A TBX5 dosage-sensitive gene regulatory network for human congenital heart disease

Irfan S. Kathiriya[1,2,3]*, Kavitha S. Rao[1,2,3], Giovanni Iacono[4], W. Patrick Devine[1,5], Swetansu K. Hota[1,2], Michael H. Lai[1,2], Bayardo I. Garay[1,2,3], Reuben Thomas[1], Andrew P. Blair[1,8], Henry Z. Gong[8], Lauren K. Wasson[6,7], Piyush Goyal[1,2], Tatyana Sukonnik[1,2], Gunes A. Akgun[1,2], Laure D. Bernard[1,2], Brynn N. Akerberg[9], Fei Gu[9], Kai Li[9], William T. Pu[9,10], Joshua M. Stuart[8], Christine E. Seidman[6,7], J. G. Seidman[6], Holger Heyn[4,11], Benoit G. Bruneau[1,2,12,13]*

1. Gladstone Institutes, San Francisco, CA, USA.

2. Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA, USA

3. Department of Anesthesia and Perioperative Care, University of California, San Francisco, CA, USA

4. CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain;

5. Department of Pathology, University of California, San Francisco, CA, USA

6. Department of Genetics, Harvard Medical School, Boston, MA USA

7. Howard Hughes Medical Institute, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA USA

8. Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, CA, USA

9. Department of Cardiology, Boston Children's Hospital, Boston, MA, USA

10. Harvard Stem Cell Institute, Harvard University, Cambridge, MA, USA

11. Universitat Pompeu Fabra, Barcelona, Spain

12. Department of Pediatrics, University of California, San Francisco, CA, USA

13. Cardiovascular Research Institute, University of California, San Francisco, CA, USA

* Authors for correspondence: irfan.kathiriya@ucsf.edu, benoit.bruneau@gladstone.ucsf.edu

Abstract

Haploinsufficiency of transcriptional regulators causes human congenital heart disease (CHD) [1]. This observation predicts gene regulatory network (GRN) imbalances [2], but the nature of dosage-vulnerable GRNs and their contribution to human cardiogenesis and CHDs are unknown. Here, we define transcriptional consequences of reduced dosage of the CHD transcription factor TBX5 during human cardiac differentiation from induced pluripotent stem (iPS) cells. Single cell RNAseq revealed that transcriptional responses to reduced TBX5 levels are not homogeneous, and instead, discrete sub-populations of cardiomyocytes exhibit dysregulation of distinct TBX5 dose-sensitive genes related to cellular phenotypes and CHD-associated genetics. Cellular trajectory inference revealed TBX5 dosage-dependent differentiation paths, with implications for cardiac developmental identity. GRN analysis of the single cell RNAseq data identified vulnerable nodes enriched for CHD genes, implicating a critical sensitivity to TBX5 dosage in cardiac network stability. A novel GRN-predicted genetic interaction between *TBX5* and *MEF2C* was validated in mouse, revealing a highly dosage-sensitive pathway for CHD. Our results reveal unforeseen complexity and exquisite sensitivity to TBX5 dosage in discrete sub-populations of iPSC-derived cardiomyocytes, providing mechanistic insights into human CHDs and quantitative transcriptional regulation in disease.

Introduction

CHDs are a leading cause of childhood morbidity and mortality, and incidence of CHD is estimated to be ten-fold higher in human fetuses than live births [3,4]. Many human mutations

linked to CHD predict reduction in dosage of transcriptional regulators, including transcription
52 factors (TFs) and chromatin-modifying genes [1]. Despite advances in tracing the roles of
individual factors, how altered dosage of transcriptional regulators translates to altered
54 transcriptional activity is not known, nor is it known how these altered GRNs disrupt heart
development, resulting in CHD.

56 Heterozygous mutations in the T-box TF gene *TBX5* cause Holt-Oram syndrome (HOS)
[5,6], which presents with upper limb defects, CHDs, diastolic dysfunction, or arrhythmias. In
58 humans, homozygous *TBX5* loss of function is presumed to cause fetal demise. Altered
expression of select genes in mice *in vivo* demonstrate a stepwise sensitivity to reductions in
60 *Tbx5* dosage [7,8]. These findings suggest that a reduction in TBX5 dosage perturbs downstream
gene expression, but the overall disrupted networks and mechanisms are not understood.

62 Mechanistic investigation of human CHD has been hampered by a lack of relevant and
tractable models. Human heart tissue is largely inaccessible for molecular analysis, and
64 pathological or surgical specimens are limited. However, induced pluripotent stem (iPS) cells,
genome editing, directed differentiation, and single cell RNAseq together provide a human
66 cellular platform for gene-centered cardiac disease modeling at single cell resolution. In
considering how transcriptional regulator haploinsufficiency might cause CHD, at least two
68 scenarios are possible: 1. That reduced dosage affects genes only in specific anatomical
locations, such as the atrial septum. 2. That reduced dosage affects cardiac gene expression
70 broadly, but altered programs manifest as morphologic defects only in structures most sensitive
to the disturbance. The first scenario would make it challenging to discover altered gene
72 expression in 2-dimensional cultures if susceptible cell types are absent. The second may
predict that homogeneous changes in gene expression might be detected in that context, as
74 corroborated by bulk RNAseq studies of heterozygous human iPS cell models of CHD [9-11].
However, it is likely that relevant, discrete alterations in a complex cell mixture were not

76 detected in those studies. Therefore, how reduced dosage of a transcriptional regulator alters GRNs in susceptible cells leading to CHD remains unclear.

78

Impaired human cardiac differentiation and function by reduced TBX5 dosage

80 To determine a role for TBX5 dosage in human cardiac biology, we created an allelic series of human iPS cells mutant for *TBX5* by using CRISPR/Cas9-mediated genome editing, to 82 target exon 3 of *TBX5* at the start of the essential T-box domain (Fig. 1a). We isolated targeted iPS cell lines: control (*TBX5*$^{+/+}$), heterozygous (*TBX5*$^{in/+}$) and homozygous (*TBX5*$^{in/del}$ and 84 *TBX5*$^{PuR/PuR}$) mutants (Fig. 1b, Extended Data 1). TBX5 protein levels in cardiomyocytes differentiated from these lines were diminished in *TBX5*$^{in/+}$ and absent in *TBX5*$^{in/del}$ and 86 *TBX5*$^{PuR/PuR}$ cells (Fig. 1c), consistent with a dosage-step allelic series of mutant *TBX5* loss-of-function cell lines.

88 We observed reduced cardiomyocyte differentiation efficiency, a delay in onset of beating, and worsening sarcomere disarray, by graded loss of *TBX5* (Fig. 1d-j, Extended Data 90 2). Patch clamp analysis revealed lengthened action potentials (AP) in *TBX5*$^{in/del}$ cells (Fig. 1k, l; APD90, adj p-value<0.04 by Holm-Sidak test[12]), consistent with previous findings [13-15]. Fluo-4 92 imaging measurements uncovered protracted calcium transient duration in *TBX5*$^{in/del}$ and *TBX5*$^{PuR/PuR}$ cells (T90 down, adj p-value<9E-4 by Holm-Sidak test), with an intermediate defect 94 in *TBX5*$^{in/+}$ cells (adj p-value<0.01) (Fig. 1m, n), consistent with findings in murine postnatal atrial CMs [16]. These findings may underlie diastolic dysfunction in HOS in mice and humans [17].

96

Resolving TBX5 dosage-sensitive transcriptomes during human cardiac differentiation

98 To determine if reduced TBX5 dosage alters gene expression in potentially vulnerable cell types, we used a droplet-based single cell RNAseq method with cells collected from

100 parental WTC11, control *TBX5*$^{+/+}$, and mutant *TBX5* (*TBX5*$^{in/+}$, *TBX5*$^{in/del}$) genotypes during directed cardiac differentiation (Fig. 2a). Heterogeneous cell populations composed of cardiac
102 precursors (CP) at day (d) 6, and cardiomyocytes (CMs) at d11 or d23, were harvested. We interrogated 64,396 cells with an average normalized read depth of 88,234 reads per cell (Fig.
104 2b-d; Extended Data 3). At d6, we identified 11 cell clusters, representing at least four cell types, including *POU5F1*$^{+}$ pluripotent cells, *MESP1*$^{+}$ mesoderm, *ISL1*$^{+}$ cardiac precursors and
106 nascent *TNNT2*$^{+}$ cardiomyocytes (Extended Data 3). At d11 and d23, differentiated cell types were identified by specific marker genes (Extended Data 3). This included a diversity of *TBX5*$^{+}$
108 cell types, comprising very few *PLVAP*$^{+}$ endothelial cells or *TTR*$^{+}$ endodermal cells, some *COL1A1*$^{+}$ fibroblasts and, most abundantly, *TNNT2*$^{+}$ cardiomyocytes (Extended Data 3, 4). All of
110 these major cell types were present and detected in each *TBX5* genotype at d11 or d23 (Extended Data 3).

112 We applied this classification of human cell types to learn more about the genes that contribute to CHDs. We examined the expression of 376 CHD-associated candidate genes,
114 compiled from [18-25], across the major cell types regardless of *TBX5* genotypes at d23 (Extended Data 4, Supplemental Table 1). For many CHD-associated genes, we observed enrichment of
116 gene expression in CMs (e.g.*TBX5, FOXH1*, *ARID1A*, *EP300*, *KDM6A*, *PKD1*; adj p-value<0.05 by Wilcoxon Rank Sum test), fibroblasts (e.g. *SMARCA2, PITX2, ZEB2, HAND2*; adj p-
118 value<0.05), endoderm (e.g. *TBX3, WDR5, SALL4, NODAL*; adj p-value<0.05), endothelial cells (e.g. *FGFR1, TGFBR2, SHANK3, KCNJ2*; adj p-value<0.05) or other cell types (Extended Data
120 4). This gene expression dataset provides a resource that helps identify progenitors to these cell types and may allow further categorization of human CHD-associated candidate genes.

122

TBX5 dose-dependent cellular trajectories during human cardiac differentiation

To assess if altered transcriptional states in *TBX5* mutant cells reflected changes in paths of differentiation, we inferred cellular trajectories of each *TBX5* genotype in our allelic series using the R package URD [26], which predicts cellular chronology based on user-determined origins and ends. We defined *POU5F1*$^{+}$ cells from d6 as a root and each d23 cluster as an individual tip in a pseudotime tree (Fig. 2b-d, Extended Data 5). *TBX5*$^{in/+}$ cells followed a path similar to wildtype and control (collectively "WT) cells during differentiation to CMs (Fig. 2e, f, dashed lines), but within the shared branch, a tip occupied by *TBX5*$^{in/+}$ CMs (red arrowhead) was disparate from wildtype (black arrowhead) and control, demonstrating *TBX5*$^{in/+}$ CMs distinct from WT. In contrast, *TBX5*$^{in/del}$ cells deviated from the differentiation path (Fig. 2g, blue dashed line, blue arrowheads). We identified several genes (e.g. *NAV1, TECRL, LAMA4,* and *NPPA*), which were positively correlated with pseudotime (rho≥|0.4|, p-value<0.05 by two-sided *t* test) in the WT/*TBX5*$^{in/+}$ branch, but not in the *TBX5*$^{in/del}$ branch (Z-score≥15 by difference in rho), suggesting that these genes were not activated properly in *TBX5*$^{in/del}$ cells (Fig. 2h, i, Extended Data 5). Likewise, we identified several genes (e.g. *DES, TBX2, RSPO3,* and *LBH*) that positively correlated with pseudotime in *TBX5*$^{in/del}$ cells (rho≥|0.4|, p-value<0.05), but not in wildtype cells (Z-score≥15) (Fig. 2j, Extended Data 5), signifying that these genes are inappropriately deployed in *TBX5*$^{in/del}$ cells.

