SUPPLEMENTAL MATERIAL

TBX20 Improves Contractility and Mitochondrial Function during Direct Human Cardiac Reprogramming

Yawen Tang*[1], Sajesan Aryal*[2,3], Xiaoxiao Geng[1], Xinyue Zhou[2,3], Vladimir G. Fast[1], Jianyi Zhang[1], Rui Lu[2,3#], and Yang Zhou[1#],

[1] Department of Biomedical Engineering, Heersink School of Medicine, School of Engineering, University of Alabama at Birmingham, Birmingham, AL, United States

[2] Department of Medicine, Division of Hematology and Oncology, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[3] O'Neal Comprehensive Cancer Center, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

Short title: TBX20 promotes direct human cardiac reprogramming

* Contributed equally

Corresponding author

Yang Zhou, Ph.D.

Department of Biomedical Engineering

Heersink School of Medicine & School of Engineering

University of Alabama at Birmingham

1670 University Blvd

Birmingham, AL 35243

Email: yangzhou@uab.edu

Phone: (205) 934-1227

Rui Lu, Ph.D.

Division of Hematology/Oncology, Heersink School of Medicine

O'Neal Comprehensive Cancer Center

University of Alabama at Birmingham

1824 6th Ave S

Birmingham, AL 35243

Email: ruilu1@uabmc.edu

Phone: (205) 975-9878

SUPPLEMENTAL METHODS

The data, analytic methods, and study materials will be made available to other researchers for purposes of reproducing the results or replicating the procedure. No animals were used in this study. The Institutional Review Board of the University of Alabama at Birmingham reviewed our use of human cells in this study and provided a determination of not human subjects research.

Fibroblast Culture

Four fibroblast cell lines were included in this study. The H9F was the human ESC-derived fibroblast line[6]. H9Fs were cultured in DMEM medium supplemented with 20% FBS and 1% Pen-strep (HDF medium). The human primary cardiac fibroblasts (HCF) and human primary dermal fibroblasts (HFF CCD-1079SK and HFF CCD-1112SK) were purchased from ATCC. HCFs were cultured with FBM2 bullet kit purchased from ATCC (HCF medium), and HFFs were cultured using IMDM with 10% FBS and 1% Pen-strep (HFF medium).

Virus Package

For reprogramming, pMXs-puro-hMGT[7], pMXs-MEF2C, -HAND2, -GATA4, -TBX5[6], pBabe-miR-133 plasmids[34], pBabe-TBX20 or pBabe-EV (constructed in lab) were used as indicated. The package plasmids gag/pol and VSV.G were purchased from Addgene. For lentiviruses expressing desired shRNAs, pMD2G and psPAX2 (purchased from Addgene) were used as package plasmid. Lentiviruses for CRISPR KO experiments were packaged using pLentiCRISPRv2Blast (Addgene #98293) together with pMD2.G and psPAX2. Corresponding viral and package plasmids were transfected with NanoFect (ALSTEM) into 293T cells. Retrovirus or lentivirus particles were collected on 72 and 96 hours after transfection and precipitated with 8% PEG6000.

Direct Cardiac Reprogramming

Fibroblasts were seeded into 24-well plates at a density of $4X10^4$ cells per well or scale-up into 6-well plates and 10-cm dishes accordingly one day before reprogramming. On day 0, fibroblasts

were infected with MGT133-EV, MGT133-TBX20 cocktails or other combinations as indicated in DMEM medium supplemented with 10% FBS and 20% Medium 199 (iCM medium). For the shRNA knockdown and CRISPR KO assays, lentiviruses were applied on reprogramming day 1. On day 2, cells were selected with puromycin at a concentration of 10 μg/mL for 4 days. On day 6, puromycin was removed from iCM medium. On day 10, the culture medium was changed to RPMI-1640 medium supplemented with 2% B27 supplement, 2% FBS, 0,05% bovine serum albumin (BSA), 50 μg/mL ascorbic acid, and 1X NEAA (hiCM Medium). Further evaluations were performed on day 14 of reprogramming.

Immunofluorescence Staining

On day 14 of reprogramming, hiCMs were re-seeded in gelatin-coated 8-well chamber slides (MilliporeSigma™ Millicell™ EZ Slides). Cells were fixed with 4% PFA for 15 min at room temperature and permeabilized with 0.01% Triton X-100. The slides were then washed with 0.1% TWEEN20 in PBS (PBST) three times, blocked with 10% donkey serum (Sigma-Aldrich; D9663) for 30 min at room temperature. The blocked slides were then incubated with corresponding primary antibodies at 4 °C overnight. Antibodies used were anti-cTnI antibody (1:400, Abcam, ab38210), anti-α-Actinin antibody (1:200, Sigma-Aldrich, A7811), anti-αMHC antibody (1:100, DSHB, MF 20), anti-cTnT antibody (1:200, R&D systems, MAB1874), anti-TBX20 antibody (1:100, Invitrogen, PA-40669), and anti-Ki67 antibody (1:200, Abcam, ab16667). After incubation, wash the slides three times with PBST and incubate in corresponding secondary antibodies (1:500, Jackson ImmunoResearch) at room temperature for ~1 hour. Wash the slides with PBST three times and mount the slides with DAPI-containing mounting medium (Vector Laboratories; H-1200). For TUNEL staining, prepare TUNEL reaction mixture as described in manufacturer guidance (In Situ Cell Death Detection Kit, Millipore Sigma #12156792910). After cell permeabilization, rinse slides twice with PBS. Following wash, dry the area around the sample and add 50 μL TUNEL reaction mixture to each sample. Incubate slide in a humidified atmosphere for 60 min at 37 °C in

the dark. Rinse the cells with PBS three times and mount the slides with DAPI-containing mounting medium. Images were obtained using Olympus IX83 fluorescence microscope or Olympus confocal microscope, and data analysis was performed with Image J software. Quantification was performed by counting the ratio of positively stained cells compared with DAPI positive total cells (~40 images were randomly taken under 20x magnification at the same exposure setting blindly).

Flow Cytometry

Cells were trypsinized with 0.05% trypsin on day 14 of the reprogramming and fixed with BD cytofix/cytoperm solution at 4 °C for 30 minutes. Wash the cells with 1X wash buffer and incubate in desired primary antibodies at room temperature for 1 hour. Antibodies used were anti-αMHC antibody (1:50, DSHB, MF 20), anti-α-Actinin antibody (1:400, Sigma-Aldrich, A7811), and anti-cTnT antibody (1:200, Abcam, ab91605). Following incubation, wash the cells once with 1X wash buffer and incubate in corresponding secondary antibodies (1:500, Jackson ImmunoResearch) for 1 hour at room temperature. The cells were then washed with 1X wash buffer and resuspended in 1% PFA. The flow cytometry assay was run with BD flow cytometer and data were processed with FlowJo.

RT-qPCR and Western Blotting

Total RNA was extracted using TRIzol reagent and reverse-transcribed into cDNA using Protoscript II Reverse Transcriptase (NEB). qPCR was set up using Power UP SYBR Green PCR Mix and run on the QuantStudio Real-Time PCR system. All primers used were listed in Table S1. Protein was collected using M-PER™ Mammalian Protein Extraction Reagent (Fisher Scientific) with 1X protease inhibitor and quantified using BCA assay. Samples were then denatured at 99 °C for 10 min and run through a 4-20% precast gel (BIORAD # 4561093). After transfer, the membrane was blocked with 5% dry milk (BIORAD # 1706404) at room temperature for 1 hour. Following, the membrane was incubated with desired primary antibodies at 4°C overnight.

Antibodies used were anti-MYOM2 antibody (1:1000, Abcam, ab233263), anti-MYH6 antibody (1:2000, ProteinTech, 22281-1-AP), anti-MYH7 antibody (1:2000, ProteinTech, 22280-1-AP), anti-MYBPC3 antibody (1:1000, ProteinTech, 19977-1-AP), anti-GAPDH antibody (1:2000, Abcam, ab22555), anti-MEF2C antibody (1:1000, Cell Signaling, 5030S), anti-GATA4 antibody (1:1000, Cell Signaling, 36966), anti-TBX5 antibody (1:1000, Santa Cruz, sc-515536), and anti-TBX20 antibody (1:1000, R&D systems, MAB8124-SP). The membrane was washed with 1X TBST (1X TBS with Tween 20) three times and incubated with corresponding secondary antibodies (1:2000, ProteinTech) for 1 hour at room temperature. After incubation, the membrane was washed with 1X TBST for three times and subjected to image.

Co-culture of hiCMs and iPSC-CMs

iPS derived cardiomyocytes (iPS-CMs) were differentiated and purified as published previously for further use[36]. On day 14 of reprogramming, hiCMs were labeled with GFP using GFP lentivirus. On day 16 of reprogramming, hiCMs and iPS-CMs were trypsinized with 0.05% Trypsin and seeded into gelatin coated 24-well plate at ratio of 1:20. The co-cultured cells were cultured in RPMI 1640 supplement with 10% B27 and 1% Pen Strep. Change medium every other day for 4 weeks. After co-culturing the cells for 4 weeks, videos of cell beating were obtained. Observed beating or beating-like GFP+ cells were counted, and percentages were calculated as beating GFP+ cells/total GFP+ cells in each video taken. 12 videos were taken for each well with biological triplicates.

Identification of TBX20 via Transcriptomic Comparison

To identify the possible missing key regulators contributing to the complete acquisition of cardiomyocyte identity, we compared the transcriptome profiles of human embryonic stem cell (ESC) derived fibroblasts (H9Fs), human dermal fibroblasts (HDFs), iCMs generated by 5 or 7 reprogramming factors[6], and cardiomyocytes derived from fetal hearts (enCMs) or differentiated from stem cells (H9CMs) from previously published dataset[6,34]. We identified more than 200

transcription factors and epigenetic regulators that were expressed significantly higher in enCMs and H9CMs than in H9Fs and HDFs[6]. Among them, 38 factors were not fully activated in iCMs direct reprogrammed with different reprogramming cocktails, compared to enCMs or H9CMs (also see details in **Supplemental Table S1**).

