/ |
| Adobe illustrator CC2017 | Adobe | https://www.adobe.com/product/photoshop.html |
| Graphpad Prism 6        | Graphpad software | https://www.graphpad.com |

Other

| Other | SOURCE | IDENTIFIER |
|-------|--------|------------|
| mTeSR1 Complete Kit | Stem Cell Technologies | #85850 |
| Penicillin-Streptomycin | Gibco Thermo Fisher | #15070063 |
| GlutaMAX Supplement | Thermo Fisher Scientific | #35050079 |
| ITS-X          | Thermo Fisher Scientific | #51500056 |
| Accutase       | Stemcell Technologies | # 07920 |
| ReleSR         | Stemcell Technologies | # 05872 |
| B27 minus insulin | Thermo Fisher Scientific | # A1895601 |
| Matrigel       | Coming | #354234 |
| N2             | Stem Cell Technologies | #7156 |
| B27            | Life Technologies | #12587-010 |
| DMEM/F12       | Thermo Fisher Scientific | # 10565018 |

Deposited data

| Deposited data | SOURCE | IDENTIFIER |
|----------------|--------|------------|
| Single cell RNA-seq | GEO | GSE119682  |
| RNA-seq         | GEO | GSE118085  |
RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Shuibing Chen (shc2034@med.cornell.edu).

Materials availability
The dual knock-in SHOX2:GFP; MYH6:mCherry reporter line will be available under MTA.

Data and code availability
scRNA-seq and RNA-seq data are publicly available from the GEO repository database with accession number GSE119682 (scRNA-seq) and GSE118085 (RNA-seq), respectively.

Code availability: None.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines
hESC line H9 (Female) and the derivative (SHOX2:GFP; MYH6:mCherry) line, and human iPSC clones (iPSC #5, Female; iPSC #33, Female, New York Stem Cell Foundation) were grown on matrigel-coated plates and maintained in mTeSR medium (Stem Cell Technologies). hESCs and iPSC work has been approved by Tri-SCI ESCRO committee and conforms to the relevant regulatory standards.

METHOD DETAILS

Sinoatrial and ventricular myocyte induction
Differentiation was initiated 72 h after plating when the culture was approximately 80% confluent.

Step 1: Cells were differentiated with 1.5 μM CHIR99021 (CHIR, Stem-RD), 20 ng/mL BMP4 and 20 ng/mL Activin A in RPMI (Cellgro) supplemented with B27 minus insulin, 2 mM GlutaMAX, 1x NEAA and 1x Pen/Strep for 3 days (RB27-INS).

Step 2: For hPSC-V cardiomyocyte differentiation, cells were treated for an additional 3 days with 5 μM XAV939 (Tocris). From day 6 onward, hPSC-V differentiation were carried out in RPMI supplemented with B27, 2 mM GlutaMAX, 1x NEAA and 1x Pen/Strep (RB27+INS). The differentiation efficiency was approximately 85–90% (Tsai et al., 2020).

For hPSC-SAN differentiation, Step 1 was followed by addition of 0.1 μM cucurbitacin (Sigma Aldrich), 1 μM retinoic acid, 5 μM SU5402 (SU, Tocris) in RB27-INS from day 3–6. 5 μM XAV939 was added from day 5–6. From day 6–9, cells were cultured with 5 μM Tyrophostin AG 490 (Sigma Aldrich) in RB27+INS. At day 7–9, 5 μM chidamide was added to the differentiation cocktail. After day 9, hPSC-SAN differentiation was carried out in RPMI supplemented with B27, 2 mM GlutaMAX, 1x NEAA and 1x Pen/Strep (RB27+INS).

FACS and immuno-fluorescence
For FACS analysis, the cells were dissociated with Accutase (Innovative Cell Technologies) for 15 min at 37°C and fixed and permeabilized with the FOXP3 Fixation/Permeabilization set (eBiosciences). Subsequently, they were washed, blocked and permeabilized using FOXP3 permeabilization buffer (eBiosciences) according to the manufacturer’s protocol. The cells were stained with primary (overnight 4 degrees) and secondary (45 min, room temperature) antibodies and analyzed using a flow cytometer. For immuno-fluorescence, cells were fixed with 4% PFA (eBiosciences) for 20 min at room temperature. Next, cells were blocked and permeabilized in 5% horse serum (Invitrogen), 0.3% Triton X in PBS for 1 h at room temperature, followed by primary antibody incubation overnight. After several rounds of washing, secondary antibodies were added for 1 h at room temperature. Nuclei were stained with DAPI. Primary antibodies and working dilutions are listed in Table S3.