To identify genes that may determine one cell trajectory route from another, we assessed differential gene expression between imputed intermediates, which we considered as potential precursors, within branches proximal to *TBX5* genotype-specific tips. We compared these intermediate branches of cells that distinguish the cell trajectory route to CMs of wildtype, control and *TBX5*$^{in/+}$ from *TBX5*$^{in/del}$ (Fig. 2k, segments 69 vs. 70). We identified several differentially expressed genes, including structural genes (e.g. *MYL3*, *TTN*, *MYH6*, *TNNT2*, *TNNI1*; adj p-value<0.05 by Wilcoxon Rank Sum test), calcium handling genes (*PLN*, *ATP2A2, RYR2*; adj p-value<0.05), secreted factors (*WNT2*, *FGFR1*; adj p-value<0.05), and cardiac TFs (*IRX4*, *HAND2*; adj p-value<0.05) (Extended Data 5). Of note, expression of the CHD-

150 associated cardiac transcription factor *NKX2-5*, a transcriptional partner of TBX5 [7,27,28], was differentially expressed between intermediate branches of the URD tree that distinguish
152 wildtype/control/*TBX5*$^{in/+}$ *TNNT2*$^{+}$ cells from *TBX5*$^{in/del}$ *TNNT2*$^{+}$ cells (Fig. 2l; adj p-value<1E-300 by Wilcoxon Rank Sum test). Onset of *NKX2-5* expression was delayed in *TBX5*$^{in/del}$ cells (adj p-
154 value<0.05 by Bonferroni-Holm multiple testing correction), along with a module of genes (e.g. *PARP1, RPL37, KIAA1462* and *ATP1A1*; adj p-value<0.05), concomitant with *NKX2-5*
156 expression (Fig. 2l, Extended Data 5), providing a potential molecular explanation for a delay of onset of beating by *TBX5* loss (Fig. 1e).
158

Quantitative transcriptomic responsiveness to reduced TBX5 dosage in cardiomyocytes

160 *TBX5* genotype-specific clusters emerged among cardiomyocytes at d11 (Fig. 2c), and *TBX5* genotype-specific segregation was more striking at d23, particularly in *TNNT2*$^{+}$ cells (Fig.
162 2d). We focused on *TNNT2*$^{+}$ clusters at d23 (Fig. 3a, b) and d11 (Extended Data 6). We constructed a phylogenetic tree relating 25 different *TNNT2*$^{+}$ cell clusters (Fig. 3c), and
164 observed heterogeneity of gene set enrichment among WT or *TBX5*$^{in/+}$ clusters (Fig. 3d), implying putative CM subtypes. Some CM subtypes did not fit conventional categorization
166 based on anatomy. We searched for differentially expressed genes, by pair-wise comparisons of related subtypes between *TBX5* genotypes. To identify *TBX5* genotype-specific differences
168 within a CM subtype, we compared clusters enriched for WT to related clusters enriched for *TBX5*$^{in/+}$ or *TBX5*$^{in/del}$ cells (Fig. 3d-k). We discovered only a few common changes in gene
170 expression amongst pair-wise cluster comparisons of WT vs *TBX5*$^{in/+}$ clusters, such as *NPW* and *PLN* (Fig. 3d-i; adj p-value<0.05 by Wilcoxon Rank Sum test). We also discerned
172 differences in gene expression based on cluster-specific comparisons (adj p-value<0.05 by Wilcoxon Rank Sum test), implying discrete cell subtype-sensitivity to TBX5 haploinsufficiency
174 (Fig. 3d-i) or *TBX5* loss (Extended Data 7). We also compared related *TBX5* genotype-specific

cell clusters at a lower resolution of Louvain clustering and discovered genes that were
176 differentially expressed across reduced TBX5 dosage in a variety of patterns (Fig. 3j, k).
Quantitative changes in gene expression level were apparent across *TBX5* genotypes (for
178 example, *PLN, FYXD1*; adj p-value<0.05 by Wilcoxon Rank Sum test), as were changes in both
expression level and percentage of expressing cells (e.g. *RSPO3, NPPA, TECRL, DES*) (Fig.
180 3k).

This differentially expressed gene set was enriched for electrophysiology (EP)-related
182 genes (FDR<0.05, Fig. 3l), which are implicated in membrane depolarization (*SCN5A*), calcium
handling (*RYR2*, *ATP2A2*, and *PLN*) and arrhythmias (*TECRL*) (Fig. 3e, g, i, k, l). Several
184 altered transcripts were encoded by other genes implicated in CHD (*CITED2, MYOCD,*
ANKRD1) (Fig. 3e, g, i, k, l). Some TBX5-dependent genes that were previously associated with
186 CHD or arrhythmias by genome-wide association studies (GWAS) were identified (Fig. 3l). We
uncovered *IGFBP7, FRAS1, MYH7B* and *SMCHD1* for CHD (odds ratio 3.2, FDR<0.03) and 47
188 reported genes for arrhythmias (for example, *PLN, HCN4, SCN5A, GJA1*, *PITX2* and *TECRL*;
odds ratio 9.0, FDR<1.9E-39) among TBX5-sensitive genes at day 23 (Fig. 3m, Supplementary
190 Table 2).

Using orthogonal assays at single cell resolution, we found that *TBX5* dosage-
192 dependent downregulation of *NPPA* was evident in CMs by RNAscope ($TBX5^{in/+}$, p<0.05;
$TBX5^{in/del}$ or $TBX5^{PuR/PuR}$, p<1E-4 by Student's t-test), consistent with a TBX5-dependent
194 rheostatic regulation of this gene in mouse [7,8]. Loss of DES in $TBX5^{in/+}$ and recovery in $TBX5^{in/del}$
was detected in CMs by flow cytometry (Extended Data 8, p-value<1E-4 by Chi-Square test),
196 corroborating some human *TBX5* dose-dependent sequelae.

Comparisons of biological replicates at d23 showed significant overlap of TBX5-
198 dependent genes for WT vs. $TBX5^{in/+}$ (p-value<2.466e-9 by hypergeometric test) or WT vs.
$TBX5^{in/del}$ (p-value<5.410e-6) (Extended Data 9, 10). The differences in gene expression
200 between biological replicates likely reflect a combination of technical variability and biological

stochasticity that, as in patients with *TBX5* mutations, may explain variable expressivity of disease for a given mutation. Since modifiers in different genetic backgrounds might modulate phenotypic effects, we assessed iPSC lines of *TBX5* mutants in an independent genetic background (PGP1, from a Caucasian male, compared to WTC11 from a Japanese male, Extended Data 11). Comparisons of TBX5-dependent genes in d23 *TNNT2*$^{+}$ cells showed significant overlap between WTC11 vs. *TBX5*$^{in/+}$ and PGP1 vs. *TBX5*$^{in/+}$ (p-value<1.174e-8 by hypergeometric test) or WTC11 vs *TBX5*$^{in/del}$ and PGP1 vs. *TBX5*$^{del/del}$ (p-value<7.260e-15) (Extended Data 11, 12), proportionally similar to biological replicates (Extended Data 9, 10). These results demonstrate reproducible TBX5 dosage-dependent alterations from independent experiments, genetic backgrounds, and gene targeting strategies.

To assess if changes to TBX5 dose-sensitive genes were largely direct or indirect TBX5 targets, we examined TBX5 occupancy in human iPSC-derived CMs from a published dataset [10]. We found correlations of TBX5 occupancy near TBX5 dosage-vulnerable gene sets (Fig. 3l, m, and Extended Data 13, 14; for example, odds ratio 6.9 for all TBX5-dependent human genes; FDR<1.9E-178), suggesting our identified genes may be direct TBX5 targets. TBX5 cooperates with GATA4 for cardiac gene regulation [10,28,29]; we also observed that GATA4 occupancy [10] with TBX5 was highly associated near TBX5-dependent genes (Fig. 3l, m, and Extended Data 13, 14; for example, odds ratio 6.4 for all TBX5-dependent human genes; FDR<1.3E-155), indicating that GATA4 may have a role modulating TBX5 dosage-sensitive genes.

Susceptible human gene regulatory networks for congenital heart disease

Since CHD-associated and arrhythmia-related genes were enriched among TBX5-dependent genes, we deduced that TBX5 may be vital for preserving the integrity of cardiac gene regulatory networks (GRNs). To evaluate the role of TBX5 dosage for regulating GRNs,

226 we used bigSCale2 [30] to independently infer putative GRNs from single cell expression data of *TNNT2*$^{+}$ cells at d6, d11 and d23 of each *TBX5* genotype (Fig. 4a, b, Extended Data 15). This
228 allows for quantitative prediction of the biological importance of genes in a regulatory network reconstructed with bigSCale2, by applying the concept of "pagerank" [30], first devised to rank the
230 importance of websites via numerical weighting [31], to discover genes with a loss in importance (i.e. centrality) among *TBX5* genotypes. By comparing nodes from inferred networks of WT to
232 *TBX5*$^{in/+}$ or *TBX5*$^{in/del}$, we uncovered several nodes that displayed reduced pagerank centrality at specific time points and levels of reduced TBX5 dosage (Fig. 4a-c, Extended Data 15). These
234 included *RYR2* and twenty CHD-associated genes (for example, *GATA6*, *HAND2*, *SMAD2,* $p<2.2e-5$ by hypergeometric test,) (Fig. 4c, d, Extended Data 15), consistent with independent
236 analysis from differential gene expression. For example, at d11, pagerank centrality of the CHD-associated TF *SMAD2* was absent in *TBX5*$^{in/+}$ cells (Fig. 4a-c, top 5% of all changes), indicating
238 an impairment of *SMAD2* function. Centrality of the cardiac development-related TF *MEF2C*, which is necessary for heart development [32], was substantially reduced by heterozygosity or
240 loss of *TBX5* at d11 (Fig. 4a-c, top 5% cutoff). *MEF2C* gene expression was unchanged by reduced TBX5 dosage, suggesting a change in *MEF2C* functional connectivity. These altered
242 GRNs show that TBX5 dosage is critical for maintaining cardiac network stability, and unveil potential genetic interactions disrupted in TBX5-dependent CHDs.

244 To further investigate the predicted relationship between *TBX5* and *MEF2C* within a dosage-sensitive CHD-associated GRN, we used a complementary approach using bigSCale2,
246 to identify gene-gene correlations with *TBX5* expression in individual *TNNT2*$^{+}$ cells across timepoints and *TBX5* genotypes (Fig. 4e). Genes highly co-expressed with *TBX5* (Pearson
248 coefficient >0.5) regardless of *TBX5* genotype suggested potential cell autonomous effects by *TBX5* dosage (for example, *PLN*, *RYR2*, *TTN*), while those with high anti-correlation (Pearson
250 coefficient <-0.5) suggested non-cell autonomous effects (for example, *HES1, TLE1, CBX1,*

ETV4, ID4, FGFR1). *MEF2C* expression was among the highest correlated with *TBX5*
252 expression and demonstrated greatest *TBX5*-dependent decrease of pagerank (Fig. 4c, e) or degree (i.e. putative number of target genes, Supplemental Table 3) at d11, further suggesting
254 *MEF2C* as a candidate for mediating TBX5 dose-sensitive regulatory effects. Multiple genes that correlated with *MEF2C* displayed diminished levels of degree by reduced TBX5 dosage (for
256 example, *SMYD1, PDK1, MYOCD, TTN, RYR2*; top 5% cutoff) (Fig. 4f, Supplementary Table 3), supporting a potential explanation for changes to *MEF2C* functional connectivity in a TBX5
258 dosage-sensitive GRN.

We evaluated the predicted genetic interaction between *TBX5* and *MEF2C in vivo*. We
260 used a hypomorphic allele of *Tbx5* (*Tbx5*^{CreERT2IRES2xFLAG}, abbreviated *Tbx5*^{CreERT2}) [33], which we further characterized (Extended Data 16) to probe for a potentially highly-sensitive genetic
262 interaction between *Tbx5* and a null allele of *Mef2c (Mef2c*^{del}*)* [32]. *Tbx5*$^{CreERT2/+}$*;Mef2c*$^{del/+}$ mice were underrepresented at weaning (Fig. 5a). By histology, we detected muscular (n=3 of 4) or
264 membranous (n=1 of 4) VSDs in compound heterozygous embryos at E14.5, which were not apparent in *Tbx5*$^{CreERT2/+}$ or *Mef2c*$^{del/+}$ embryos (Fig. 5b-k). These findings demonstrate a highly-
266 sensitive genetic interaction between *Tbx5* and *Mef2c* in mouse *in vivo*, consistent with prediction from the human TBX5 dose-sensitive GRN.

268 We speculated that MEF2C may play a direct role to co-regulate TBX5-dependent gene expression. Chromatin occupancy [34] of TBX5, MEF2c, and MEF2a (also predicted to be part of
270 the TBX5-dependent GRN) was highly correlated near TBX5-sensitive human or mouse genes (for example, *HAND2*, *FHL2*, *TECRL*, *NPPA*/*NPPB*; Fig 5i-q, Extended Data 17, 18, 19;
272 FDR<0.05 for multiple comparisons). Thus, direct co-regulation of target genes by TBX5, MEF2c, and MEF2a, in addition to previously known interactions with NKX2-5 and GATA4, may
274 be a potential TBX5 dosage-dependent mechanism for CHDs.

Discussion

Our studies with a human cellular model of CHD has defined consequences of reduced TBX5 dosage during cardiac differentiation at single cell resolution. Of potential relevance to a range of anatomical and functional manifestations of *TBX5* haploinsufficiency, we uncovered discrete responses to reduced TBX5 dosage in susceptible cardiomyocyte subtypes. The quantitative specificity of TBX5-dependent cell types underscores cellular complexity in response to transcription factor dosage. Analysis of human TBX5 dose-sensitive GRNs identified vulnerable nodes enriched for CHD-associated or cardiac development genes, suggesting a vital role for TBX5 dosage to maintain cardiac network stability. The sensitivity of a GRN to transcription factor dosage has been observed in *Drosophila* embryo patterning [35] for example, but has not been linked to human disease to date. The biophysical rules regulating transcription factor dosage sensitivity are only now becoming understood, and our results in a human CHD model may bring immediate pertinence of human disease to this biological context.