RNA Sequencing and Data Processing

Total RNA was extracted using TRIzol. RNA sequencing (RNA-seq) was performed in triplicate using independent biological samples. Sequencing libraries were generated using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) following the manufacturer's manual and applied on a HiSeq 4000 sequencer (Illumina) for a pair-end 150 bp reads sequencing run. RNA-seq reads were aligned to the Ensemble reference genome and gene model annotation using Tophat program. Differentially expressed genes were determined by Deseq2[37] with default parameters (fold change $>$ 2 and FDR adjusted P value $<$ 0.01). Gene ontology enrichment analyses have been performed with Enrichr (R package version 3.0)[38]. The expression changes of selected cardiac contractility genes were visualized by heatmaps and histograms. GSEA has been performed with *gseGO* function in clusterProfiler (R package version 4.4.1).

MitoTracker Immunofluorescence Staining

Cells were seeded onto 8-well chamber slide on day 14 of reprogramming and cultured for 24 hours. The cells were incubated in 50nM MitoTracker Red CMXRos (Cell Signaling #9082) for 45 min at 37 °C. After incubation, the cells were washed once with 1X PBS and fixed in ice-cold methanol for 15 min at -20 °C. Following fixation, the cells were washed three times with PBS for 5 min each. The fixed cells were blocked with 10% donkey serum for 1 h at room temperature and then incubated in cTnT (R&D System #MAB1874, 1:200) at 4 °C overnight. The primary antibody was rinsed three times with PBS for 5 min each and incubated with secondary antibody (Alexa fluor 488) for 1 h at room temperature. The cells were then washed three times with PBS and followed by nuclear staining using DAPI. Images were obtained using Olympus IX83 confocal

microscope. The Mitotracker signals were measured as Mitotracker positive pixel area for each DAPI identified using ImageJ. Differences between groups were examined for statistical significance.

Mitochondria DNA Assessment

Total DNA of cells was extracted using QIAmp DNA Mini kit (Qiagen #51304). The extracted DNA were then proceeded to real-time qPCR using primers mt-DNA_F 5′-CACCCAAGAACAGGGTTTGT; mt-DNA_R 5′-TGGCCATGGGTATGTTGTTA; B2_F 5′-TGCTGTCTCCATGTTTGATGTATCT; B2_R 5′-TCTCTGCTCCCCACCTCTAAGT with SYBR Green mix. The obtained mitochondrial DNA CT values were normalized to corresponding nuclear DNA CT values and then normalized to the control group.

Seahorse Metabolic Profiling

The Seahorse XF96 extracellular flux analyzer (Agilent Technologies) was used for mitochondria function assessment. The cells were seeded onto 96-well seahorse plate 48 hours prior to the assay with a density of 20,000 cells/well. One hour before the assay, culture media were changed to assay base media and incubated at 37 °C in a non-CO_2 incubator. The assay was then run according to manufacturer instructions (Agilent). The collected data were analyzed with Agilent Wave software and were graphed using Graphpad Prism 6.

Calcium Imaging and Quantification

After co-culture the hiCMs with hiPSC-CMs for 1 month, the cells were stained with Calbryte 520 AM dye (AAT Bioquest, #20650) with a final in-well concentration of 5 μM according to manufacturer instruction. After staining, 10 videos per well were obtained using microscope. Each video was converted into a format readable for ImageJ using FFmeg tool. The mean fluorescence intensities (F) of cells and background ($Background$) were measured using ImageJ for all slides obtained from video (0.03 s/slide). Then calcium transients were calculated using the equation

$\Delta F/F_0 = \frac{F-F_{base}}{F_{base}-Background}$, in which, F_{base} is the minimum fluorescent signal among all measurement. Peak $\Delta F/F_0$ is the maximum $\Delta F/F_0$ obtained from each calcium trace.

Optical Mapping of Membrane Potential

After one month of co-culture, hiCMs generated with MGT+EV or MGT+TBX20 were selected with puromycin (1 ug/ml) for four days and applied to optical mapping as previously reported[39]. Briefly, the cells were transferred into a perfusion chamber mounted on an inverted microscope and perfused with Hank's balanced salt solution at a temperature of 36 °C. To measure action potentials, cells were stained with 5 μM of a fluorescent voltage-sensitive dye RH-237 for 5 min. Dye fluorescence was excited within a spectral range of 560/55 nm and measured at >650 nm. Optical signals were recorded using a 16×16 photodiode array (Hamamatsu) at a spatial resolution of 55 μm per diode as previously described[39] and digitally filtered to increase the signal-to-noise ratio. Activation times were measured at 50% of the maximum action potential amplitude. The durations of the action potentials were measured as time intervals between the activation and recovery times determined at 50% and 80% levels of signal recovery (APD50 and APD80, respectively).

Dual-luciferase Reporter Assay

MYBPC3 and *MYL4* enhancer sequences (supplemental table) were PCR amplified from human genomic DNA and cloned into pGL3 promoter luciferase vectors (Promega). All sequencing validated constructs were co-transfected with a Renilla expressing plasmid, pRL CMV (Promega), at ratio of 1000:1 using NanoFect into 293T cells. 48 hours after transfection, luminescence was then assessed using Dual Glo luciferase kit according to the manufacturer's protocol (Promega). All data shown are average from at least two biological replicates and two technical replicates, representing fold-change in luciferase activity compared to Empty controls and normalized for Renilla transfection control.

Single Cell RNA-sequencing Data Analysis

Single-cell RNA-seq outputs were processed using the Cell Ranger (10X Genomics) suite versions 6.0.1. Raw reads were mapped to a merged and modified genome of hg 19 (GRCh37). For mapping reads from we modified the hg19 reference genome by merging sequences of MGTpuro-, mir-133- and TBX20- expressing vectors. These sequences were also annotated in the GTF (gene transfer format) file and rebuilt the reference genome by the *mkref* command implemented in the Cell Ranger software. Raw base count files were demultiplexed and processed by *mkfastq* implemented in the Cell Ranger software.

Raw reads were then mapped to the customized reference genome by using Cell Ranger *count*. Secondary analysis was conducted using the Seurat (version 4.0.5) package[40]. We filtered the data for cells expressing more than 200 genes. EV was then filtered with MGTpuro expression > 0 and miR-133 expression > 0. Similarly, TBX20 was filtered with MGTpuro, miR-133 and TBX20 expression each > 0. To avoid analyzing doublets or dead cells, barcodes with a high mitochondrial content (> 35%), a low number of genes detected (< 1,000), or a high number of UMIs (> 50,000) were removed from downstream analysis. In total we analyzed 1552 EV- and 1048 TBX20- treated cells. EV and TBX20 cell populations were merged to obtain a matrix. The combined matrix was log-normalized and scaled. Highly variable feature selection was obtained using *FindVariableFeatures* in Seurat using default settings. To visualize the data, we performed Uniform Manifold Approximation and Projection (UMAP) to achieve dimensional reduction and unbiased clustering.

The expression distribution of selected genes was demonstrated by violin plots in different clusters representing EV and TBX20 treated cells with various expressions of MGT. Wilcox test was performed between the comparison groups. Differentially expressed genes for each of the identity classes in the combined EV and TBX20 cell population was obtained using *FindAllMarkers* command returning only positive markers and log scale fold change set to 0.05.

Gene set variation analysis (GSVA) was performed to assess gene expression signatures and gene sets from Molecular Signatures Database (MSigDB) were used. Top gene gets and their expression were visualized with violin plots and/or with heatmaps.

Differential trajectory inference and pseudotime analysis was performed with the condiments workflow[41]. Briefly, EV and TBX20 cell population expression data were integrated to obtain an integrated data assay using Seurat's *IntegrateData* function. The data was then scaled, clustered, and visualized as UMAP. The data was converted to single cell analysis format object using *SingleCellExperiment* command in Seurat. Slingshot[42] was used for trajectory inference and the tradeSeq gene expression model was fitted using the inferred trajectory[43]. Heart Global dataset (HHA)[21] with combined single cell and single nuclei RNA-seq data of 485K cells with annotations was downloaded as h5ad file from heartatlas.org. Similarly, Heart Atrial Cardiomyocytes (HA) and Heart Ventricular Cardiomyocytes (HV) data were also downloaded as h5ad files. The h5ad file was converted to h5Seurat file and loaded in Seurat object to do further analysis. Symphony[22] was used to map our hiCMs dataset (query) to the human cardiomyocyte dataset (reference) in HHA[21]. For sub-population markers in Atrial (HA) and Ventricular (HV), *FindAllMarkers* in Seurat package was used and markers of each identity class in HA or HV were filtered (pct.1 > 0.3) and combined to make a gene list. This gene list was used to perform (GSVA) on the combined data matrix of EV- and TBX20- treated cells. Violin plots of GSVA scholar were used to view expression between comparison groups and p-value was reported using the Wilcox test where required.

Network analysis was executed using the Weighted Gene Co-expression Network Analysis (WGCNA) R package[23]. Top 2,000 most variable genes were used to define the transcriptome co-expression clusters. A signed (i.e. with direction) co-expression network was built using the Pearson correlation matrix, power was set to 6, the minimum height to merge modules was set to 0.15. Expression of each co-expression cluster genes was viewed in UMAP

using *FeaturePlot* in the Seurat package. Cytoscape[44] was used for gene network visualization.

CUT&Tag, CUT&RUN, and Data Analysis

CUT&Tag assay was performed as previously described[45]. In brief, 0.5 million cells were collected for each sample. The following primary antibodies were used: rabbit anti-TBX20 antibody (1:50; Thermo Fisher, PA5-40669), mouse anti-TBX20 antibody (1:50; R&D systems, MAB8124). Guinea pig anti-rabbit antibody (1:100; Antibodies Online ABIN101961) or rabbit anti-mouse antibody (1:100; Abcam, ab46540) was used as the secondary antibody. pAG-Tn5 was purchased from EpiCypher (15-1017) and used at 1:50 dilution for each 50 µL CUT&Tag reaction. CUT&RUN assay was performed as previously described[46,47]. In brief, 0.5 million cells were collected for each sample. The H3K4me1 antibody (Abcam, ab8895) was used at a 1:100 dilution for binding of the primary antibody. pAG-MNase was purchased from EpiCypher (15-1116) and used at 1:20 dilution for each 50 µL CUT&RUN reaction. Library construction was performed using the NEBNext Ultra II DNA Library Prep Kit from NEB (E7645S). Indexed samples were run using the Illumina HiSeq 4000 platform with paired-end sequencing. CUT&Tag or CUT&RUN reads were processed with ChIP-seq pipeline as previously described with modifications[47,48]. Briefly, raw reads were subjected to adapter removal (cutadapt 2.10) and mapped genome hg19 (bowtie2 2.4.1)[49] using --end-to-end --very-sensitive -–no-mixed -–no-discordant --phred33 -I 10 -X 700 parameters. Bigwig files were generated by bamCoverage[50] (version 3.3.0) and used for visualized in IGV[51] (version 2.8.6). Heatmaps were generated by deeptools[50] (version 3.3.0). SEACR[52] was used for peak calling. HOMER[53] was used for *de novo* motif analysis. GREAT[54] (version 4.0.4) was used for annotation analysis of identified peaks with distal distance set up to 500 kb. Genes bound and upregulated by TBX20 (MGT+TBX20 versus MGT+EV) were used to define Class I and Class II target genes: Class I genes were defined by genes not upregulated by MGT+EV as compared to H9F; Class II genes were defined by genes upregulated by MGT+EV as compared to H9F. In all cases, upregulation was defined by fold change > 2 and FDR adjusted *P* value < 0.01 as

determined in Deseq2.