Gene expression analysis
Total RNA was isolated using the Qiagen RNeasy mini kit according to manufacturer instructions. cDNA synthesis was performed using TAKARA PrimeScript 1st strand cDNA synthesis kit. qRT-PCR reactions
were generated using Roche SYBR green PCR mix (Roche). Each data point represents at minimum 3 independent biological replicates. Primary antibodies and working dilutions are listed in Table S2.

**Generation of the SHOX2:GFP; MYH6:mCherry dual hESC reporter line**

The α-MHC:mCherry reporter line was created in our previous published studies (Tsai et al., 2020). For targeting the SHOX2 locus, sgRNA sequences were designed using the website http://crispr.mit.edu/ and targeted the sequence: 5’- GGCGTTGGCGTCACAGACCC-3’. sgRNA was cloned into the PX459 vector (Addgene, plasmid # 42230 (Cong et al., 2013)). For constructing the SHOX2:GFP donor plasmid, the NLS-eGFP (from Tol2 kit plasmid), codon optimized P2A, and left and right homology arms (from H9 genomic DNA) were PCR amplified and assembled using the In-Fusion kit (Clontech) and then cloned into the zero-blunt plasmid (Thermo Fischer). H9 hESCs were disassociated into single cells with Accutase for 5-7 min. Cells were electroproporated using human stem cell nucleofector kit2 (Lonza, #VPH-5022) according to the manufacturer guidelines. Briefly, 1 million cells were resuspended in 100 μL nucleofector mix to which was added 2 μg CRISPR targeting plasmid and 4 μg donor plasmid. Cells were re-plated on matrigel coated plates with ROCK inhibitor. Two days after electroporation cells were treated with puromycin (0.5 μg/mL) for two days. About two weeks after electroporation individual colonies were picked and expanded for PCR genotyping and sequencing for the reporter lines. Note that both reporter genes are expressed with a self-cleaving sequence and do not generate fusion proteins. The sequences for all primers used for generating the donor plasmids are listed in Table S2.

**Bulk RNA sequencing**

For RNA sequencing, total RNA was isolated using the Agilent nano kit according to manufacturer instructions. The quality of RNA samples was examined using an Agilent bioanalyzer. cDNA libraries were generated using TruSeq RNA Sample Preparation (Illumina). Each library was sequenced using a single-read 50 bp in HiSeq4000 (Illumina). Gene expression levels were analyzed with TopHat and Cufflinks by the Weill Cornell Genomic Core facility. Raw data were normalized to that from SHOX2:GFP negative cells to define fold-change. Gene lists were filtered based on expression level differences ≥4 or ≤-4. Murine sinoatrial and atrial tissues were obtained from a published database (GSE65658). False discovery rate (FDR) q values <0.25 or nominal (NOM) p values <0.05 were considered significant. The expression data were normalized for each gene by subtracting their mean across samples and then dividing by their standard deviation. The normalization was performed separately on three sets of samples (D25 samples, D40 samples and public samples) to remove batch effects introduced by experiments. The heatmap plot was generated using heatmap.2 in the R “gplots” package.