From TBX5-sensitive GRNs, we discovered several important nodes linking many CHD-related genes. For example, the reduced centrality of *MEF2C* in the TBX5-dependent GRN suggests an important and sensitive link between these cardiac transcription factors. Indeed, we validated this predicted genetic interaction in mouse, which yielded muscular VSDs, a type of CHD rarely observed in mouse models but common in humans. It is likely that other sensitive nodes in a TBX5-dependent network will illuminate novel aspects of CHD and cardiac gene regulation. Variability in CHDs with monogenic inherited or *de novo* mutations could be explained by additional mutations in genes that form part of these functional networks, as illuminated by our findings, and as evidenced by oligogenic inheritance of CHD-causing variants [11]. Our results point to a genomic framework that will guide genetic insights into the underpinnings of CHD.

Contributions. I.S.K. and B.G.B conceived and designed the project. B.I.G., L.W., L.B., T.S.
302 and I.S.K. performed gene targeting and isolation of mutant iPSCs. B.I.G., K.S.R, P.G., T.S.,
and I.S.K. performed in vitro differentiation and harvested samples. P.G. performed the
304 Western analysis. M.H.L. performed electrophysiology analyses. R.T. performed statistical
analyses for electrophysiology. K.S.R. performed immunostaining and scoring of
306 cardiomyocytes. G.A.A. performed RNAscope and flow cytometry. K.S.R. and I.S.K. performed
Seurat analyses. K.S.R., A.P.B., H.G., and I.S.K. performed pseudotime analyses. H.G.
308 implemented the cell browser. G.I. performed gene regulatory network analyses. W.P.D. and
I.S.K. performed phenotype analyses of mutant mice. I.S.K. performed bulk RNAseq analyses.
310 B.N.A., F.G., K.L., and W.T.P. performed ChIP-seq experiments and peak calling. S.K.H. and
R.T. performed association analyses of co-occupancy, gene expression and disease
312 candidates. J.M.S., W.T.P., C.E.S., J.G.S., and H.H. provided advisement. I.S.K. and B.G.B.
wrote the manuscript, with comments and contributions from all authors.

314

Acknowledgements. We thank Dario Miguel-Perez and Sarah Wood for mouse genotyping
316 and colony maintenance, Jeff Farrell for input on URD, David Joy and Todd McDevitt for sharing
in-house imaging software, Brian Black for mouse lines, Kathryn Claiborn for editorial
318 assistance, and members of the Bruneau lab for discussions and comments. We also thank the
Gladstone Bioinformatics, Genomics, Histology and Microscopy, and Stem Cell Cores, R. Sarah
320 Elmes at the UCSF Laboratory for Cell Analysis, the UCSF Center for Advanced Technology,
the Salk Institutes Center of Excellence for Stem Cell Genomics, Matthew Speir and Maximilian
322 Haeussler at cells.ucsc.edu, and the UCSC Stem Cell Data Center Hub for their invaluable
assistance. This work was supported by grants from the National Institutes of Health (NHLBI
324 Bench to Bassinet Program UM1HL098179 to B.G.B. and UM1HL098166 to J.G.S. and W.T.P.,
2T32HL007731-27 to S.K.H), , the Foundation for Anesthesia Education and Research

326 (Mentored Research Training Grant to I.S.K.), Society for Pediatric Anesthesia (Young
Investigator Award to I.S.K.), Hellman Family Fund (I.S.K.), UCSF REAC Grant (I.S.K.) and
328 UCSF Department of Anesthesia and Perioperative Care (New Investigator Award to I.S.K.).
H.H. is a Miguel Servet (CP14/00229) researcher supported by the Spanish Institute of Health
330 Carlos III (ISCIII) and Ministerio de Ciencia, Innovación y Universidades (SAF2017-89109-P;
AEI/FEDER, UE). This work was also supported by the California Institute for Regenerative
332 Medicine (RB4-05901 to B.G.B), an NIH/NCRR grant (C06 RR018928) to the J. David
Gladstone Institutes, the Younger Family Fund (B.G.B.), and the Office of the Assistant
334 Secretary of Defense for Health Affairs through the Peer Review Medical Research Program
under Award No. W81XWH-17-1-0191. Opinions, interpretations, conclusions and
336 recommendations are those of the author and are not necessarily endorsed by the Department
of Defense. In conducting research using animals, the investigator(s) adheres to the laws of the
338 United States and regulations of the Department of Agriculture. In the conduct of research
utilizing recombinant DNA, the investigator adhered to NIH Guidelines for research involving
340 recombinant DNA molecules.

342 **Competing Interests**: B.G.B. is a co-founder and shareholder of Tenaya Therapeutics. None of
the work presented here is related to the interests of Tenaya Therapeutics.

344

346

348 **Methods**

Gene targeting and genotyping of human iPS cells mutant for *TBX5*. sgRNAs for *TBX5*
350 exon 3 (sgRNA1, TCCTTCTTGCAGGGCATGGA) or exon 7 (sgRNA2,
CCTTTGCCAAAGGATTTCG), which encode the T-box domain, were selected using
352 crispr.genome-engineering.org, and cloned by annealing pairs of oligos into a plasmid
containing humanized *S. pyogenes* Cas9, as described in [36] (px330-U6-Chimeric_BB-CBh-
354 hSpCas9 was a gift from Feng Zhang, Addgene #42230).

For WTC11-derivatives $TBX5^{+/+}$ (control), $TBX5^{in/+}$ or $TBX5^{in/del}$, the induced pluripotent
356 stem (iPS) cell line WTC11 (gift from Bruce Conklin, available at NIGMS Human Genetic Cell
Repository/Coriell #GM25236) [37] was electroporated (Lonza #VPH-5012) with a cloned
358 nuclease construct containing a guide RNA (sgRNA1) targeting exon 3 of *TBX5*, as described in
[37,38]. Cells were plated on human ESC-grade Matrigel (Corning #354277) and cultured in
360 mTeSR-1 (StemCell Technologies Cat #05850) with 10μM ROCK inhibitor (StemCell
Technologies, Y-27632). For screening of *TBX5* exon 3 non-homologous end-joining (NHEJ)
362 mutations, genomic DNA flanking the targeted sequence was amplified by PCR (For1:
ATGGCATCAGGCGTGTCCTATAA and Rev1: CCCACTTCGTGGAATTTTAGCCA), amplicons
364 underwent digestion by NlaIII, and then evaluated for loss of NlaIII by gel electrophoresis
(wildtype band 800bp, mutant band 880bp). Clones with no change, a heterozygous or
366 homozygous loss of NlaIII were sequenced (For1: ATGGCATCAGGCGTGTCCTATAA, Rev1:
TTCCGGGCTTGAACTTCTGG, Seq1: ATAGCCTTGTGCTGATGGCA).

368 For generation of $TBX5^{PuR/PuR}$, a puromycin resistance gene cassette (Frt-PGK-EM7-
PuroR-bpA-Frt) containing homology arms of 469bp (5' homology arm) and 466bp (3'
370 homology arm) around the sgRNA1 target site at +9bp from the start of *TBX5* exon 3 was
cloned by Cold Fusion (System Biosciences #MC010B) using amplicons from genomic DNA of
372 WTC11 into a construct that was a modification of plasmid pEN114[39]. WTC11 cells were

electroporated with a cloned nuclease construct containing a guide RNA targeting exon 3, along
374 with the *TBX5* exon3 homology arm-Frt-PGK-EM7-PuroR-bpA-Frt cassette and plated as a
serial dilution in mTeSR-1 with Rock inhibitor, as described in [38]. On day 2 and subsequent
376 days, cells were grown in media containing mTeSR-1, Rock inhibitor and puromycin (0.5ug/mL),
to select for puromycin-resistant cells. For screening of *TBX5* exon 3 homology-directed repair
378 (HDR) mutations, genomic DNA flanking the targeted sequence was amplified by PCR (For1:
ATGGCATCAGGCGTGTCCTATAA, and Rev2: CCCACTTCGTGGAATTTTAGCCA for
380 wildtype, 797 bp, For1: ATGGCATCAGGCGTGTCCTATAA, Rev3:
GTTCTTGCAGCTCGGTGAC [39] for PuroR, 1631 bp). Positive 5' arm clones were genotyped by
382 PCR for the 3' arm (For2: ATTGCATCGCATTGTCTGAG [39], Rev4:
TTTGACAATCGGGTGGGACC, 829 bp).

384 For PGP1-derivatives $TBX5^{in/+}$ or $TBX5^{del/del}$, the iPS cell line PGP1 (gift from George
Church, available at NIGMS Human Genetic Cell Repository/Coriell #GM23338) [40] was
386 electroporated with a cloned nuclease construct containing a guide RNA (sgRNA2) targeting
exon 7 of *TBX5*, as described in [41]. For screening of *TBX5* exon 7 NHEJ mutations, the targeted
388 sequence was amplified using PCR primers (For3: GCTTCTTTTGGTTGCCAGAG, Rev5:
CATTCTCCCCATTTCCATGT, Seq2: AGAGGCTGCATTTCCATGAT), Illumina compatible-
390 libraries from clones were generated and multiplex-sequenced on a MiSeq for purity of
homogeneity of clones for heterozygous or homozygous mutations, as described in [41].

392

Isolation of homogenous iPS cell clones. Isolation of homogenous colonies for WTC11-
394 derivatives $TBX5^{+/+}$ (control), $TBX5^{in/+}$ or $TBX5^{in/\mathrm{del}}$ was performed by modification of methods
described previously [38,42]. Briefly, single cell suspension of electroporated iPS cells was plated
396 on Matrigel-coated 6 well plates (WP) (BD Bioscience #351146). Once cultures were adherent
and recovered to ~80% confluency, cells were detached by Accutase Cell Detachment Solution
398 (Stemcell Technologies #07920), diluted with 1X DPBS without Ca^{2+}/Mg^{2+} and singularized

using a P1000 filtered tip, and centrifuged. The cell pellet was resuspended in mTeSR-1, Rock
400 inhibitor and Gentamicin (Life Technologies #15750-060) media, incubated with DAPI (1:1000
from a 1mg/mL stock) for 5 min, centrifuged and resuspended at a concentration of at least
402 1.0E6 cells/mL in mTeSR-1, Rock inhibitor and Gentamicin media without DAPI. After filtering
cells with a 40-micron mesh into FACS tubes, remaining cells (about 120,000 cells per well)
404 were plated on to 6WP for maintenance. Single cells were then sorted for DAPI negativity using
a BD FACS AriaII or AriaIII, with a 100-micron nozzle at the lowest flow rate available, into
406 individual wells of a 96WP coated with Matrigel containing media of mTeSR-1, Rock inhibitor
and Gentamicin. Upon recovery at 37°C, each well was evaluated one day later for no cells, one
408 cell or more than one cell. All cells were maintained with mTeSR-1, Rock inhibitor and
Gentamicin media for at least 5 days, then with mTeSR-1 alone for an additional 5-7 days. Each
410 well at 25% confluency was harvested and re-plated upon singularization with P200 tips in
96WP for more efficient cell growth. When the cell confluency of each well from “single” cells
412 was nearly 100%, then 90% of cells were harvested for genotyping using QuickExtract DNA
lysis solution (Epicentre #QE0905T), while 10% of cells were re-plated for the next round of cell
414 selection for wells of interest by FACS sorting again or by serial dilution of cells for manual
picking of colonies, as described in [37,38] from apparent “single” cells. Rounds were repeated until
416 every daughter well showed the same genotype, consistent with homogeneity. Genomic DNA
from individual wells of interest were amplified using high fidelity *Taq* polymerase, TA-cloned
418 and sequenced to confirm genotype and homogeneity.

Isolation of a homogenous colonies for PGP1-derivatives $TBX5^{in/+}$ or $TBX5^{del/del}$ was
420 performed, as described in {Byrne:2007fn}. Isolation of a homogenous colonies for WTC11-
derivative $TBX5^{PuR/PuR}$ was performed, as described in [38].

422 After sequencing confirmation of respective genotypes, karyotypically-normal cells from
each iPS cell line were expanded for subsequent studies.