ATAC-seq and Data Analysis

ATAC-seq was performed as previously described[55]. Briefly, 50,000 cells were washed once with cold phosphate-buffered saline (PBS) and resuspended with 50 µl cold ATAC-Resuspension Buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM Mg2Cl, 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin, and protease inhibitor), followed by centrifugation at 500 g for 5 mins to collect the cell nuclei. Tn5 (Illumina) transposition of nuclei pellets was carried out at 37°C for 30 min. The transposed DNA was purified using DNA purification columns (Zymo Research) and subjected to PCR amplification for library generation. ATAC-seq libraries were sequenced on the Illumina HiSeq 4000 platform for paired-end 150 cycles. All sequencing reads were mapped to the human genome (hg19) using the Bowtie2 (v2.2.3) and reads with a mapping score > 20 were selected for data analysis, after duplicated reads removed by SAMtools (v1.9). MACS2 were used to call ATAC-seq peaks, using default parameters in paired-end mode. Bigwig files were generated by bamCoverage[50] (version 3.3.0) and used for visualized in IGV[51] (version 2.8.6). Heatmaps were generated by deeptools[50] (version 3.3.0). SEACR[52] was used for peak calling.

Data Availability

All genomic datasets have been deposited to the Gene Expression Omnibus with accession number GSE193810 and are publicly available.

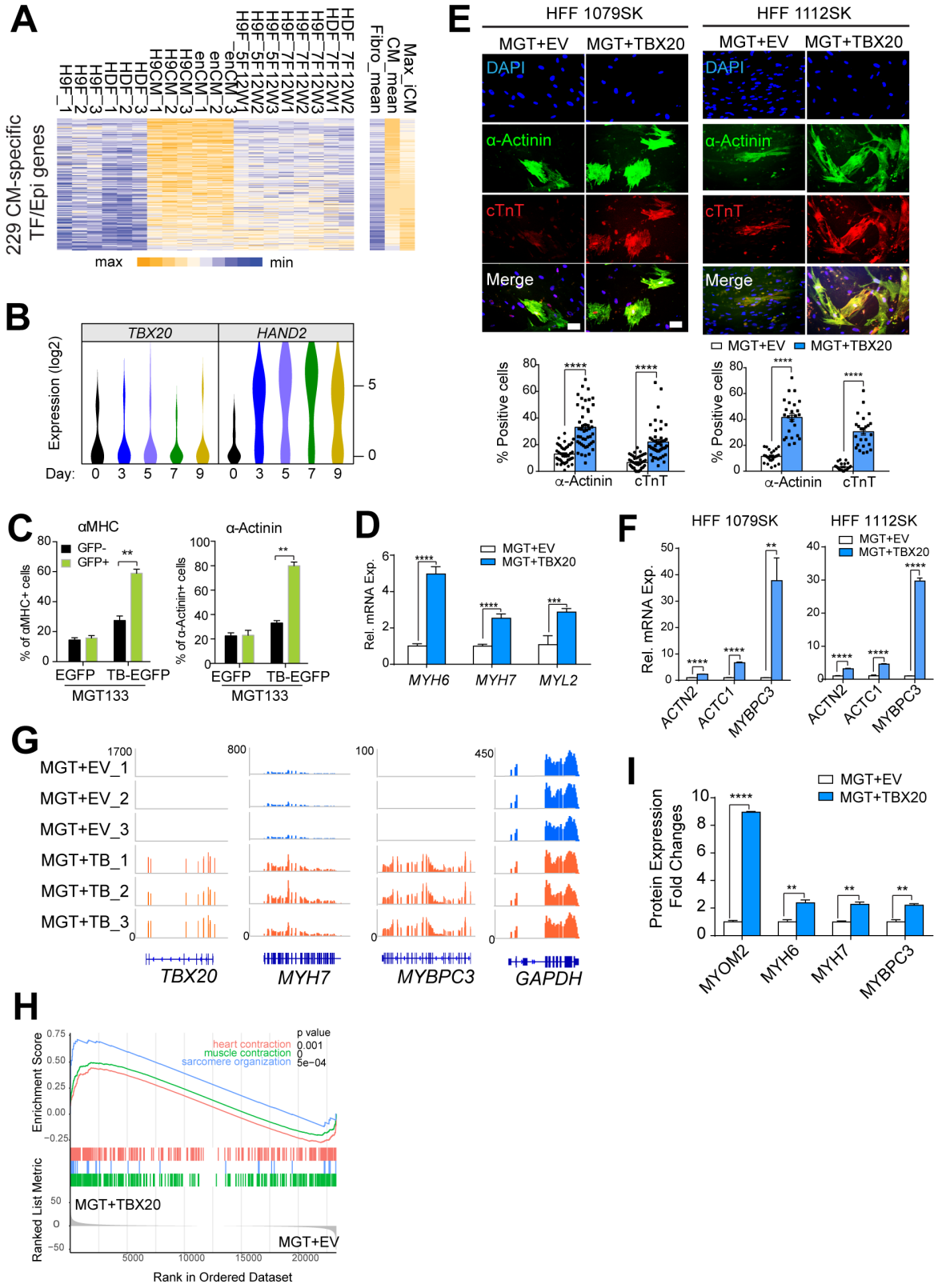
Statistical Analysis

Data were analyzed using Prism 8 (GraphPad) or the statistical language R (www.R-project.org) using RStudio. All results are reported as mean ± SEM. Data distribution and homogeneity of variance were tested using the Shapiro–Wilk and Levene tests, respectively. Differences between groups were examined for statistical significance using Student's *t*-test, or one-way ANOVA followed by Dunnett's multiple comparisons test (to a single control group), or Tukey's multiple comparisons test (among groups) where appropriate. For data that failed to Levene test, Welch

t-test or Welch ANOVA was used. For multiple group comparisons with more than two variables, two-way ANOVA was conducted, followed by either Tukey or Holm-Šídák's multiple comparisons test.Replicates and statistical tests are described in the figure legends. Differential gene expression in the scRNA-Seq data was determined using a two-sided Wilcoxon rank-sum test. The difference in scRNA-seq trajectory was determined by a Kolmogorov–Smirnov test. The overlapping enrichment analysis was determined by Fisher's exact test. $P < 0.05$ was considered significant.