**Single cell RNA sequencing**

Cells were differentiated into hESC-SAN or cardiomyocyte (V-cell) lineages. hESC-SAN cells were sorted based on GFP, while HPSV cells were sorted based on mCherry. Single-cells were captured using a Drop-Seq system. The two aqueous suspensions - the single-cell suspension and the barcoded primer beads suspended in a lysis buffer, were loaded into 3 mL plastic syringes (BD) respectively. Droplet generation oil (BioRad) was loaded into a 10 mL plastic syringe (BD). The three syringes were connected to a 125 μm co-flow device by 0.38 mm inner-diameter polyethylene tubing (Scientific Commodities, Inc.), and injected using syringe pumps (KD Scientific), resulting in ~125 μm emulsion drops with a volume of ~1 nL each. The flow was monitored under an inverted microscope. Droplets were collected in 50 mL falcon tubes; the collection tube was changed out after every 1 mL of combined aqueous flow volume. During droplet generation, the beads were kept in suspension by continuous, gentle magnetic stirring (V&P Scientific). The uniformity in droplet size and the occupancy of beads were evaluated by observing aliquots of droplets under the microscope. The oil from the bottom of each aliquot of droplets was removed after which 30 mL 6X SSC (Life Technologies) at room temperature was added. Collected droplets were further purified by adding perfluorooctanol in 30 mL 6X SSC to destabilize the oil-water interface and the microparticles were disrupted. RNA-hybridized beads were extracted. The extracted beads were then washed and resuspended in a reverse transcriptase mix, followed by a treatment with exonuclease I to remove unhybridized, non-extended primers. The beads were then washed, counted, and aliquoted into PCR tubes for PCR amplification. 15-18 cycles of PCR amplification were applied to reach cDNA concentration at 400-1000 pg/μL at pre-PCR test step. The final PCR products were purified and pooled, and the amplified cDNA quantified on an Agilent 2100 BioAnalyzer High Sensitivity Chip. The cDNA was fragmented and amplified for sequencing with the Nextera XT DNA sample prep kit (Illumina) using custom primers. The libraries were purified, quantified, and sequenced on the Illiuna NextSeq 500 at > 50,000 reads/cell.
After quality assessment of the sequencing with FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), all single-cell RNA-seq samples were processed according to the Drop-seq Computational Protocol with the Drop-seq Toolkit v1.2 created by the McCarroll Lab (http://mccarrolllab.com/dropseq/). Sequences were first tagged with their corresponding cell and molecular barcodes (unique molecular identifiers, UMI) using dropseq TagBamWithReadSequenceExtended. Cell and molecular barcodes were required to have a minimum base quality of 10 across their lengths, otherwise they were discarded. SMART adapter sequences were trimmed from the 5’ end of each read whenever there were at least five continuous bases exactly matching the SMART adapter (dropseq TrimStartingSequence). Putative poly(A) tails were trimmed from the 3’ end whenever there were stretches of at least 6 adenosines with zero mismatches (dropseq PolyATrimmer). The UMI-tagged and trimmed reads were then aligned with the human reference genome (GRCh38) using STAR (v2.4.2a) (Dobin et al., 2013). The final file containing alignment information plus the cell and UMI tags was generated with Picard MergeBamAlignment and Picard MergeBamAlignment. Sequences that mapped to more than one locus were excluded from further analysis by filtering reads with a mapping quality lower than 10.

To determine the numbers of detected transcripts per gene and cell, the overlaps of the aligned reads with gene annotation (Ensembl release 76) were counted and summarized using dropseq TagReadWithGeneExon and dropseq DigitalExpression. Individual gene expression matrices from each sample were combined in R to create a single gene count matrix for all cells and genes.

The count matrix was read into R; basic quality metrics such as the distributions of the numbers of gene counts per cell and gene were determined using in-house scripts and the Scater package (McCarthy et al., 2017). We excluded cells with fewer than 100 UMI or fewer than 200 genes per cell, and more than 15% of the UMI mapping to mitochondrial genes. This resulted in 1,571 GFP-negative and 2,057 GFP-positive cells (representing about equal numbers per technical replicate). After filtering the cells, we excluded genes that were expressed in fewer than 5 cells per condition. Additionally, genes had to be covered by at least 10 UMI across all cells of the same condition. 13,448 genes met these filtering criteria and were used for downstream analyses. The UMI counts were adjusted for the differences in the individual cell coverage by dividing every gene count by the total number of UMI counts of the respective cell (library size) and multiplying with the median of all library sizes. These scaled UMI counts were then further de-noised using the MAGIC algorithm. All expression values shown in the manuscript are log10-transformed values (with an offset of 0.01 to avoid taking the log of zeros) after applying MAGIC (https://doi.org/10.1101/111591). t-SNE was performed with the Rtsne package (https://github.com/jkrijthe/Rtsne). Hierarchical clustering was applied to all cells and a subset of genes and visualized with the pheatmap function (https://CRAN.R-project.org/package=pheatmap). Genes and cells were clustered with the base R hclust function using Euclidean distances and the agglomeration function “complete”.

**Adult heart single cell analysis**

We used an existing single cell RNA-seq dataset generated from heart tissue biopsies of adult healthy individuals (Tucker et al., 2020). Briefly, cells with poor quality were removed if 1) they had less than 100 genes or greater than predefined upper outlier cutoff, 2) the number of UMI for the cell was greater than predicted upper outlier cutoff, or 3) the percentage of mitochondrial gene content was greater than 5%. Please refer to the original manuscript for detailed information about the cutoffs.

Data normalization, variable gene identification and data scaling was done using Seurat V3. During data scaling, variability due to read coverage (nCount_RNA), cell cycle effects (S.Score and G2M.score) and the percentage of mitochondrial and ribosomal gene reads were regressed out. The Seurat featureplot and Dotplot functions were used to visualize genes of interest, and the uniform manifold approximation and projection (UMAP) coordinates provided by the original authors were used for featureplot visualizations.