424

Maintenance of iPS cells and differentiation to cardiomyocytes. All iPS cell lines were
426 transitioned to and maintained on growth factor-reduced basement membrane matrix Matrigel
(Corning #356231) in mTeSR-1 medium. For directed cardiac differentiations, iPS cells were
428 dissociated using Accutase and seeded on to 6WP or 12WP. The culture was allowed to reach
80-90% confluency and induced with the Stemdiff Cardiomyocyte Differentiation Kit (Stemcell
430 Technologies #05010), according to the manufacturer's instructions. Starting on day 7,
differentiations were monitored daily for beating cardiomyocytes and onset of beating was
432 recorded as the day when beating was first observed.

434 **Flow Cytometry.** iPS-derived cardiomyocytes from WTC11, Control, *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$
lines were dissociated using Trypsin-EDTA 0.25% on day 15 or day 23 after induction of the
436 differentiation protocol and fixed with 4% methanol-free formaldehyde. Cells were washed with
PBS and permeabilized using FACS buffer (0.5% w/v saponin, 4% Fetal Bovine Serum in PBS).
438 For evaluation of differentiation efficiency, cells were stained with a mouse monoclonal antibody
for cardiac isoform Ab-1 Troponin at 1:100 dilution (ThermoFisher Scientific #MS-295-P) or the
440 isotype control antibody (ThermoFisher Scientific #14-4714-82). For analyzing levels of Desmin
protein, cells were co-stained with the mouse monoclonal antibody for cardiac isoform Ab-1
442 Troponin at 1:100 dilution and recombinant rabbit anti-Desmin antibody at 1:70 dilution (Abcam
#ab32362), or normal rabbit IgG antibody (Millipore Sigma #NI01) for 1 hour at room
444 temperature. After washing with FACS buffer, cells were stained with the following secondary
antibodies - goat anti-mouse IgG Alexa 594 at 1:200 dilution (ThermoFisher Scientific #A-
446 11005) and donkey anti-rabbit IgG Alexa 488 at 1:200 dilution (ThermoFisher Scientific
#A21206) for 1 hour at room temperature. Cells were then washed with FACS buffer, stained
448 with DAPI for 5 minutes, rinsed, and filtered with a 40-micron mesh. At least 10,000 cells were
analyzed using the BD FACSAriaII or AriaIII, and results were processed using FlowJo (BD
450 Bioscience). Gating strategy is shown in Extended Data 20.

452 **Western blotting.** iPS-derived cardiomyocytes were harvested on day 15, pelleted and flash frozen. Protein was isolated from supernatant in RIPA buffer with EDTA-free protease and
454 phosphatase inhibitor (ThermoFisher Scientific) after sonication (15 second pulse on, 15 second pulse off, for four pulses). After quantification by BCA assay (ThermoFisher Scientific), 150µg of
456 total protein was loaded per well for each genotype. After running on SDS-PAGE and wet transfer with NuPage Transfer buffer (ThermoFisher Scientific) to a PVDF membrane, the blot
458 was washed in PBST and incubated in primary antibodies of rabbit polyclonal anti-TBX5 at a 1:400 dilution (Sigma #HPA008786) and mouse monoclonal anti-cTNT at 1:1000 dilution
460 (ThermoFisher Scientific #MS-295-P), followed by secondary antibody incubation with donkey anti-rabbit IgG IRDye680 at 1:2000 dilution (Licor #926-68073) and donkey anti-mouse IgG
462 IRDye800 at 1:2000 dilution (Licor #926-32212). The blot was imaged on an Odyssey FC Dual-Mode Imaging system (Licor).
464

Fluorescent *in situ* hybridization. iPS cell-derived cardiomyocytes from WTC11, Control,
466 *TBX5*$^{in/+}$,*TBX5*$^{in/del}$ and *TBX5*$^{PuR/PuR}$ were dissociated using Trypsin-EDTA 0.25% on day 23 after induction of the differentiation protocol, and 25,000-40,000 cells were plated on to 8-well
468 chambered slides (Ibidi #80826), to obtain a relatively sparse monolayer of cardiomyocytes. Cells were fixed the following day with 10% Neutral Buffered Formalin for 15 minutes at room
470 temperature. Cells were then serially dehydrated in 50%, 70% and 100% ethanol and stored at -20°C until ready to be hybridized. *In situ* hybridization was performed using the RNAscope
472 Multiplex Fluorescent v2 Assay kit (Advanced Cell Diagnostics #323100) with probes for *Hs-TNNT2* (#518991) and *Hs-NPPA* (#531281). Slides were imaged at 10X and 40X magnification
474 on the Keyence BZ-X710 All-in-One Fluorescence Microscope. Mean intensity of *NPPA* signal was measured in each *TNNT2*+ cell from every group. Unpaired t-tests were used to calculate
476 statistical significance.

478 **Replating cardiomyocytes for single cell electrophysiology.** iPS cell-derived cardiomyocytes (day 15 or older) from WTC11, Control, *TBX5*$^{in/+}$, *TBX5*$^{in/del}$ and *TBX5*$^{PuR/PuR}$
480 were gently dissociated in Trypsin-EDTA 0.25% and quenched using StemDiff Maintenance Medium with 10% FBS. Cell suspension was centrifuged at 800 rpm for 5 minutes. The pellet
482 was resuspended in StemDiff Maintenance Medium with Rock inhibitor at a 1:1000 dilution. Cardiomyocytes were counted, and 25,000-35,000 cells were plated on to growth factor-
484 reduced Matrigel-coated 15mm round glass coverslips (Warner Instruments #64-0703) to obtain a sparse distribution. Cardiomyocytes were then maintained on coverslips in StemDiff
486 Maintenance Medium.

488 **Patch Clamp Electrophysiology.** Patch clamp recordings were made on single iPSC-derived CMs using the perforated-patch configuration. Experiments were performed at 30°C under
490 continuous perfusion of warmed Tyrode's solution containing (in mM): 140 NaCl, 5.4 KCl, 1 $CaCl_2$, 1 $MgCl_2$, 10 glucose, and 10 HEPES, with the pH adjusted to 7.4 with NaOH.
492 Recordings were conducted using borosilicate glass pipettes (Sutter Instruments) with typical resistances of 2 to 4MΩ. The pipette solution consisted of (in mM): 150 KCl, 5 NaCl, 5 MgATP,
494 10 HEPES, 5 EGTA, 2 $CaCl_2$, and 240 μg/mL amphotericin B, with the pH adjusted to 7.2 with KOH. Spontaneous action potentials were acquired in a zero-current current clamp
496 configuration using an Axopatch 200B amplifier and pClamp 10 software (Axon Instruments). Data was digitized at 20 kHz and filtered at 1kHz. Action potential parameters from each cell
498 were derived using Clampfit 10 software (Axon Instruments).

500 **Fluorescent recordings of calcium flux.** iPSC-derived CMs on glass coverslips were loaded with Ca^{2+} indicator dye Fluo-4 AM (Thermo Fisher Scientific #F14201) to record Ca^{2+} flux, as

502 previously described [43]. Measurements were made on spontaneously firing single or small clusters of iPSC-derived CMs using a 10X objective on a Zeiss Axio Observer Z1 inverted
504 microscope. For experiments, cells were placed in Tyrode's solution containing 1.8 mM Ca^{2+} within a 37°C heated stage-top imaging chamber (Okolab). Images were acquired at 100 fps
506 using an ORCA-Flash 4.0 camera (Hamamatsu, Bridgewater, NJ). Data was processed using ZEN (Zeiss) or Image J software (http://rsbweb.nih.gov/ij/) and analyzed using custom in-house
508 software [44].

510 **Graphing and statistics for electrophysiology.** For electrophysiology and calcium imaging experiments, graphs were generated using Prism 8.2.0 (GraphPad Software). Significance
512 between parental and experimental groups was determined with a custom R-script using unpaired two-sided Welch's *t* tests with Holm-Sidak correction for multiple comparisons [12].
514 Adjusted $P < 0.05$ was considered statistically significant.

516 **Immunostaining and scoring of cardiomyocytes.** iPSC-derived cardiomyocytes from WTC11, Control, $TBX5^{in/+}$ and $TBX5^{in/del}$ were replated on coverslips placed in 12-well plates on
518 day 23, as described above for replating for electrophysiology. Cells were fixed in 4% formaldehyde for 20 minutes at room temperature, followed by washes in PBS. Cells were then
520 treated with a blocking buffer containing 5% goat serum and 0.1% Triton X-100 in PBS for 1 hour at room temperature. A mouse monoclonal antibody for cardiac isoform Ab-1
522 Troponin (ThermoFisher Scientific #MS-295-P) was added to the coverslip-containing wells at a 1:100 dilution in blocking buffer and incubated on a rocker for 2 hours at room temperature.
524 Following washes with 0.1% Triton X-100 in PBS, coverslips were treated with a donkey anti-rabbit IgG Alexa 488 antibody (ThermoFisher Scientific #A21206) at a 1:200 dilution for 2 hours
526 at room temperature. Coverslips were then washed with 0.1% Triton X-100 in PBS and stained

with DAPI at a 1:1000 dilution for 2 minutes. Coverslips were washed and stored in PBS at 4C.
528 Images were acquired on a Zeiss LSM 880 with Airyscan and processed by ImageJ [45].

Myofibrillar arrangement in cardiomyocytes was manually scored on a scale of 1-5,
530 similar to [46]. A score of 1 represents cells with intact myofibrils in a parallel arrangement. A score of 2 represents cells that have intact myofibrils, but many are not parallel. Scores of 3 and
532 of 4 include cells with increasing degrees of myofibrillar fragmentation or aggregation. A score of 5 represents cells without visible myofibrils. No cells were apparent among our samples with
534 a score of 5. Violin plots were generated in Prism (GraphPad) to show distribution of scored cells from each group. Fisher's exact test was used to determine statistical significance.

536

Cell harvesting for single cell RNAseq. Cells from day 6, day 11 or day 23 of the
538 differentiation protocol were collected from 3 independent differentiations. Wells for dissociation were chosen based on typical differentiated morphology on day 6 or robust beating on day 11
540 and day 23. Cells were singularized with Trypsin-EDTA 0.25%. After quenching, the single cell suspension was centrifuged at 800 rpm for 5 minutes. The pellet was resuspended in 1X PBS
542 with 0.04% w/v Ultrapure BSA (MCLAB #UBSA-500) and counted. A 30μL cell suspension containing 10,000 cells was used to generate single cell droplet libraries with the Chromium
544 Single Cell 3′ GEM, Library & Gel Bead Kit v2 according to manufacturer's instructions (10X Genomics). After KAPA qPCR quantification, a shallow sequencing run was performed on a
546 NextSeq 500 (Illumina) prior to deep sequencing on a NextSeq 500, HiSeq 4000, or NovaSeq (Illumina) for a read depth of >100 million reads per cell.

548

Data processing using Cellranger. All datasets were processed using Cellranger 2.0.2.
550 FASTQ files were generated using the mkfastq function. Reads were aligned to hg19 reference (version 1.2.0). Cellranger aggr was used to aggregate multiple GEM libraries.

552

Seurat analysis. Outputs from the Cellranger pipeline were analyzed using the Seurat package
554 (version 2.3.4) in R (version 3.5.1) [R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
556 URL https://www.R-project.org/]. Datasets from day 6, day 11 or day 23 experiments were analyzed as separate Seurat objects. Quality control steps were performed to remove dead
558 cells or doublets, and cells with a UMI count between 10,000 to 80,000 were retained. After normalizing the data, sources of unwanted variation, such as differences in the number of UMI,
560 number of genes, percentage of mitochondrial reads and differences between G2M and S phase scores were regressed using the ScaleData function. Next, principal component analysis
562 (PCA) was performed using the most highly variable genes. Cells were then clustered based on the top 25-30 principal components and visualized using a dimensionality reduction method
564 called Uniform Manifold Approximation and Projection (UMAP) [47]. The resolution parameter was set, so that cluster boundaries separated the major cell types.

566 For day 11 or day 23 cardiomyocyte datasets, *TNNT2*$^{+}$ clusters were defined as containing a majority of cells expressing *TNNT2* on a feature plot and extracted using the
568 SubsetData function and re-clustered. Subsequently, the resolution parameter was set to partition clusters enriched for a particular genotype. A phylogenetic tree was generated by
570 relating the “average” cell from each cluster by centroid analysis, using the BuildClusterTree function. Differential gene expression tests were run between closely related clusters, using the
572 FindMarkers function with min.pct set to 0.1 and logfc.threshold set to 0.25. Selected differentially expressed genes with an adjusted p-value less than 0.05 from the Wilcoxon Rank
574 Sum test were then displayed using the Dotplot function. As Seurat log normalizes gene expression counts and scales values for each gene (mean is 0, std dev of +/-1), dot plots and
576 heatmaps are based on scaled expression values.