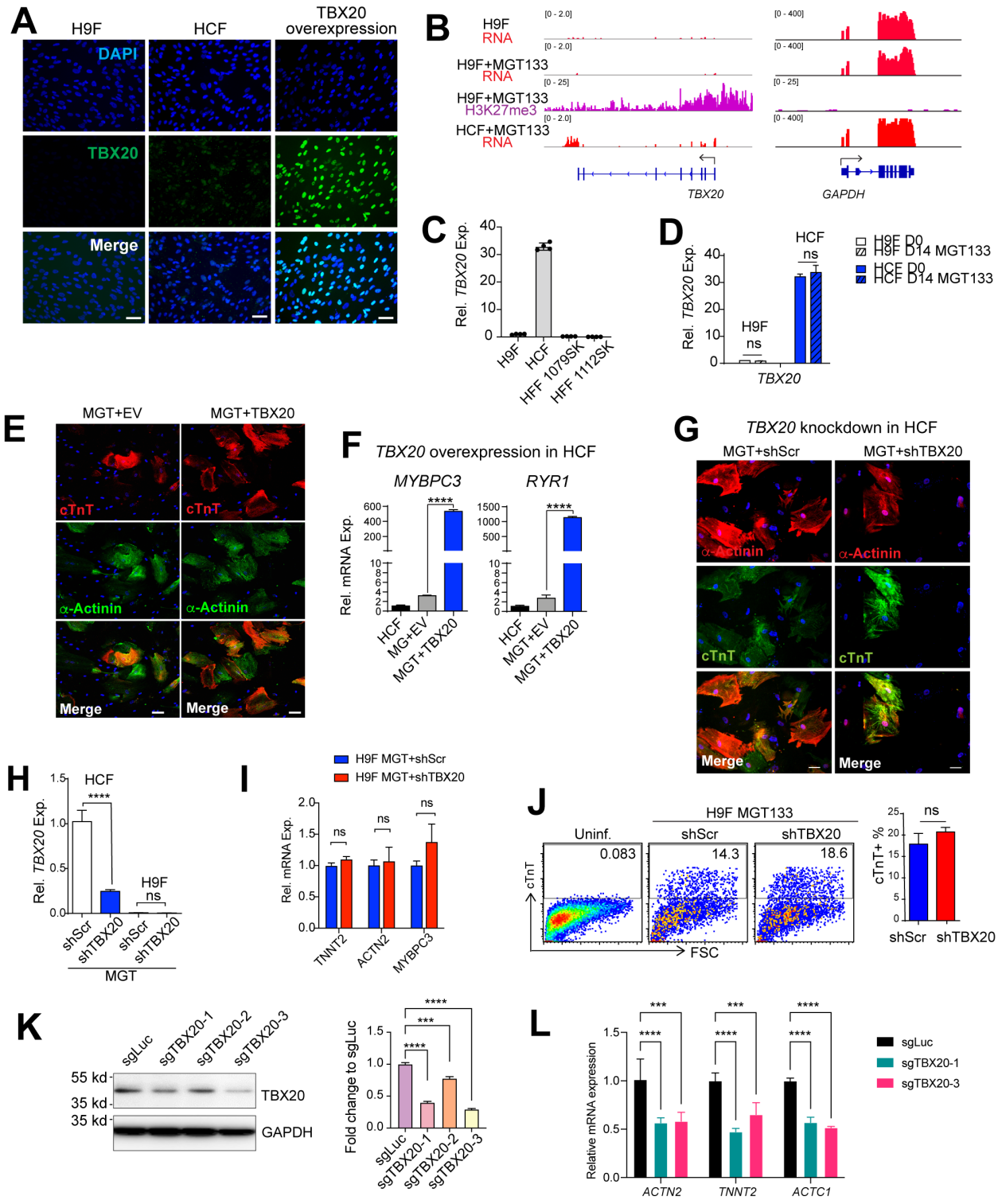
Supplemental Figure S1-S10

Figure S1


A
229 CM-specific TF/Epi genes
max
min
Fibro_mean
CM_mean
Max_iCM
B
TBX20
HAND2
Expression (log2)
Day:
C
αMHC
α-Actinin
GFP-
GFP+
% of αMHC+ cells
% of α-Actinin+ cells
EGFP
TB-EGFP
MGT133
D
MGT+EV
MGT+TBX20
Rel. mRNA Exp.
MYH6
MYH7
MYL2
E
HFF 1079SK
HFF 1112SK
MGT+EV
MGT+TBX20
DAPI
α-Actinin
cTnT
Merge
% Positive cells
F
HFF 1079SK
HFF 1112SK
ACTN2
ACTC1
MYBPC3
G
MGT+EV_1
MGT+EV_2
MGT+EV_3
MGT+TB_1
MGT+TB_2
MGT+TB_3
TBX20
MYH7
MYBPC3
GAPDH
H
Enrichment Score
heart contraction
muscle contraction
sarcomere organization
p value
0.001
0
5e−04
Ranked List Metric
MGT+TBX20
MGT+EV
Rank in Ordered Dataset
I
Protein Expression Fold Changes
MYOM2
MYH6
MYH7
MYBPC3

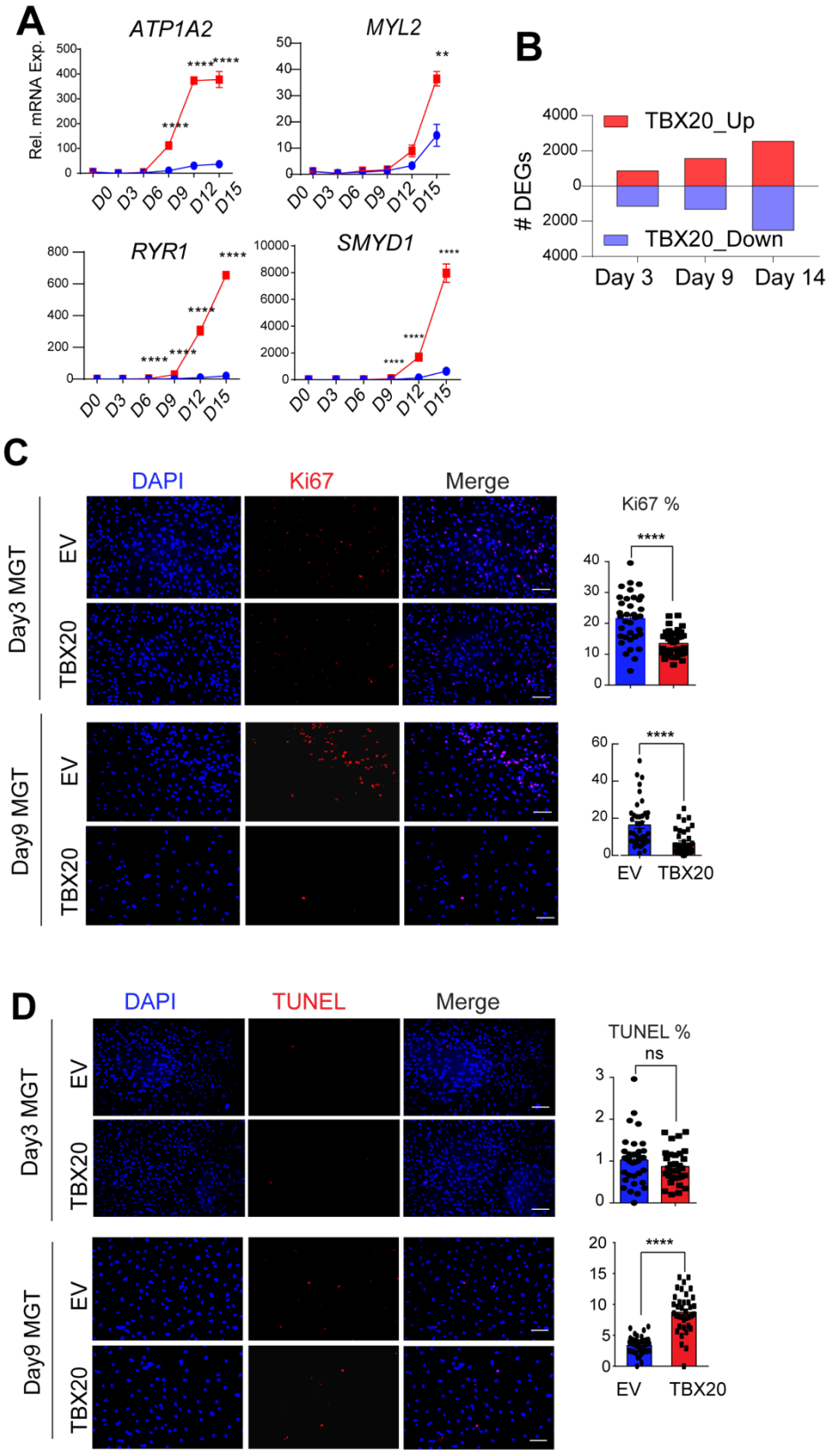
Supplemental Figure S1. TBX20 enhances human direct reprogramming and promotes sarcomere formation. A, Heatmap showing the differences in transcription factors and chromatin modulators among human direct cardiac reprogramming cells (H9F_5F12W and H9F_7F12W), endogenous cardiomyocytes (enCM), and stem cell-derived cardiomyocytes (H9CM). These genes are largely under-expressed in reprogrammed cells. **B,** Violin plots showing dynamics of *TBX20* and *HAND2* expression in every single cell along with indicated reprogramming days. **C,** Quantification of flow cytometry for αMHC+ and α-Actinin+ hiCMs transduced with TBX20-2A-EGFP or EGFP control at day 14. **D,** RT-qPCR evaluated the mRNA expression level of selected cardiac markers *MYH6*, *MYH7,* and *MYL2* (n=4 per group). **E,** IF staining images and quantification of cTnT+ or α-Actinin+ cells in hiCMs transduced with MGT+TBX20 or MGT+EV from HFF CCD-1079SK and HFF CCD-1112SK (n=20 per group). Scale bars, 100 μm. **F,** RT-qPCR results showing relative mRNA expression of cardiac markers as indicated in MGT+EV- or MGT+TBX20-induced hiCMs derived from HFF CCD-1079SK and HFF CCD-1112SK (n=4 per group). **G,** RNA-Seq signal mapping to the reference genome (hg19) of the genes *TBX20*, *MYH7* and *MYBPC3* for both samples in triplicate. Transcript per million (TPM) values are indicated on the y-axis. **H,** GSEA of DEGs using gene lists from GO terms associated with cardiac muscle contractility. **I,** Western Blotting quantification of MYOM2, MYH6, MYH7 and MYBPC3 in hiCMs transduced with MGT+EV or MGT+TBX20 from H9Fs (n=3 per group). GAPDH serves as a loading control. All data are expressed as mean ± SEM (**C-F, I** Student *t* test); *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

Figure S2


A
H9F
HCF
TBX20 overexpression
DAPI
TBX20
Merge
B
H9F RNA
H9F+MGT133 RNA
H9F+MGT133 H3K27me3
HCF+MGT133 RNA
TBX20
GAPDH
C
Rel. TBX20 Exp.
H9F
HCF
HFF 1079SK
HFF 1112SK
D
Rel. TBX20 Exp.
H9F ns
HCF ns
TBX20
H9F D0
H9F D14 MGT133
HCF D0
HCF D14 MGT133
E
MGT+EV
MGT+TBX20
cTnT
α-Actinin
Merge
F
TBX20 overexpression in HCF
MYBPC3
RYR1
Rel. mRNA Exp.
HCF
MG+EV
MGT+TBX20
G
TBX20 knockdown in HCF
MGT+shScr
MGT+shTBX20
α-Actinin
cTnT
Merge
H
HCF
Rel. TBX20 Exp.
H9F ns
shScr
shTBX20
MGT
I
H9F MGT+shScr
H9F MGT+shTBX20
Rel. mRNA Exp.
TNNT2
ACTN2
MYBPC3
J
H9F MGT133
Uninf.
shScr
shTBX20
0.083
14.3
18.6
cTnT
FSC
cTnT+ %
K
sgLuc
sgTBX20-1
sgTBX20-2
sgTBX20-3
55 kd
35 kd
TBX20
GAPDH
Fold change to sgLuc
L
Relative mRNA expression
ACTN2
TNNT2
ACTC1
sgLuc
sgTBX20-1
sgTBX20-3