Clusters 0 to 16 corresponds to 0) Adipocyte; 1) Atrial Cardiomyocyte; 2) Cytoplasmic Cardiomyocyte I; 3) Cytoplasmic Cardiomyocyte II; 4) Endothelium I; 5) Endothelium II; 6) Fibroblast I; 7) Fibroblast II; 8) Fibroblast III; 9) Lymphocyte; 10) Macrophage; 11) Neuronal; 12) Pericyte; 13) Vascular smooth muscle; 14) Ventricular Cardiomyocyte I; 15) Ventricular Cardiomyocyte II; and 16) Ventricular Cardiomyocyte III.
Calcium imaging

hESC-derived cells were plated on ibidi plates coated with gelatin. The cells were loaded with 2 μM Fluo-4 AM dissolved in 1:1 (v/v) of 20% Pluronic®-F127 and DMSO with stock concentration of 1 mM for 45 min at RT in Tyrode solution consisting of: 140 mM NaCl, 5.4 mM KCl, 1 mM MgCl₂, 1.8 mM CaCl₂, 10 mM glucose and 10 mM HEPES at pH 7.4. Calcium transients of hESC-derived beating cardiac clusters were recorded on a heated stage using a confocal scanning microscope (Zeiss LSM 710) at intervals of 200 ms (5 frames per second). They were then quantified as the background subtracted fluorescence intensity changes normalized to the background subtracted baseline fluorescence using MetaXpress software. For the experiments involving drug treatment (Isoproterenol 1 μM, Carbachol 5 μM and Ivabradine 1 μM), the cells were treated with either vehicle or drug and the imaging was performed within 10 s (for Isoproterenol and Carbachol) or 5 min (Ivabradine) of adding the drug.

Cellular electrophysiology and characterization

Spontaneous action potentials were recorded over 10 s with the perforated patch clamp technique using an AM-Systems (WA, USA) model 2400 amplifier in current-clamp mode and the software platform Real-Time eXperiment Interface (RTXI; www.rtxi.org) (Ortega et al., 2014; Patel et al., 2017). Cells were superfused at 35 ºC with a Tyrode’s solution containing in mM: 137 NaCl, 5.4 KCl, 2 CaCl₂, 1 MgSO₄, 10 HEPES, and 10 glucose at pH 7.35 (NaOH). Whole cell access was achieved using 480 μg/mL Amphotericin-B (Sigma-Aldrich, MO, USA) in a pipette solution also containing in mM: 5 NaCl, 20 KCl, 120 K-aspartate, and 10 HEPES at pH 7.2 (KOH). A liquid junction potential of 14 mV was corrected. Pipettes were pulled from 1.5 mm capillary tubes (AM-Systems, WA, USA) to a resistance between 2.5 – 3.5 MΩ. Cells were characterized as nodal-like if they exhibited spontaneity and a slow upstroke (dV/dt<sub>max</sub> < 30 V/s). hESC-V voltage recordings were used for comparison (Tsai et al., 2020).

Whole-cell currents were recorded with an extracellular solution containing (mM) 140 NaCl, 5.4 KCl, 1.8 CaCl₂, 1 MgCl₂, 10 glucose, and 10 HEPES, pH 7.4. Borosilicate glass patch pipettes (resistance 1.5–2.5 MΩ) were filled with internal solution containing (mM): 125 K-gluconate, 20 KCl, 5 NaCl, 1 MgCl₂, 5 MgATP, and 10 HEPES, with pH adjusted to 7.2 with KOH. All chemicals were from Sigma-Aldrich. The signal was filtered at 5 Hz. The liquid junction potentials were not corrected. Data are presented as mean ± SEM.

Chemical screen

The epigenetic library was purchased from Cayman Chemicals (Item N° 11076). The Epigenetics Screening Library contains compounds that modulate the activity of methyltransferases, demethylases, histone acetyltransferases, histone deacetylases, and acetylated histone binding proteins. The dual-reporter hESC line was differentiated into mesoderm lineage by adding Activin, BMP4 and CHIR99021 to the differentiation medium for 3 days and then switched to differentiation medium containing Retinoic Acid, Cucurbitacin, and SU5402 for another 3 days. After 6 days, the cells were replated into 384 well plates and the medium was changed to RPMI+ complete B27 containing Tyrphostin AG490. After 24 h, small molecules from the chemical library were added to the media and incubated for 48 h. The medium was changed to RPMI plus complete B27 on day 9 and it was changed every 48 h afterwards. Cells were imaged using a high content imaging system on day 14 and assessed for GFP percentage. The compound with least toxicity and highest percentage of GFP was selected as potential epigenetic modulator compound to enhance differentiation of sinoatrial nodal cells.

QUANTIFICATION AND STATISTICAL ANALYSIS

N = 3 independent biological replicates were used for all experiments unless otherwise indicated. n.s. indicates a non-significant difference. p-values were calculated by unpaired two-tailed Student’s t-test or one-way ANOVA test unless otherwise indicated. *p < 0.05, **p < 0.01 and ***p < 0.001. Data are represented as mean ± SEM.