Major cell type categories were defined by their expression of certain marker genes -
578 cardiomyocytes (*TNNT2*), dividing cardiomyocytes (*CENPF*+*TNNT2*), fibroblasts (*COL1A1*),

endoderm (*TTR* alone or *TTR+AFP*) and endothelial cells (*PLVAP*). Clusters of cells not defined
580 by any of these markers were labeled as 'Others'. The numbers of cells in each major cell type category in each genotype were then calculated. Sunburst plot was generated in Excel using
582 the percentage of cells in each cell type category per genotype. We used FindAllMarkers to generate a list of top marker genes for each cluster and presented two selected genes in a dot
584 plot to display potential diversity of subtypes among these major cell types.

586

Congenital Heart Disease-Associated or Electrophysiology-Related Gene Lists and Cell-
588 **Type Expression**. A list of 376 CHD-associated candidate genes were manually curated from [18-25], including inherited, *de novo*, syndromic or non-syndromic CHD-associated genes. A list of
590 76 EP-linked genes were manually curated. A list of cardiac development-related factors is from [48]. Lists can be found in Supplementary Table 1.
592 To determine cell-type enrichment of each CHD-associated gene, cells in the day 23 dataset were assigned one of five cell-type labels (cardiomyocytes, fibroblasts, endoderm,
594 endothelial and others). A dot plot was generated to visualize the expression of CHD-associated genes in the 5 cell types. Genes were first manually curated for enrichment in a certain cell type,
596 based on expression. Statistical significance of the enrichment of a certain gene in one cell type over the others was then evaluated by Wilcoxon Rank Sum test (adjusted p-value < 0.05).

598

Cell trajectories and pseudotime analysis. Pseudotime analysis was performed using the
600 URD package (version 1.0.2). A single Seurat object with data from three timepoints and four genotypes was processed, as described in the previous section, and then converted to an URD
602 object using the seuratToURD function. Cell-to-cell transition probabilities were constructed, by setting the number of nearest neighbors (knn) to 211 and sigma to 8. Pseudotime was then
604 calculated by running 80 flood simulations with *POU5F1+* clusters as the 'root' cells. Next, all

day 23 clusters were set as 'tip' cells and biased random walks were simulated from each tip to
606 build an URD tree.

To identify genes that correlate with pseudotime in one genotype but not the other, we
608 calculated Spearman rank correlation, using Python (v3.7.3, and libraries Pandas 0.25.0, Numpy 1.17.1, and SciPy 1.3.1), to find genes that share a significant monotonic relationship (p-
610 value less than 0.05) with pseudotime. To determine if these monotonic relationships differ between WT and *TBX5*$^{in/del}$ paths to cardiomyocytes, we used a Fisher z-transformation to test
612 the null hypothesis that there is no significant difference in correlation [49]. To illustrate these results, we use a scatter plot for all genes with a significant rho value (p-value<0.05 by two-
614 sided *t* test), or heat maps for genes with a |rho|≥0.4 to pseudotime and Z-score≥15 as a difference between WT and *TBX5*$^{in/del}$ paths.ki

616 To identify differential expressed genes in inferred cardiac precursors (stems in the URD tree) that are affected by *TBX5* loss, cell barcodes from each precursor segment (69 for
618 wildtype/control/*TBX5*$^{in/+}$ path and 70 for *TBX5*$^{in/del}$ path) were extracted from the URD object and assigned new identities in the corresponding Seurat object. Differential gene test was then
620 performed between the two segments using Wilcoxon Rank Sum test with min.pct set to 0.1 and logfc.threshold set to 0.25. Selected genes with an adjusted p-value less than 0.05 were plotted
622 on the URD tree to visualize their expression.

To compare the trident (*TNNT2*$^{+}$ distal branch for WTC11, control and *TBX5*$^{in/+}$) and fork
624 (*TNNT2*$^{+}$ distal branch for *TBX5*$^{in/del}$) during pseudotime, we subdivided the pseudotime from the common branchpoint to the tips of the trident and fork into twenty uniform windows. Within each
626 window, we then calculated the t-test, difference of means, and fold change between the trident and fork for all genes. We filtered the statistics by gene-window combinations with p<0.05 after
628 Bonferroni-Holm multiple testing correction. Then, we hierarchically clustered the genes on t-test p-values and plotted statistics using the R pheatmap library.

630

Gene ontology analysis. Gene ontology analysis was performed using 2018 biological process terms from org.Hs.eg.db (v3.8.2) and clusterProfiler (v 3.12.0), with a significance threshold set at p-value<0.05. To prevent high false discovery rate (FDR) in multiple testing, we used the default cutoff for the estimated q-value<0.05. Figures were produced using Matplotlib (v3.1.0). Gene ratios indicate the number of differentially expressed genes that overlap with each GO term divided by the total number of differentially expressed genes. For comparing *TNNT2*$^{+}$ clusters at d23, enriched gene ontology (GO) terms were determined using genes computed from a one-versus rest differential test (logfc>0.25; adj p-value<0.05 by Wilcoxon Rank Sum test) among clusters that were predominantly composed of the same *TBX5* genotype (WT-21, 12, 3, and Control-4, or *TBX5*$^{in/+}$- 9, 7, 6, and 0). Enriched GO terms associated with pseudotime in a WT- or *TBX5*$^{in/del}$ path were generated using genes positively correlated with pseudotime (rho≥|0.4|, and Z-score≥15) from each path. Several significant GO terms are highlighted.

Cell browser implementation. The cell browser at cells.ucsc.edu was developed by Maximilian Haeussler. We created a cell browser session that allows the user to interrogate the spatial distribution of metadata and expression across data, in multiple reduced dimensionality spaces including the URD trajectory. Using a Scanpy python pipeline, we generated PCA, tSNE, UMAP, PAGA, and drl transforms. We also imported the URD trajectory mapping and WGCNA transform from their respective packages. We ran the scoreCT algorithm to assign cell types to cell clusters using a marker gene set.

Gene regulatory network analysis. bigSCale2 (https://github.com/iaconogi/bigSCale2) [30] was used with default parameters to infer gene regulatory networks and “correlomes” from single cell RNAseq expression data for *TNNT2*$^{+}$ cells, simply passing as input expression counts and gene names. Details of each data set can be found in Supplementary Table 3. To evaluate significant

changes in pagerank (or degree) centrality, we computed all pairwise differential differences in
658 pagerank (or degree) between baseline (wildtype or control) and *TBX5* mutants (*TBX5*$^{in/+}$ or *TBX5*$^{in/del}$) (12 total differences, from 2 *TBX5* mutants * 2 baselines * 3 stages) and used these
660 values to determine the top 5% upper change cutoff from 8,704 genes of all networks. Classification of Pearson correlations were empirically chosen at >0.5 for correlation and <-0.05
662 for anti-correlation.

Mice. All mouse protocols were approved by the Institutional Animal Care and Use Committee
664 at UCSF. *Tbx5*$^{del/+}$ [7] and *Tbx5*^{CreERT2IRES2xFLAG} (abbreviated here as *Tbx5*^{CreERT2}) [33] mice were described previously. *Tbx5*$^{fl\text{-}bio/fl\text{-}bio}$[50] mice were obtained from Frank Conlon. *Mef2c*$^{del/+}$ mice [32]
666 were obtained from Brian Black. *Mef2a*$^{fl\text{-}bio}$ and *Mef2c*$^{fl\text{-}bio}$ (Jackson #025983) were described in [34]. *Rosa26BirA* mice were obtained from Jackson labs (Jackson #010920) [51].

668 **ChIP-seq.** Combined peaks of human TBX5 or GATA4 ChIP-seq from hiPSC-derived CMs were used [10]. bioChIP-seq of mouse TBX5, MEF2c and MEF2a from E12.5 hearts were from [34].
670 Single replicates of TF bioChIP peaks, which were IDR normalized (IDR_THRESHOLD=0.05 between each set of replicates), were defined as the summit of the peak with the strongest ChIP
672 signal ± 100bp of the individual replicate with the greatest peak intensity. Mouse H3K27ac ChIP-seq at E12.5 of embryonic cardiac ventricles was from [52].

674 **Bulk RNAseq analysis of embryonic mouse hearts.** Embryos at E10.5 from timed matings of *Tbx5*$^{del/+}$ x *Tbx5*$^{CreERT2/+}$ were harvested. Each whole heart was dissected, placed in Qiazol, and
676 homogenized with a 25-guage needle. Total RNA was extracted using the miRNeasy micro kit with on-column DNAse digestion (Qiagen) and assessed by bioanalyzer (Agilent) for RNA
678 integrity, as previously described [53]. Four individual hearts of each genotype (wildtype, *Tbx5*$^{CreERT2/+}$, *Tbx5*$^{del/+}$ or *Tbx5*$^{del/CreERT2}$) from multiple litters were used as biological
680 quadruplicates. RNAseq libraries were generated using the Ovation RNA-seq system V2 kit

(NuGen), and SPIA-amplified cDNA was made using the Ovation Ultralow System V2 kit (NuGen).
682 Libraries were evaluated by bioanalyzer (Agilent), quantified using KAPA qPCR (Kapa Biosystems), and sequenced by paired-end 100 bp reads using a HiSeq2500 (Illumina), for a
684 sequencing depth of >40 million reads per library. Reads were aligned to mm9 by TopHat2 [54] and counts per gene were tallied by featureCounts [55]. Differentially expressed genes were
686 identified by EdgeR[56], using cutoffs of log2 fold-change of >0.5 or <-0.5, and raw p-value <0.01. R functions, K-means clustering and pheatmap, were used to cluster and create bulk RNAseq
688 heatmaps.

690 **Statistical analyses for correlations.** We evaluated the pair-wise association among 24 variables, including all human genes, TBX5-dysregulated genes in cardiomyocytes from two
692 different stages of differentiation (d11 or d23), CHD-associated genes, EP-linked genes, TBX5 or GATA4 binding [10], and genome-wide association (GWAS) genes for CHDs or arrhythmias.
694 Reported GWAS genes from https://www.ebi.ac.uk/gwas/ for the terms congenital heart disease, congenital heart malformation, congenital left sided heart lesions, conotruncal heart
696 defect and aortic coarctation were used to define congenital heart disease-related (CHD) GWAS genes. Reported genes from terms such as cardiac arrhythmia, supraventricular ectopy,
698 ventricular ectopy, premature cardiac contractions, atrial fibrillation, sudden cardiac arrest and ventricular fibrillation were considered as arrhythmia-related (EP-GWAS) genes. Two nearest
700 genes within 100kb, by using GREAT (great.stanford.edu), of TBX5 or GATA4 binding sites or of reported genes from each group of GWAS were considered for the analysis. The natural
702 logarithm odds of genes associating with each one of these variables versus the odds of genes associating with every other variable were estimated using generalized linear models with
704 family="binomial" setting in R. The resulting significances of these natural log odds ratios were adjusted for multiple testing by the Benjamini-Hochberg method [57]. Significance was determined
706 using an FDR threshold of 0.05 or less.

Additional correlations were evaluated between 42 variables, including all mouse genes, mouse TBX5-dysregulated genes (from bulk RNAseq of E10.5 hearts from wildtype, *Tbx5*$^{del/+}$, *Tbx5*$^{CreERT2/+}$ or *Tbx5*$^{CreERT2/del}$ embryos), human TBX5-dysregulated genes from d11 or d23 cardiomyocytes, and TBX5, MEF2c or MEF2a binding sites from E12.5 mouse heart tissue [34]. Human gene symbols were converted to mouse gene symbols, using the getLDS() function from the biomaRt package (https://www.r-bloggers.com/converting-mouse-to-human-gene-names-with-biomart-package/). Two nearest genes within 100kb of TBX5, MEF2c or MEF2a binding sites were considered for the analysis.

For assessment of associations between binding locations of TBX5, MEF2c and MEF2a transcription factors with genes dysregulated by TBX5, analyses were performed corresponding to binding regions of each of the three TFs. First, binding regions of each TF was evaluated for association with genes, defined by nearest two genes within 100kb. Using the list of TBX5-dysregulated genes (mouse or human), binding regions of each TF associated with a TBX5-dysregulated gene was determined. Identified binding regions of each TF that overlapped with at least 50% of the binding regions of each of the other two TFs was determined, using bedops --element-of -50%. This approach defined three variables, including every binding region of the TF, if associated with a TBX5-dysregulated gene, or if it overlaps by at least 50% with the binding region of the other TFs, that were used for logistic regression in R. The resulting changes in odds are represented as natural logarithm odds ratios. Multiple testing correction was performed using the multtest package in R. All estimates are based on analyses for a given species (mouse or human TBX5-dysregulated).

Data availability. scRNAseq and bulk RNAseq datasets have been deposited at NCBI GEO, under accession GSE137876.

Code availability. R and python scripts will be available upon publication.

732 **References**.