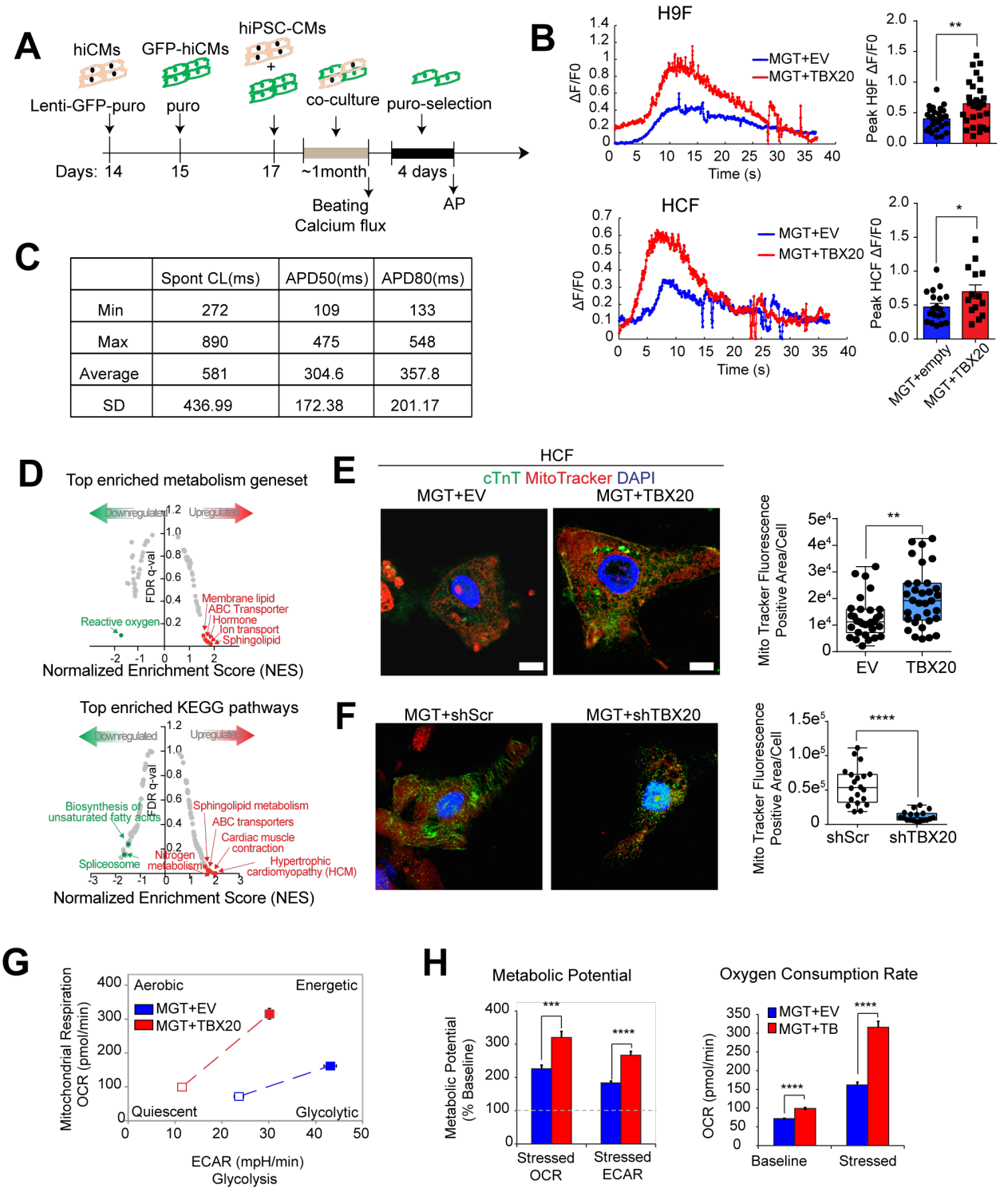
Supplemental Figure S2. TBX20 regulates hiCM reprogramming from human cardiac fibroblasts. A, IF staining for TBX20 expression in H9Fs, HCFs, and TBX20-overexpressed H9Fs. DAPI was used to label nuclei. Scale bar, 100 μm. **B**, IGV tracks showing low RNA-seq reads of *TBX20* in H9Fs and reprogramming day 14 cells, and high H3K27me3 CUT&RUN signals at *TBX20* locus in MGT133-transduced H9Fs. **C**, Relative *TBX20* mRNA expression in different types of fibroblasts as indicated (n=4 per group). **D**, RT-qPCR results showing relative expression of *TBX20* in H9F and HCF before and after reprogramming (n=4 per group). **E**, Double staining for cTnT and α-Actinin expression in HCF-derived hiCMs induced as indicated. Scale bars, 100 μm. **F**, RT-qPCR analysis of mRNA expression of *MYBPC3* and *RYR1* in HCF fibroblasts or in hiCMs transduced with MGT+EV or MGT+TBX20 from HCF (n=4 per group). **G,** Double staining for cTnT and α-Actinin expression in HCF-derived hiCMs coinfected with control shRNA (shScr) or shTBX20. Scale bars, 100 μm. **H**, RT-qPCR analysis showing knockdown efficiency using oligos targeting TBX20 (shTBX20) in HCF but not in H9F (n=4 per group). **I**, Relative changes of *TNNT2*, *ACTN2*, and *MYBPC3* mRNA expression via RT-qPCR assays in H9Fs transduced with MGT+shScr or MGT+shTBX20 (n = 4 per group). **J,** Flow plots and quantification of cTnT expression in H9Fs 14 days after MGT transduction with or without TBX20 knockdown (n=3 per group). **K,** Western blotting images and quantification showing CRISPR KO of TBX20 in HCFs. **L**, RT-qPCR analysis of cardiac gene expression after CRISPR KO of TBX20. All data are expressed as mean ± SEM (**F, K,** and **L**, one-way ANOVA with appropriate ad hoc test; **D** and **H-J**, Student *t* test;); *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001, ns, not significant.

Figure S3


A
ATP1A2
MYL2
RYR1
SMYD1
Rel. mRNA Exp.
D0 D3 D6 D9 D12 D15
B
TBX20_Up
TBX20_Down
DEGs
Day 3
Day 9
Day 14
C
DAPI
Ki67
Merge
Day3 MGT
Day9 MGT
EV
TBX20
Ki67 %
D
TUNEL
TUNEL %
ns

Supplemental Figure S3. TBX20 represses proliferation but not apoptosis during reprogramming. **A,** Time course RT-qPCR to evaluate relative mRNA expression of selected genes, *ATP1A2*, *MYL2, RYR1,* and *SMYD1* in H9F-derived hiCMs induced with MGT+EV or MGT+TBX20 on reprogramming day 0, 3, 6, 9, 12, and 15 (n=4 per group). **B,** Number of TBX20 differentially expressed genes (DEGs) identified at different time points (day 3, 9, and 14) by RNA-seq analysis during reprogramming from H9Fs. **C,** IF staining images and quantification of proliferating Ki67+ and **D,** TUNEL+ cells in H9Fs transduced with MGT+EV or MGT+TBX20 at reprogramming day 3 and 9 (n=20 per group). All data are expressed as mean ± SEM (**A,** Two-way ANOVA**; C** and **D,** Student *t* test;); *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001, ns, not significant.

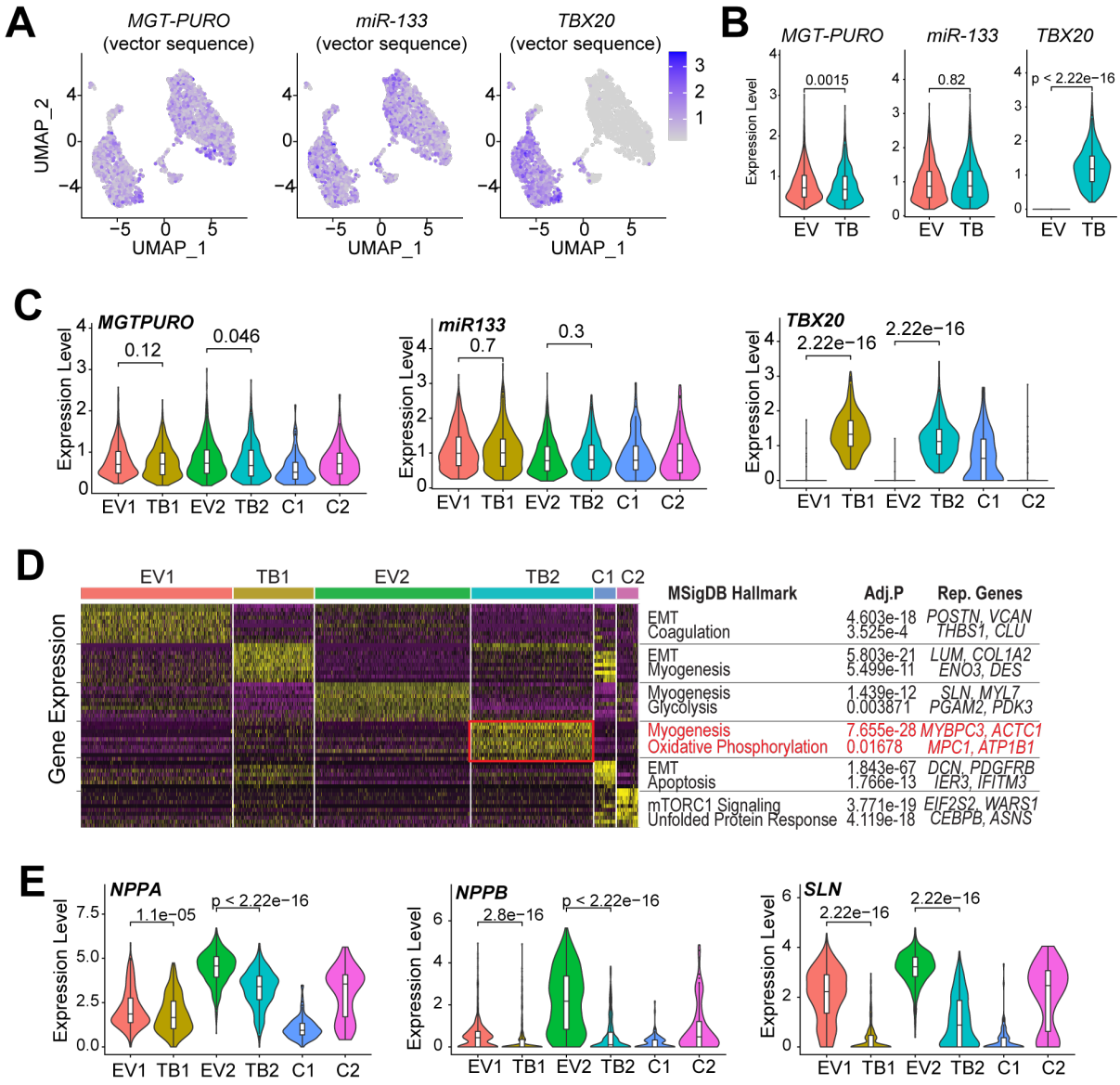
Figure S4



Supplemental Figure S4. TBX20 promotes contractility and metabolic switch in hiCMs. A,