1. Zaidi, S. & Brueckner, M. Genetics and Genomics of Congenital Heart Disease. *Circ Res* **120,** 923–940 (2017).
2. Hill, J. T., Demarest, B., Gorsi, B., Smith, M. & Yost, H. J. Heart morphogenesis gene regulatory networks revealed by temporal expression analysis. *Development* **144,** 3487–3498 (2017).
3. Hoffman, J. I. Incidence of congenital heart disease: II. Prenatal incidence. *Pediatr Cardiol* **16,** 155–165 (1995).
4. Hoffman, J. I. E. & Kaplan, S. The incidence of congenital heart disease. *J Am Coll Cardiol* **39,** 1890–1900 (2002).
5. Basson, C. T. *et al.* Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet* **15,** 30–35 (1997).
6. Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nat Genet* **15,** 21–29 (1997).
7. Bruneau, B. G. *et al.* A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. **106,** 709–721 (2001).
8. Mori, A. D. *et al.* Tbx5-dependent rheostatic control of cardiac gene expression and morphogenesis. **297,** 566–586 (2006).
9. Theodoris, C. V. *et al.* Human disease modeling reveals integrated transcriptional and epigenetic mechanisms of NOTCH1 haploinsufficiency. *Cell* **160,** 1072–1086 (2015).
10. Ang, Y.-S. *et al.* Disease Model of GATA4 Mutation Reveals Transcription Factor Cooperativity in Human Cardiogenesis. *Cell* **167,** 1734–1749.e22 (2016).
11. Gifford, C. A. *et al.* Oligogenic inheritance of a human heart disease involving a genetic modifier. *Science* **364,** 865–870 (2019).

12. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6,** 65–70 (1979).

13. Karakikes, I. *et al.* A Comprehensive TALEN-Based Knockout Library for Generating Human-Induced Pluripotent Stem Cell-Based Models for Cardiovascular Diseases. *Circ Res* **120,** 1561–1571 (2017).

14. Nadadur, R. D. *et al.* Pitx2 modulates a Tbx5-dependent gene regulatory network to maintain atrial rhythm. *Science Translational Medicine* **8,** 354ra115–354ra115 (2016).

15. Dai, W. *et al.* A calcium transport mechanism for atrial fibrillation in Tbx5-mutant mice. *eLife* **8,** 40 (2019).

16. Laforest, B. *et al.* Atrial fibrillation risk loci interact to modulate Ca2+-dependent atrial rhythm homeostasis. *J Clin Invest* **412,** 1825–15 (2019).

17. Zhu, Y. *et al.* Tbx5-dependent pathway regulating diastolic function in congenital heart disease. *Proceedings of the National Academy of Sciences* **105,** 5519–5524 (2008).

18. McCulley, D. J. & Black, B. L. Transcription factor pathways and congenital heart disease. *Curr. Top. Dev. Biol.* **100,** 253–277 (2012).

19. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. **498,** 220–223 (2013).

20. Prendiville, T., Jay, P. Y. & Pu, W. T. Insights into the genetic structure of congenital heart disease from human and murine studies on monogenic disorders. *Cold Spring Harb Perspect Med* **4,** a013946–a013946 (2014).

21. Lalani, S. R. & Belmont, J. W. Genetic basis of congenital cardiovascular malformations. *Eur J Med Genet* **57,** 402–413 (2014).

22. Homsy, J. *et al.* De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350,** 1262–1266 (2015).

23. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet* **48,** 1060–1065 (2016).

24. Priest, J. R. *et al.* De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. *PLoS Genet* **12,** e1005963–25 (2016).

25. Jin, S. C. *et al.* Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet* **49,** 1593–1601 (2017).

26. Farrell, J. A. *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **108,** eaar3131–9 (2018).

27. Hiroi, Y. *et al.* Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet* **28,** 276–280 (2001).

28. Luna-Zurita, L. *et al.* Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell* **164,** 999–1014 (2016).

29. Garg, V. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. **424,** 443–447 (2003).

30. Iacono, G., Massoni-Badosa, R. & Heyn, H. Single-cell transcriptomics unveils gene regulatory network plasticity. 1–20 (2019). doi:10.1186/s13059-019-1713-4

31. Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and Isdn Systems* **30,** 107–117 (1998).

32. Lin, Q., Schwarz, J., Bucana, C. & Olson, E. N. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276,** 1404–1407 (1997).

33. Devine, W. P., Wythe, J. D., George, M., Koshiba-Takeuchi, K. & Bruneau, B. G. Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife* **3,** (2014).

34. Akerberg, B. N. *et al.* A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nat Commun* **10,** 4907 (2019).

35. Stathopoulos, A. & Levine, M. Dorsal Gradient Networks in the Drosophila Embryo. **246,** 57–67 (2002).

36. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339,** 819–823 (2013).

37. Miyaoka, Y. *et al.* Isolation of single-base genome-edited human iPS cells without antibiotic selection. *Nat Meth* **11,** 291–293 (2014).

38. Mandegar, M. A. *et al.* CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Stem Cell* **18,** 541–553 (2016).

39. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169,** 930–933.e22 (2017).

40. Lee, J.-H. *et al.* A Robust Approach to Identifying Tissue-Specific Gene Expression Regulatory Variants Using Personalized Human Induced Pluripotent Stem Cells. *PLoS Genet* **5,** e1000718–15 (2009).

41. Byrne, S. M. & Church, G. M. Crispr-mediated Gene Targeting of Human Induced Pluripotent Stem Cells. *Current protocols in stem cell biology* **35,** 5A.8.1–22 (2015).

830 42. Peters, D. T., Cowan, C. A. & Musunuru, K. *Genome editing in human pluripotent stem cells*. (Harvard Stem Cell Institute, 2008).

832 43. Spencer, C. I. *et al.* Calcium Transients Closely Reflect Prolonged Action Potentials in iPSC Models of Inherited Cardiac Arrhythmia. *Stem Cell Reports* 1–13 (2014).
834 doi:10.1016/j.stemcr.2014.06.003

44. Hookway, T. A. *et al.* Phenotypic Variation Between Stromal Cells Differentially
836 Impacts Engineered Cardiac Tissue Function. *Tissue Engineering Part A* **25,** 773–785 (2019).

838 45. Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. *Image processing with ImageJ.* 36–42 (Biophotonics International, 2004).

840 46. Judge, L. M. *et al.* A BAG3 chaperone complex maintains cardiomyocyte function during proteotoxic stress. *JCI Insight* **2,** 83–18 (2017).

842 47. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* **37,** 38–44 (2018).

844 48. Duan, J. *et al.* Rational Reprogramming of Cellular States by Combinatorial Perturbation. *CellReports* **27,** 3486–3499.e6 (2019).

846 49. Fisher, R. A. On the‘ Probable Error’ of a Coefficient of Correlation Deduced from a Small Sample. *Metron* **1,** 3–32 (1921).

848 50. Waldron, L. *et al.* The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Dev Cell* **36,** 262–275 (2016).

850 51. Driegen, S. *et al.* A generic tool for biotinylation of tagged proteins in transgenic mice. *Transgenic Res* **14,** 477–482 (2005).

852 52. He, A. *et al.* Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat Commun* **5,** 4907 (2014).

854 53. Hota, S. K. *et al.* Dynamic BAF chromatin remodeling complex subunit inclusion promotes temporally distinct gene expression programs in cardiogenesis.
856 *Development* **146,** dev174086–21 (2019).
54. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of
858 insertions, deletions and gene fusions. *Genome Biol* **14,** R36–13 (2013).
55. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose
860 program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
862 56. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*
864 **26,** 139–140 (2009).
57. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and
866 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **57,** 289–300 (1995).
868

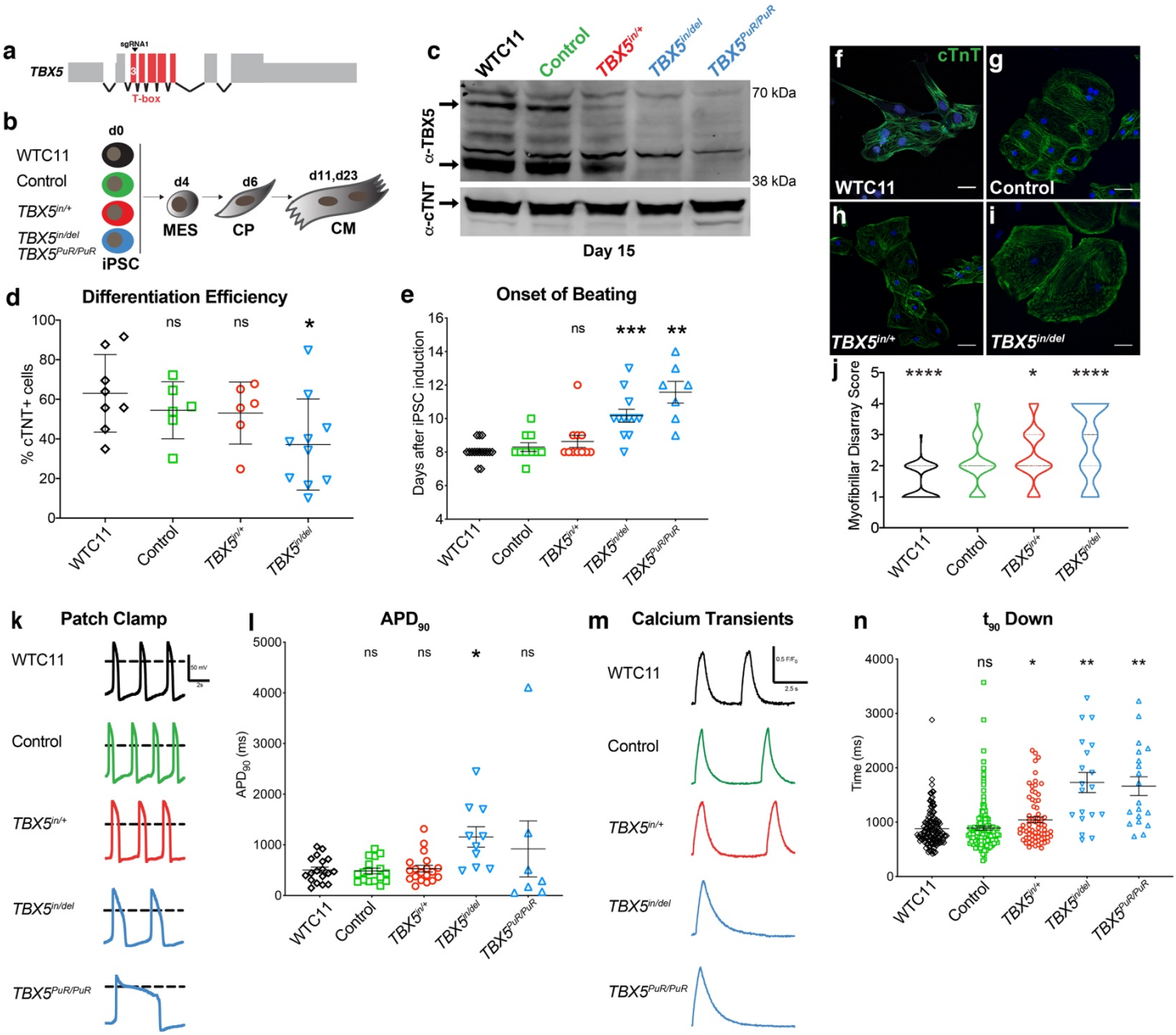


Figure 1. A human allelic series of *TBX5* mutants models features of congenital heart disease. a, Diagram of the human *TBX5* gene. Exons encoding the T-box domain of TBX5 are indicated in red. sgRNA1 was used to target exon 3 of *TBX5* by a CRISPR/Cas9 nuclease. **b**, Parental iPS cell line WTC11, control (CRISPR-treated, unmodified at exon 3 of *TBX5*) and targeted *TBX5* loss-of-function mutants (*TBX5*$^{in/+}$, *TBX5*$^{in/del}$, or *TBX5*$^{PuR/PuR}$) underwent directed differentiation to cardiomyocytes (CM) via mesoderm (MES) and cardiac precursor (CP) stages. **c**, TBX5 and cTNT protein expression for each *TBX5* genotype at day 15 (CMs). **d**, Differentiation efficiency by flow cytometry for cTNT^{+} cells (* p-value<0.05 by unpaired *t* test). **e**, Onset of beating (** p-value<0.01, *** p-value<0.001 by unpaired *t* test). **f-j**, Myofibrillar arrangement of cardiomyocytes (* p-value<0.05, **** p-value<0.0001 by Fisher's exact test). **k**,

884 Action potentials by patch clamp of single beating cells for each *TBX5* genotype. **l,** Action potential duration at 90% (APD_{90}) (* FDR<0.05). **m-n**, Traces of calcium transients from single
886 beating cells were analyzed, including time at 90% decay (t_{90} down) (* FDR<0.05, ** FDR<0.01). Error bars represent standard deviation (**d, e**) or standard error (**l, n**) of the mean.
888