Schematics of hiCM co-culture experiments with beating iPSC-CMs and following functional assessment. **B**, Calcium dye tracing and quantification of peak ΔF/F0 in MGT+EV- and MGT+TBX20-induced hiCMs derived from H9Fs or HCFs as indicated (n = at least 15). **C**, Table of optical mapping data measured in MGT+TBX20-induced hiCMs. CL, cycling length; APD, action potential duration (n = 3). **D**, GSEA analysis showing the metabolism and KEGG pathways gene sets enriched in TBX20 upregulated or downregulated genes. **E**, Left, representative images of mitochondria labeled using MitoTracker in cTnT+ hiCMs transduced with TBX20 or EV from HCFs. Scale Bar, 10μm. Right, quantification of total MitoTracker fluorescence positive pixel area per cell (n=20 per group). **F,** Left, representative images of mitochondria labeled using MitoTracker in cTnT+ hiCMs upon TBX20 knockdown (shTBX20) or control treatment (shScr) from HCFs. Scale Bar, 10μm. Right**,** quantification of total MitoTracker fluorescence positive pixel area per cell (n=20 per group). **G**, Metabolic phenotype plot summarizing ECAR and OCR measured in MGT+EV- and MGT+TBX20-induced hiCMs. **H**, Quantification of metabolic potential and oxygen consumption rate in hiCMs transduced with MGT+TBX20 or MGT+EV from H9Fs (n=10 with three biological replicates). All data are expressed as mean ± SEM (**B, E, F,** and **H,** Student *t* test;); *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

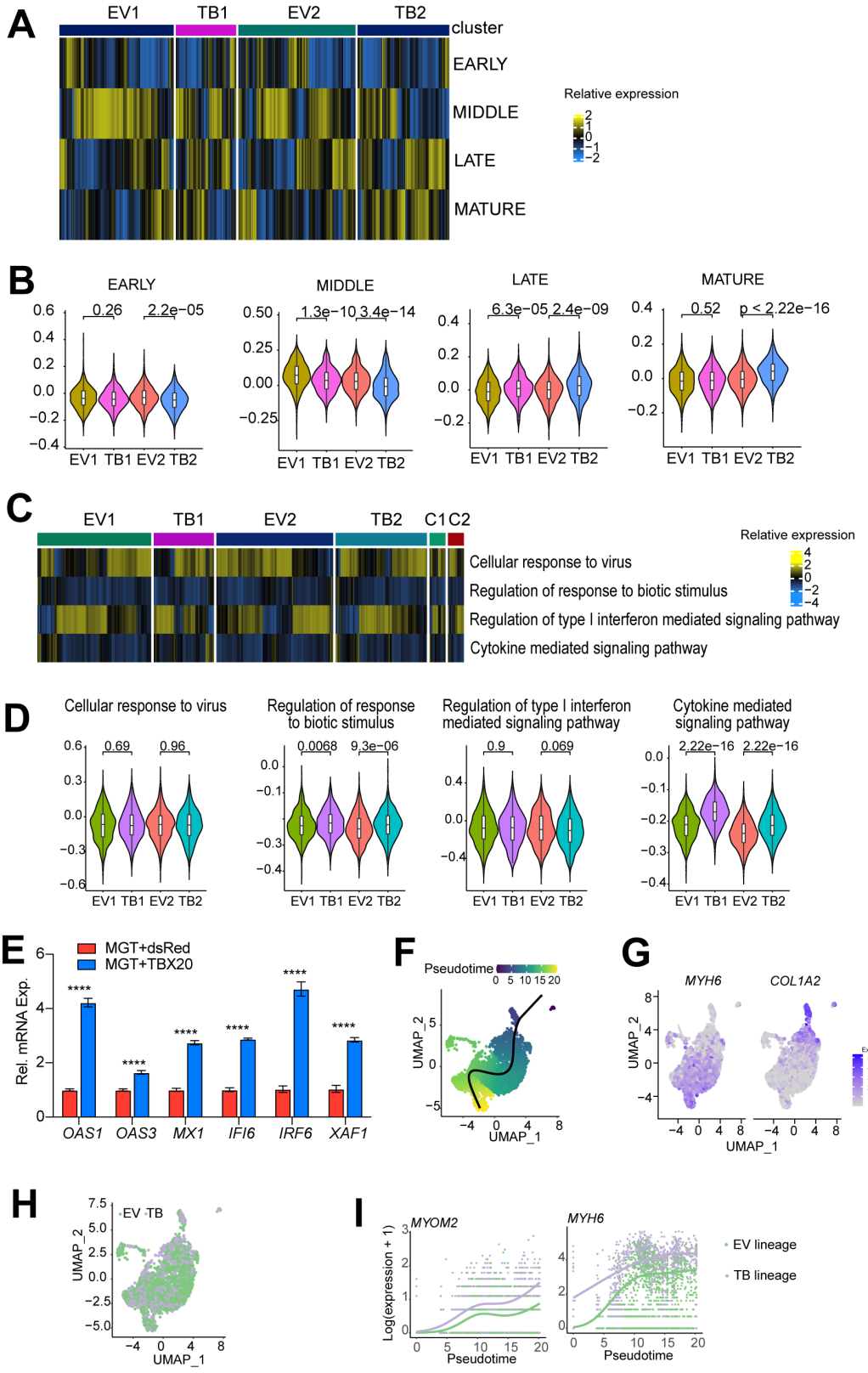
Figure S5



Supplemental Figure S5. Unbiased clustering of scRNA-seq data from hiCMs generated with or without TBX20. A, Feature plots showing expression levels of reprogramming factors (*MGT-PURO*, *miR-133,* and *TBX20*) on UMAP. **B**, Violin plots showing expression of reprogramming factors (*MGT-PURO*, *miR-133,* and *TBX20*) in EV and TB populations. **C**, Violin plots showing expression of reprogramming factors (*MGT-PURO*, *miR-133,* and TBX20) in all clusters (EV1, EV2, TB1, TB2, C1, and C2). **D,** Heatmap displaying the top enriched genes (n=10)

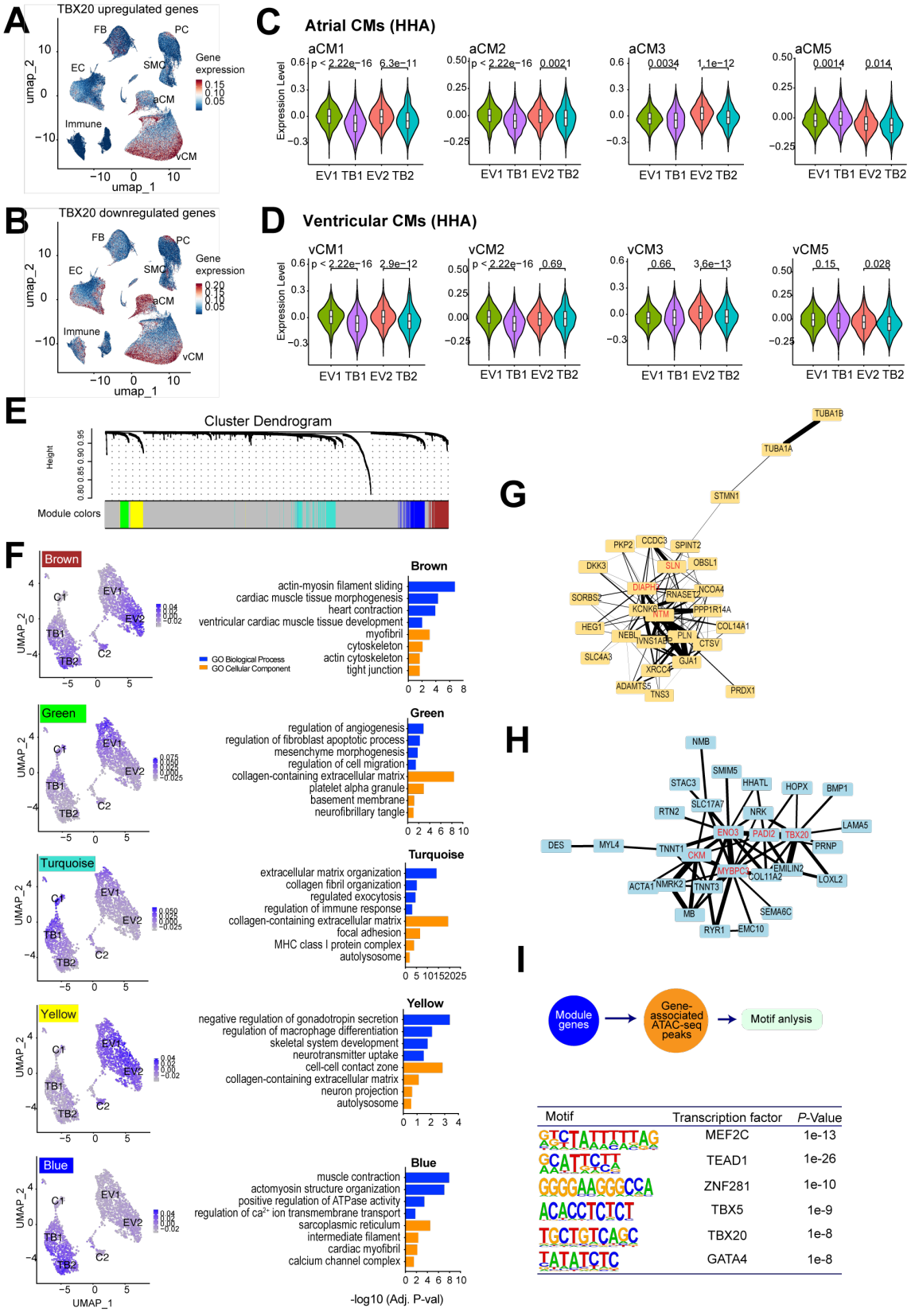
in each sub-cluster and their corresponding top molecular signature hallmark terms, adjusted p-value, and representative genes. **E**, Violin plots showing expression of selected atrial marker genes (*NPPA*, *NPPB,* and *SLN*) in all clusters (EV1, EV2, TB1, TB2, C1, and C2). **B, C, E**, P values were calculated by a two-sided Wilcoxon rank-sum test. $P < 2.2 \times 10{-}16$ represents a P value approaching 0. Box, median ± interquartile range. Whiskers, 1.5 × interquartile range.

Figure S6


A
EV1
TB1
EV2
TB2
cluster
EARLY
MIDDLE
LATE
MATURE
Relative expression
B
EARLY
MIDDLE
LATE
MATURE
0.26
2.2e−05
1.3e−10
3.4e−14
6.3e−05
2.4e−09
0.52
p < 2.22e−16
EV1 TB1 EV2 TB2
C
EV1
TB1
EV2
TB2
C1
C2
Cellular response to virus
Regulation of response to biotic stimulus
Regulation of type I interferon mediated signaling pathway
Cytokine mediated signaling pathway
Relative expression
D
Cellular response to virus
Regulation of response to biotic stimulus
Regulation of type I interferon mediated signaling pathway
Cytokine mediated signaling pathway
0.69
0.96
0.0068
9.3e−06
0.9
0.069
2.22e−16
2.22e−16
E
MGT+dsRed
MGT+TBX20
Rel. mRNA Exp.
OAS1
OAS3
MX1
IFI6
IRF6
XAF1
F
Pseudotime
UMAP_1
UMAP_2
G
MYH6
COL1A2
H
EV TB
UMAP_1
UMAP_2
I
MYOM2
MYH6
Log(expression + 1)
Pseudotime
EV lineage
TB lineage

Supplemental Figure S6. Gene set enrichment analysis and trajectory reconstitution of scRNA-seq data from MGT+EV and MGT+TBX20 hiCMs. A, Heatmap and **B,** violin plots displaying GSVA enrichment scores of indicated cardiomyocyte maturation-related gene (early, middle, late and mature) sets[20] in each sub-cluster. **C**, Heatmap and **D,** violin plots displaying GSVA enrichment scores of indicated immune response-related gene sets in each sub-cluster. **E**, RT-qPCR results showing mRNA expression changes of immune genes in MGT+dsRed or MGT+TBX20-transduced H9Fs (n=3 per group). All data are expressed as mean ± SEM. Student *t* test; ****P < 0.0001. **F**, Trajectory analysis showing the reprogramming route with pseudotime. **G,** Feature plots showing selected marker gene expression (*MYH6* and *COL1A2*) plotted on the trajectory UMAP obtained through trajectory analysis. **H**, UMAP of pseudotime analysis labeled with EV or TB clusters. **I**, Pseudotime analysis showing cardiac marker genes (*MYOM2* and *MYH6*) activation by TBX20. For **B** and **D**, P values were calculated by a two-sided Wilcoxon rank-sum test. $P < 2.2 \times 10{-}16$ represents a P value approaching 0. Box, median ± interquartile range. Whiskers, 1.5 × interquartile range.

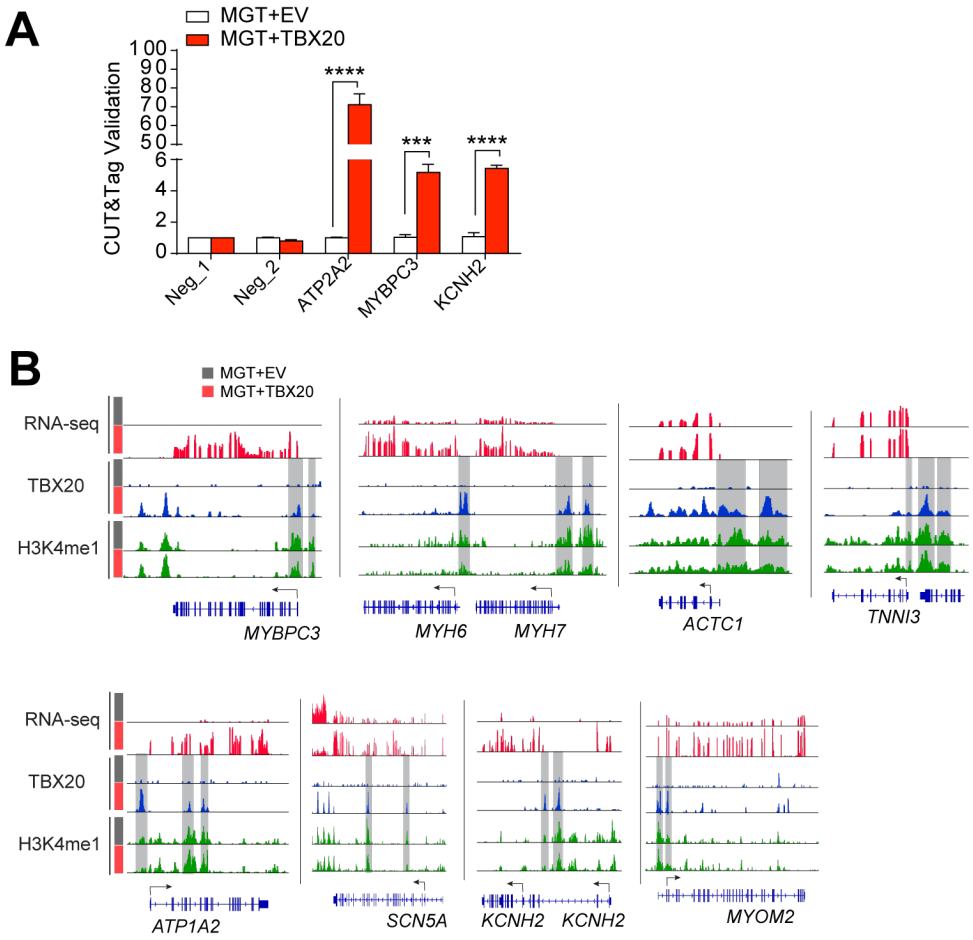
Figure S7



Motif	Transcription factor	*P*-Value
	MEF2C	1e-13
	TEAD1	1e-26
	ZNF281	1e-10
	TBX5	1e-9
	TBX20	1e-8
	GATA4	1e-8

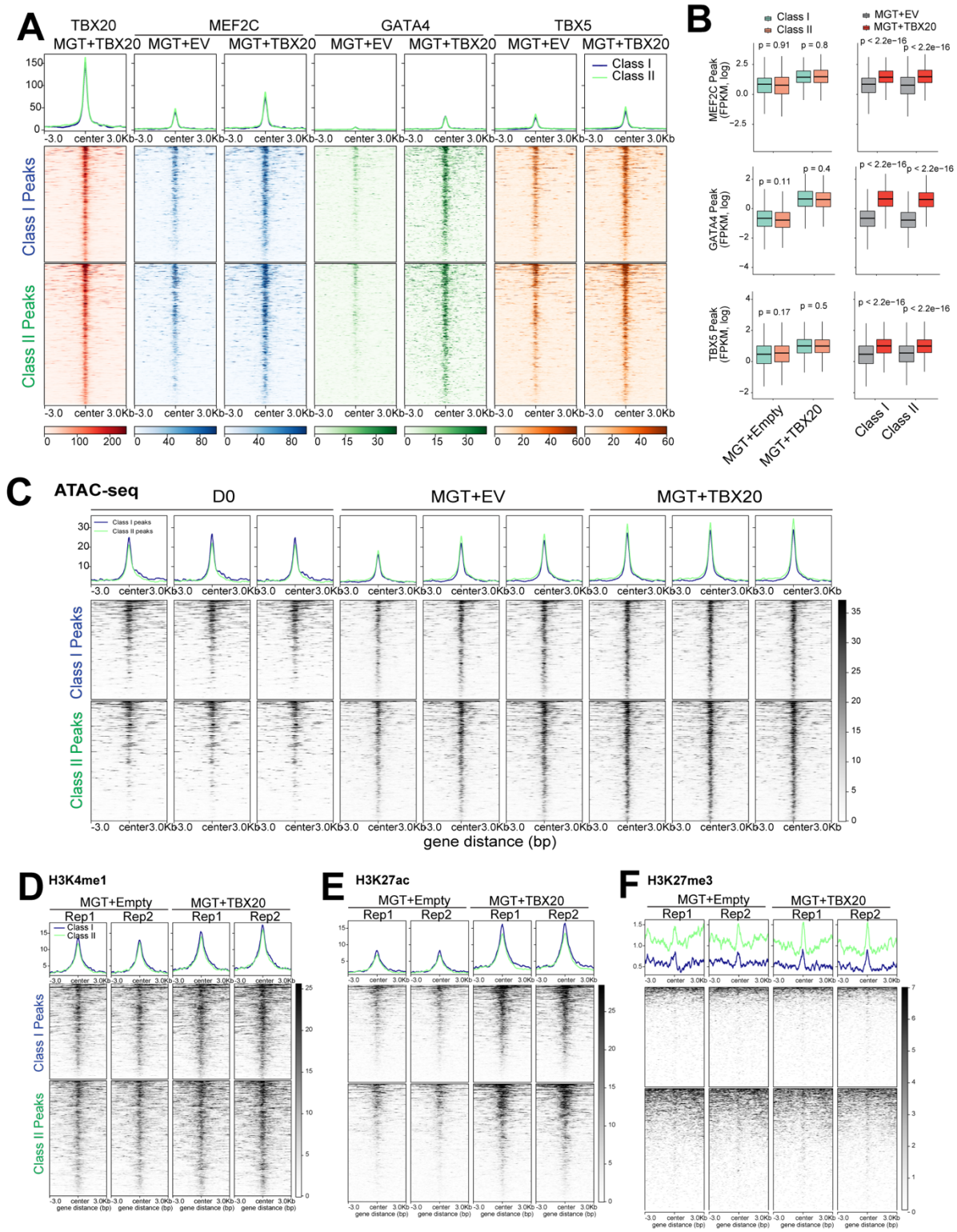
Supplemental Figure S7. Network analyses of scRNA-seq data from MGT+EV and MGT+TBX20 hiCMs. A, Feature plots of human heart atlas showing module expression of TBX20-upregulated targets and **B,** TBX20-downregulated targets. **C,** Violin plots showing expression of human heart cell atlas[21] atrial and **D**, ventricular cardiomyocytes subpopulation markers in hiCM clusters (EV1, EV2, TB1, and TB2). aCM, atrial cardiomyocytes. vCM, ventricular cardiomyocytes. **E**, WGCNA dendrogram indicating expression of five gene modules identified in scRNA-seq dataset. **F**, Left panels: feature plots showing average gene expression of each module. Right panels: selected GO terms significantly enriched in each module. **G**, Cytosape visualization for network analysis of TBX20-repressive yellow module. **H**, Cytosape visualization for network analysis of blue TBX20-responsive module. The stroke size of each line indicates connection strength. Red highlights the genes known for cardiomyocyte function. **I,** *De novo* motifs identified in cis-regulatory regions associated with blue module genes. **C**, **D**, *P* values were calculated by a two-sided Wilcoxon rank-sum test. $P < 2.2 \times 10{-}16$ represents a *P* value approaching 0. Box, median ± interquartile range. Whiskers, 1.5 × interquartile range.