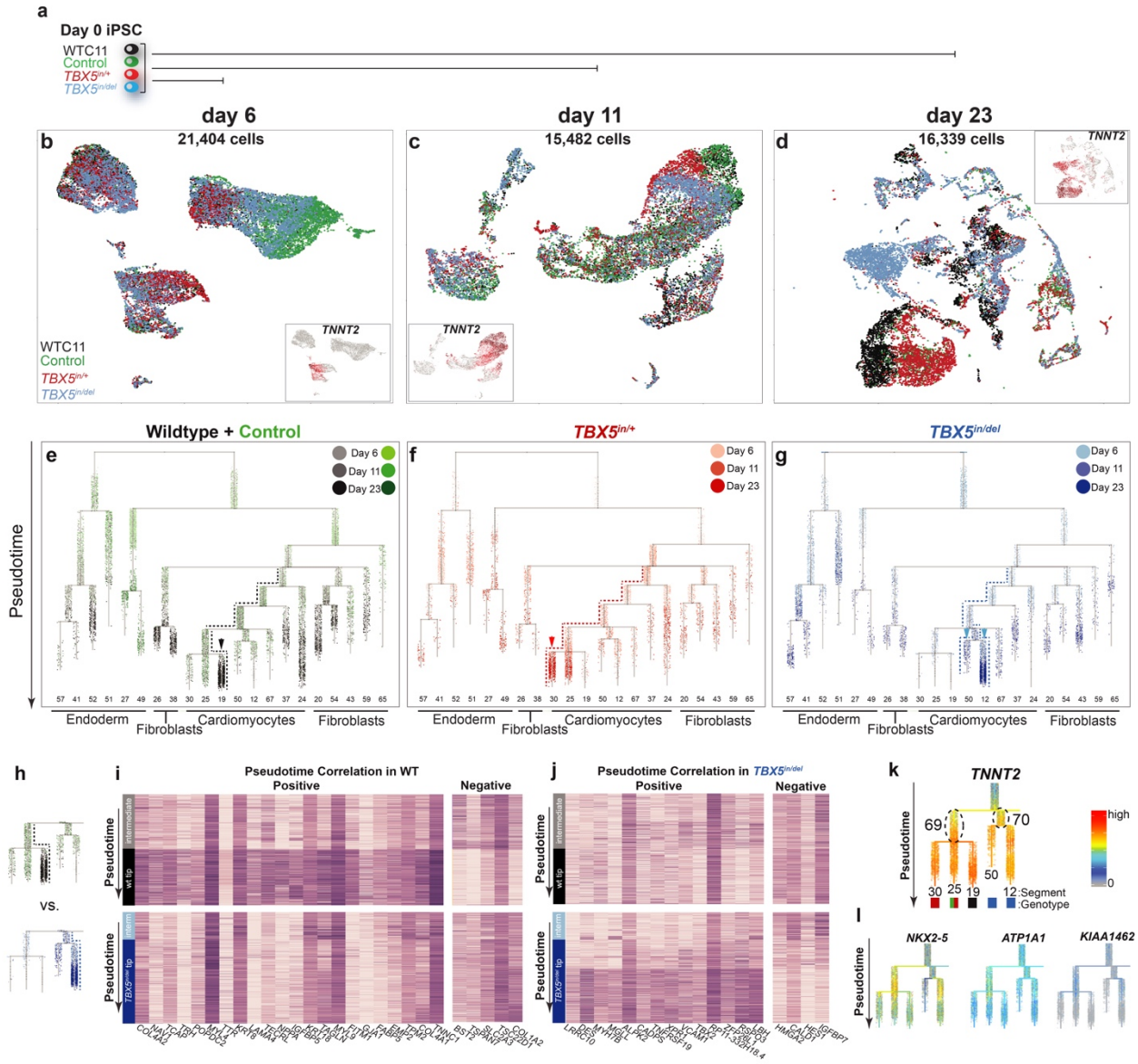


Figure 2. *TBX5* loss disturbs cell trajectories to cardiomyocyte fate. **a**, Cells of each *TBX5* genotype were harvested at specific stages during directed differentiation to cardiomyocytes (CM) for single cell RNAseq. **b-d,** UMAPs display cells at day 6, day 11 or day 23, colored by *TBX5* genotype. A feature plot of the expression of the CM structural gene, *TNNT2*, is shown in each inset. **e-g**, Reconstructed cell trajectories by the R package URD are shown as a dendrogram for each *TBX5* genotype from day 6, 11 and 23, from root (top; early pseudotime) to tips (bottom; late pseudotime). Each *TBX5* genotype is color-coded from light to dark, to indicate time point. Note the enrichment or depletion of cells from one genotype at certain branch points (arrowheads). Dashed lines show inferred path to CMs by *TBX5* genotype. **h**, Deduced paths to non-dividing CMs of WT (black dashed line) or *TBX5*$^{in/del}$ (blue dashed line) are shown, from intermediate (labeled interm) to tip. **i**, Heat maps show expression for each gene that displays a positive or negative correlation with pseudotime (|rho|≥0.4 and Z-score≥15 by difference in rho) in the WT path (above) and is altered in the *TBX5*$^{in/del}$ path (below). **j**, Heat maps show expression for each gene that displays a positive or negative correlation with pseudotime (|rho|≥0.4 and Z-score≥15) in the *TBX5*$^{in/del}$ path and is altered in the WT path

906 (below). **k**, Differential gene expression of inferred precursors for the cardiomyocyte branches (segments 69 vs. 70, dashed ovals) show several genes that display altered gene expression
908 (adj p-value<0.05 by Wilcoxon Rank Sum test) along the WT or *TBX5*$^{in/del}$ path. Colored blocks below represent the predominant *TBX5* genotypes in each tip. **l,** Feature plots show a delayed
910 onset of expression during pseudotime for *NKX2-5,* in addition to *ATP1A1* and *KIAA1462*.

912

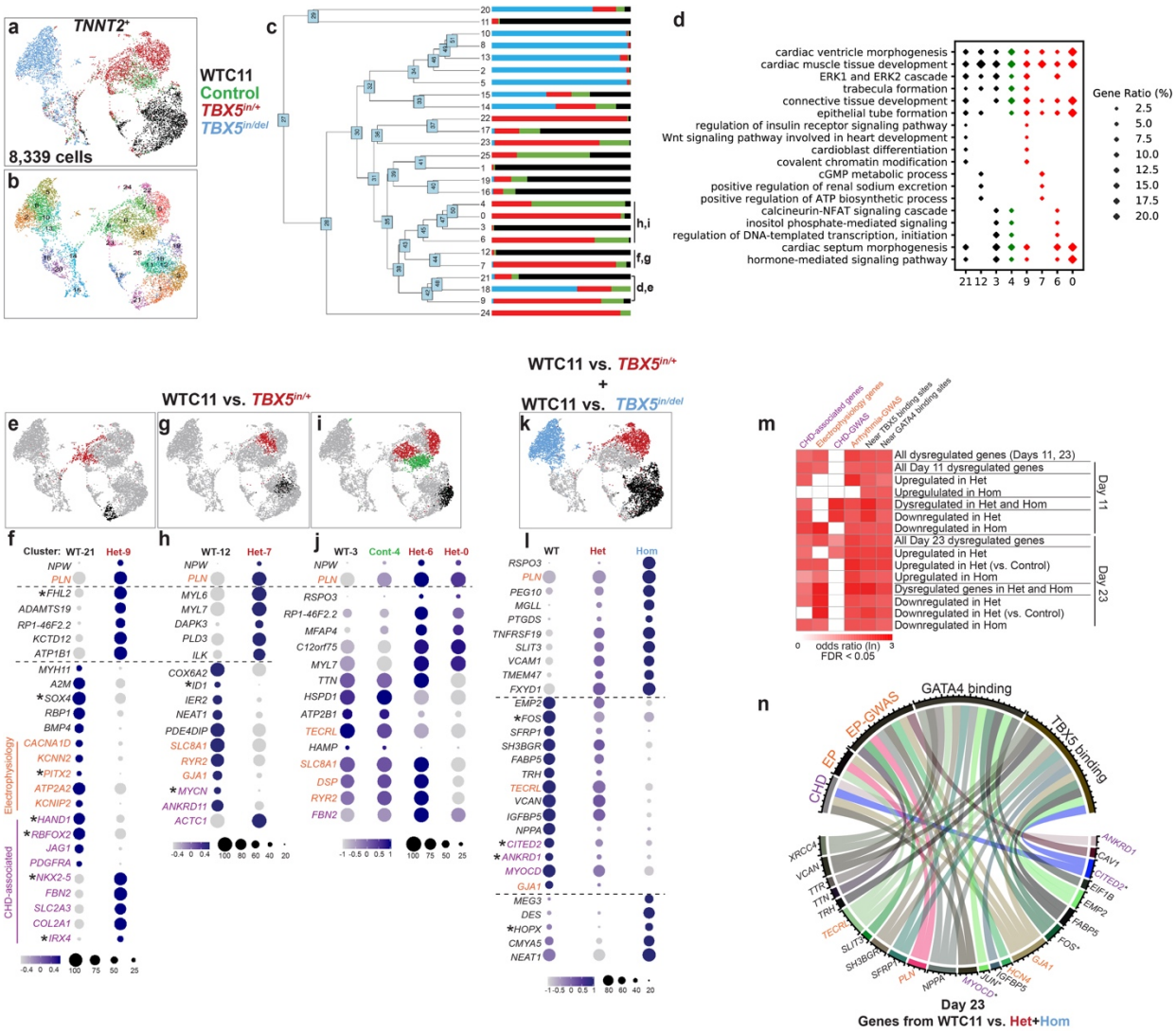


Figure 3. Subsets of cardiomyocytes respond discretely by quantitative transcriptional perturbations to reduced TBX5 dosage. **a-c,** *TNNT2*$^{+}$ clusters from day 23 were re-clustered in Seurat. UMAP shows cells colored by *TBX5* genotype (**a**) or by cluster identity (**b**). **c**, Phylogenetic tree shows the relatedness of the 'average' cell per cluster among clusters. The proportion of cells in each cluster are colored by *TBX5* genotype. Pairs of related clusters between *TBX5* genotypes were selected for differential gene tests. **d,** Gene ontology for enriched biological processes (p-value<0.05 and q-value <0.05) in cell clusters, which are selected for differential gene expression (logfc>0.25 and adj p-value<0.05 by Wilcoxon Rank Sum test), are shown. **e, g, i**, UMAPs highlight clusters used for pair-wise comparisons for differential gene tests in corresponding dot plots below. **f, h**, Dot plots of top differentially expressed genes between wildtype-enriched cluster 21 and *TBX5*$^{in/+}$-enriched cluster 9 (**f**), or between wildtype-enriched cluster 12 and *TBX5*$^{in/+}$-enriched cluster 7 (**h**) (Supplementary Table 2). Each dot size corresponds to the percentage of cells expressing the gene in the cluster, and the color intensity represents scaled expression values in the cluster. Electrophysiology (EP)-related (orange) and congenital heart disease (CHD)-associated (purple) genes are highlighted. Asterisks denote transcription factors. Significance was determined by Wilcoxon Rank Sum test (adj p-value<0.05). **j**, The most differentially expressed genes in *TBX5*$^{in/+}$-enriched clusters 0 and 6, when compared to both wildtype (WT)-enriched cluster 3 or control-enriched cluster 4,

934 are shown. **k**, **l,** UMAP highlights clusters for comparison (**k**) in dot plots of genes that are most differentially expressed in both *TBX5*$^{in/+}$- and *TBX5*$^{in/del}$-enriched clusters, when compared to
936 WT-enriched clusters (**l**). **m**, Heat map displays significant correlations (adj p-value<0.05 by Benjamini-Hochberg multiple corrections test) among comparisons of candidate lists for CHD-
938 associated genes, EP-related genes, genome-wide association studies (GWAS) for CHDs or arrhythmias (Supplementary Table 4), and TBX5-dependent gene sets. **n**, Circular plot shows
940 selected genes by category for WT-enriched vs. *TBX5*$^{in/+}$-enriched and WT-enriched vs. *TBX5*$^{in/del}$-enriched clusters at day 23.
942