Figure S8



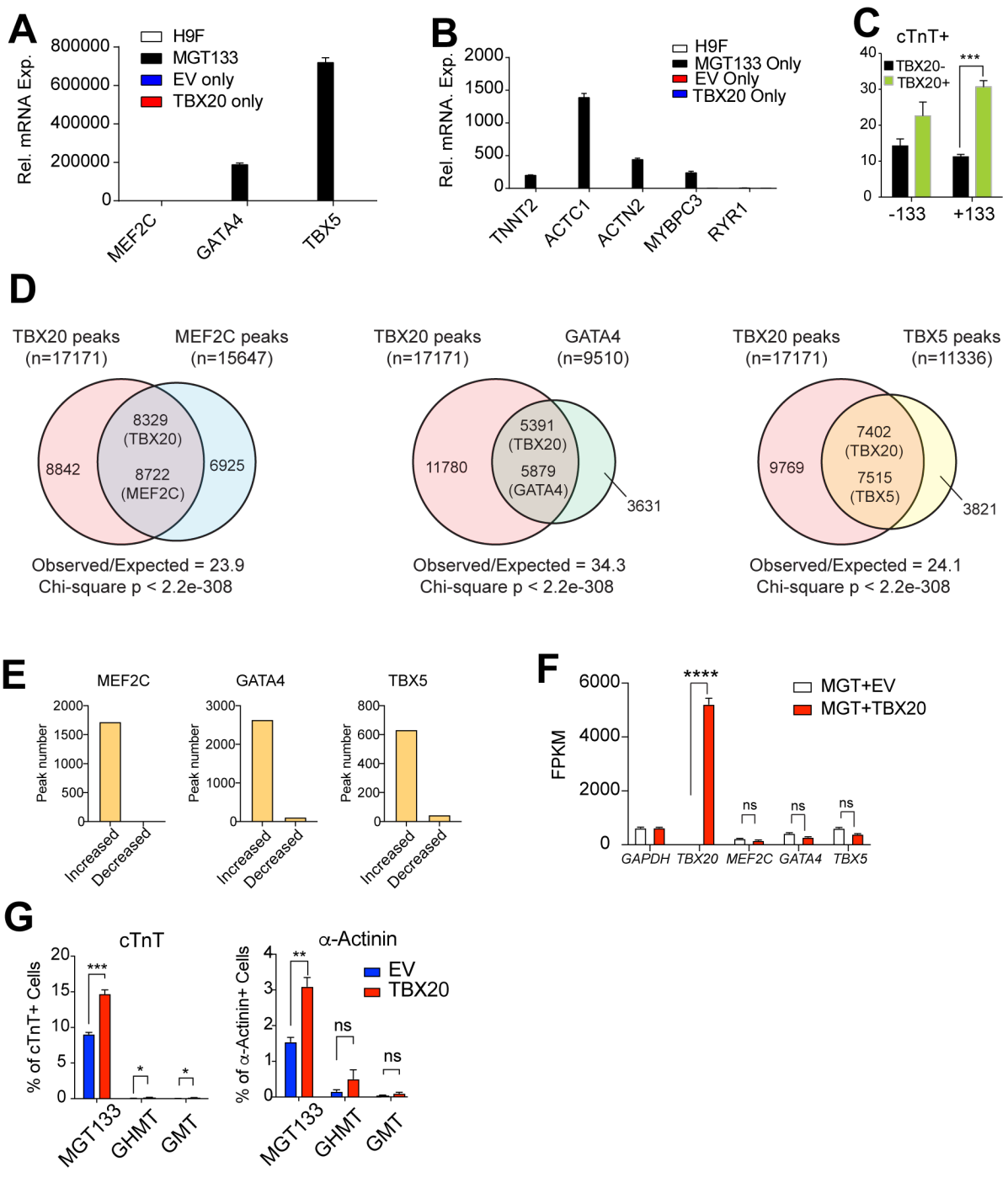
Supplemental Figure S8. TBX20 directly binds to the enhancer regions of cardiac function genes. A, QPCR validation for the selected gene loci in CUT&Tag samples (n=4 per group). All data are expressed as mean ± SEM. Student *t* test; ***P < 0.001, ****P < 0.0001. **B**, IGV browser tracks showing RNA-Seq signals, TBX20 CUT&Tag peaks, H3K27me1 CUT&RUN peaks associated with the selected cardiac genes upregulated in MGT+TBX20 hiCMs.

Figure S9


A
TBX20
MEF2C
GATA4
TBX5
MGT+TBX20
MGT+EV
Class I
Class II
Class I Peaks
Class II Peaks
B
Class I
Class II
MGT+EV
MGT+TBX20
MEF2C Peak (FPKM, log)
GATA4 Peak (FPKM, log)
TBX5 Peak (FPKM, log)
MGT+Empty
MGT+TBX20
C
ATAC-seq
D0
MGT+EV
MGT+TBX20
gene distance (bp)
D
H3K4me1
E
H3K27ac
F
H3K27me3
Rep1
Rep2

Supplemental Figure S9. Epigenetic profiling of two classes of upregulated TBX20 downstream targets. A, Heatmaps displaying the MEF2C, GATA4, TBX5 binding on Class I and Class II TBX20 peaks in H9F-derived hiCMs induced with MGT+EV or MGT+TBX20. **B**, Boxplots of peak intensities of class I and class II gene peaks (fragments per kb of peaks per million reads mapped, FPKM) at MEF2C, GATA4, TBX5 binding sites. P values were calculated by a two-sided Wilcoxon rank-sum test. $P < 2.2 \times 10{-}16$ represents a P value approaching 0. Box, median ± interquartile range. Whiskers, 1.5 × interquartile range. **C**, Density heatmaps of ATAC-seq signals of TBX20-upregulated Class I and Class II genes in D0 H9Fs, MGT+EV hiCMs, and MGT+TBX20 hiCMs. Three biological replicates were applied for ATAC-seq analysis. **D-F**, Heatmaps displaying the intensities of H3K4me1 (**D**), H3K27ac (**E**), and H3K27me3 (**F**) CUT&RUN at TBX20 binding peaks of class I and class II genes in MGT+EV- and MGT+TBX20-induced hiCMs.

Figure S10


A
Rel. mRNA Exp.
800000
600000
400000
200000
0
H9F
MGT133
EV only
TBX20 only
MEF2C
GATA4
TBX5
B
Rel. mRNA. Exp.
2000
1500
1000
500
0
H9F
MGT133 Only
EV Only
TBX20 Only
TNNT2
ACTC1
ACTN2
MYBPC3
RYR1
C
cTnT+
40
30
20
10
0
TBX20-
TBX20+
***
-133
+133
D
TBX20 peaks (n=17171)
MEF2C peaks (n=15647)
8842
8329 (TBX20)
8722 (MEF2C)
6925
Observed/Expected = 23.9
Chi-square p < 2.2e-308
TBX20 peaks (n=17171)
GATA4 (n=9510)
11780
5391 (TBX20)
5879 (GATA4)
3631
Observed/Expected = 34.3
Chi-square p < 2.2e-308
TBX20 peaks (n=17171)
TBX5 peaks (n=11336)
9769
7402 (TBX20)
7515 (TBX5)
3821
Observed/Expected = 24.1
Chi-square p < 2.2e-308
E
MEF2C
GATA4
TBX5
Peak number
2000
1500
1000
500
0
3000
2000
1000
800
600
400
200
Increased
Decreased
F
FPKM
6000
4000
2000
0
****
ns
MGT+EV
MGT+TBX20
GAPDH
TBX20
MEF2C
GATA4
TBX5
G
cTnT
% of cTnT+ Cells
20
15
10
5
0
***
*
MGT133
GHMT
GMT
α-Actinin
% of α-Actinin+ Cells
4
3
2
1
0
**
ns
EV
TBX20

Supplemental Figure S10. The synergy of TBX20 and core reprogramming TFs. A-B, RT-qPCR evaluation of reprogramming TFs (**A**) and cardiac marker genes (**B**) as indicated in H9Fs, H9Fs transduced with MGT+miR-133, H9Fs transduced with empty vector control (EV), or H9Fs transduced with TBX20 only. **C,** Flow cytometry analyses for expression of cTnT in TBX20-GFP- and TBX20-GFP+ hiCMs 14 days post-transduction of MGT+TBX20-EGFP with or without miR-133. **D**, Venn diagrams depicting the highly significant overlap between TBX20 and MEF2C, GATA4, or TBX5 peaks. P values were calculated by Pearson's chi-square test. **E**, Number of MGT peaks that showed increased or decreased changes after TBX20 overexpression. **F**, RNA-seq gene expression of reprogramming factors, *MEF2C*, *GATA4* and *TBX5* showing no difference after TBX20 overexpression. FPKM, fragments per kilobase million. **G,** Flow cytometry analyses for cTnT and α-Actinin expression in HFFs after 14 days of reprogramming using MGT133, GMT and GHMT with or without TBX20. All data are expressed as mean ± SEM (**C, F, G**, Student *t* test); *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001, ns, not significant.

Supplemental Video S1-S3

Video S1. Beating TBX20-induced hiCMs observed one month after co-culture. hiCMs transduced by TBX20 developed spontaneously-like beating after co-culture with iPSC-CMs for one month. TBX20-induced hiCMs were labeled with GFP indicator prior to co-culture. Video were played by 4X speed.

Video S2. Calcium flux recording of EV-induced hiCMs one month after co-culture.
EV-induced hiCMs were labeled with GFP indicator prior to co-culture. After one month co-culture with iPSC-CMs, calcium flux was recorded as shown.

Video S3. Calcium flux recording of TBX20-induced hiCMs 1M after co-culture.
TBX20-induced hiCMs were labeled with GFP indicator prior to co-culture. After one month co-culture with iPSC-CMs, calcium flux was recorded as shown.

Supplemental Table S1-S3

Table S1. Identified transcription factors and epigenetic regulators underexpressed in hiCMs.

Table S2. Genomic coordinates for enhancers cloned into pGL3-Promoter vector.

Table S3. Lists of RT-qPCR, CUT&Tag-qPCR, and mitochondrial DNA qPCR primers.