944

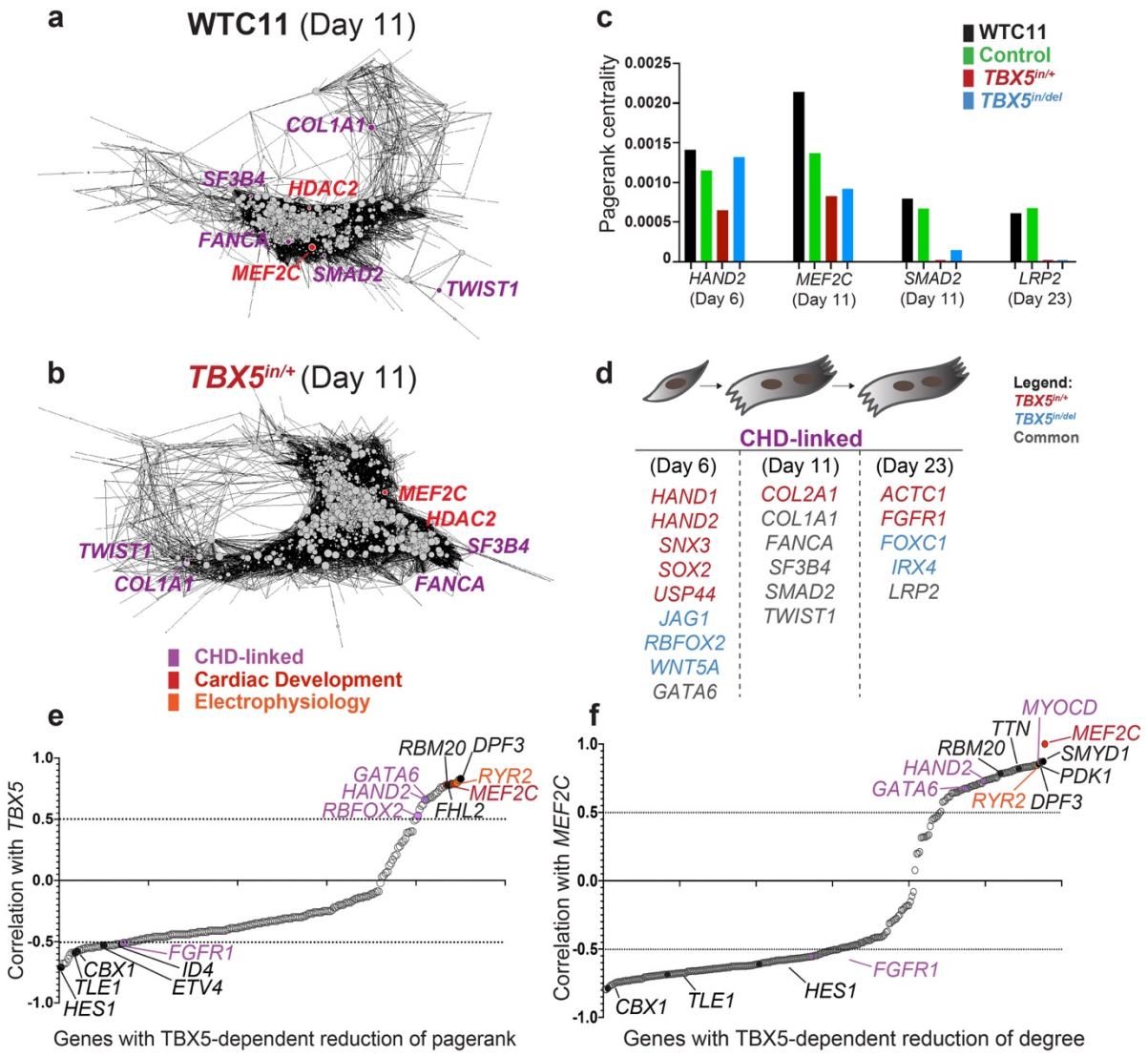


Figure 4. TBX5 dosage preserves cardiomyocyte network stability. a, b, Gene regulatory networks (GRNs) for day 11 cardiomyocytes of WTC11 (**a**) or *TBX5*$^{in/+}$ (**b**) are shown. Nodes of CHD-related (purple), heart development (red) or electrophysiology (orange) genes are shown. The size of each node represents its quantitative importance, based on pagerank centrality. Note the absence of *SMAD2* and the reduced centrality of *MEF2C* (smaller circle) in the *TBX5*$^{in/+}$ network, compared to WTC11. **c**, Pagerank centrality for significantly altered (top 5% cutoff) nodes of CHD-associated or heart development genes at specific time points are shown. **d,** Twenty CHD-associated genes display a reduction in pagerank (top 5% cutoff) in at least one mutant *TBX5* genotype at any stage. This indicates enrichment of CHD-associated genes in TBX5 dosage-sensitive networks ($p<2.2e-5$ by hypergeometric test). **e,** TBX5-dependent genes with a reduction of pagerank are correlated (correlation >0.5), anti-correlated (correlation <-0.5), or indeterminate (0.5<correlation<-0.5) with *TBX5* expression in *TNNT2*$^{+}$ cells. **f,** Correlations with *MEF2C* and TBX5-dependent genes with a reduction of degree in *TNNT2*$^{+}$ cells are plotted. Additional data can be found in Supplementary Table 3.

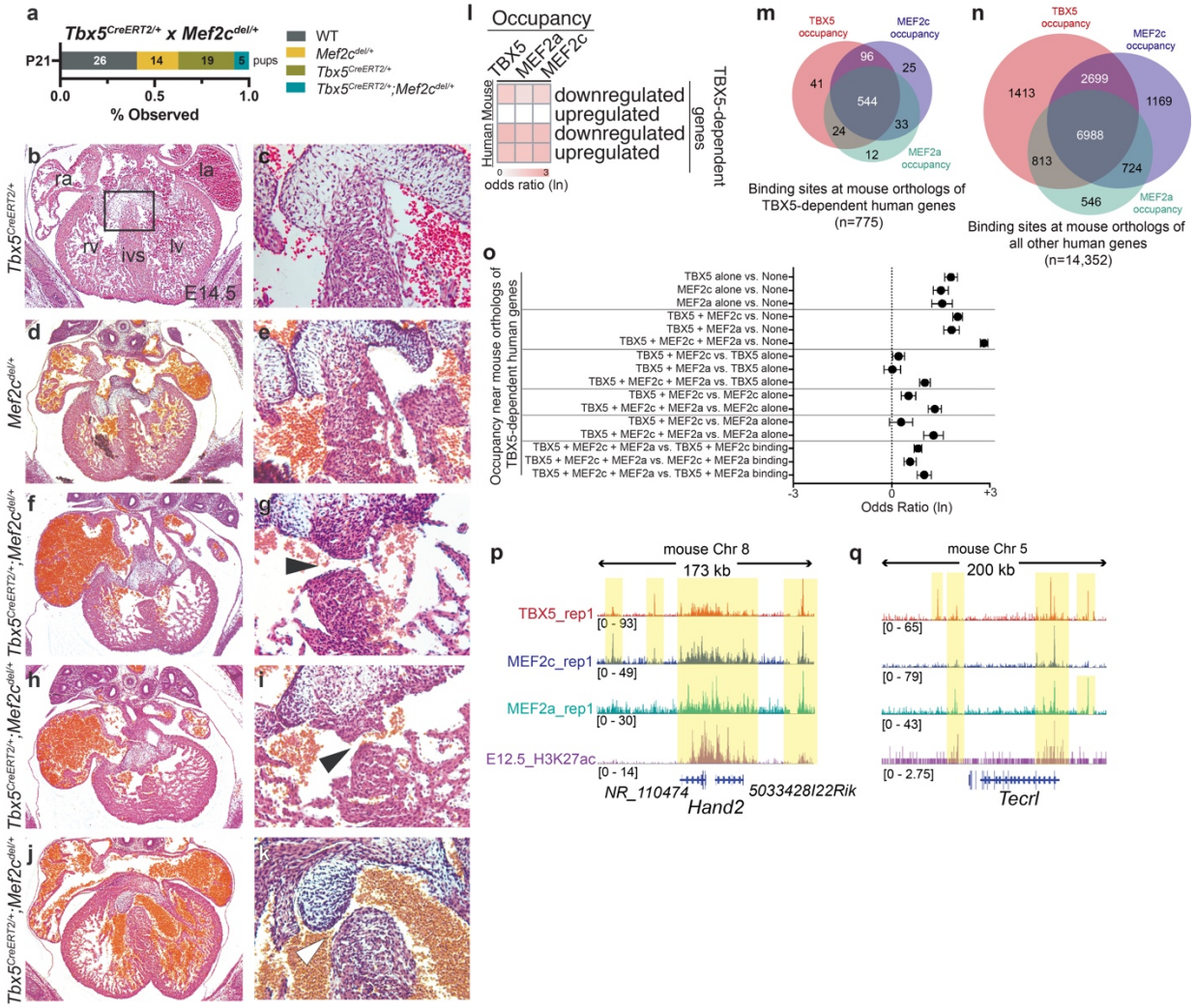
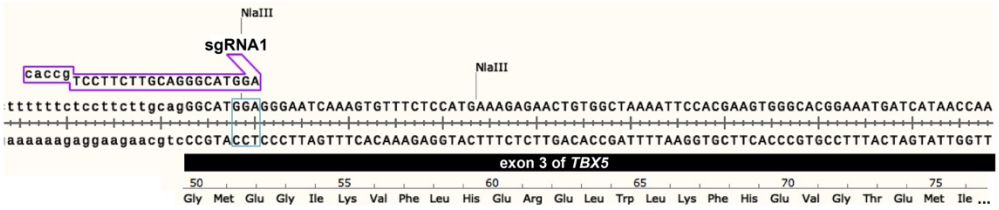


Figure 5. *Tbx5* and *Mef2c* cooperate in heart development. a, Pups at post-natal day 21 (P21) from matings of *Tbx5*$^{CreERT2/+}$ X *Mef2c*$^{del/+}$ were genotyped. Expected Mendelian ratios were not observed for *Tbx5*$^{CreERT2/+}$;*Mef2c*$^{del/+}$ at P21. **b, d, f, h, j**, Transverse sections of hearts at embryonic day 14.5 (E14.5) from each genotype are shown. **c, e, g, i, k**. Magnified views of the interventricular septa are shown. Note muscular VSDs (black arrowheads in **g, i**), a subaortic membranous VSD (**k**, white arrowhead) and dilated blood-filled atria in the *Mef2c*$^{+/-}$;*Tbx5*$^{CreERT2/+}$ embryos (**f, h, j**). **l,** Heat map indicates odds ratios (FDR<0.05) of TBX5, MEF2a or MEF2c occupancy near mouse or human TBX5-dependent genes (Supplementary Table 6). **m, n,** Venn diagrams display the overlap of TBX5, MEF2a or MEF2c occupancy near mouse orthologs of human TBX5-dependent or -independent genes, respectively. **o,** Odds ratio (FDR<0.05) of combinations of TBX5, MEF2a or MEF2c occupancy near mouse orthologs of TBX5-dependent human genes (Supplementary Table 6). **p, q,** Browser tracks for ChIP-seq data from E12.5 hearts for TBX5, MEF2c, MEF2a and H3K27ac near conserved TBX5-dependent genes, *Hand2* (**p**) or *Tecrl* (**q**). Yellow bands of shading indicate co-occupancy.

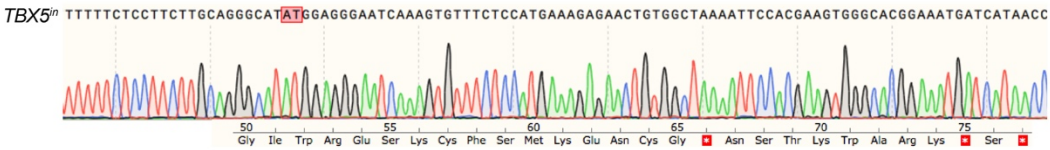
a

	Genotype	Translation
WTC11	+/+	
+ Cas9 Control	+/+	
+ Cas9 *TBX5*$^{in/+}$	ins 2bp/+	p.M51fs*16/+
+ Cas9 *TBX5*$^{in/del}$	ins 1bp / del 8bp	p.M51fs*10/p.M51fs*7
+ Cas9 *TBX5*$^{PuR/PuR}$	ins PGK-PuroR/ins PGK-PuroR	

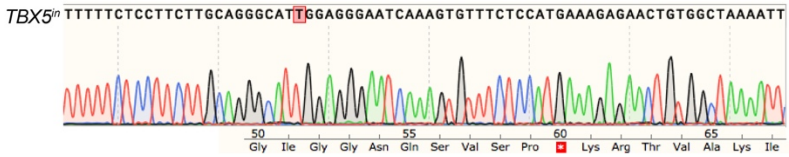
b



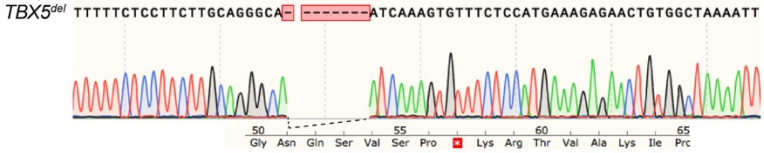
c 2bp insertion allele for *TBX5*$^{in/+}$



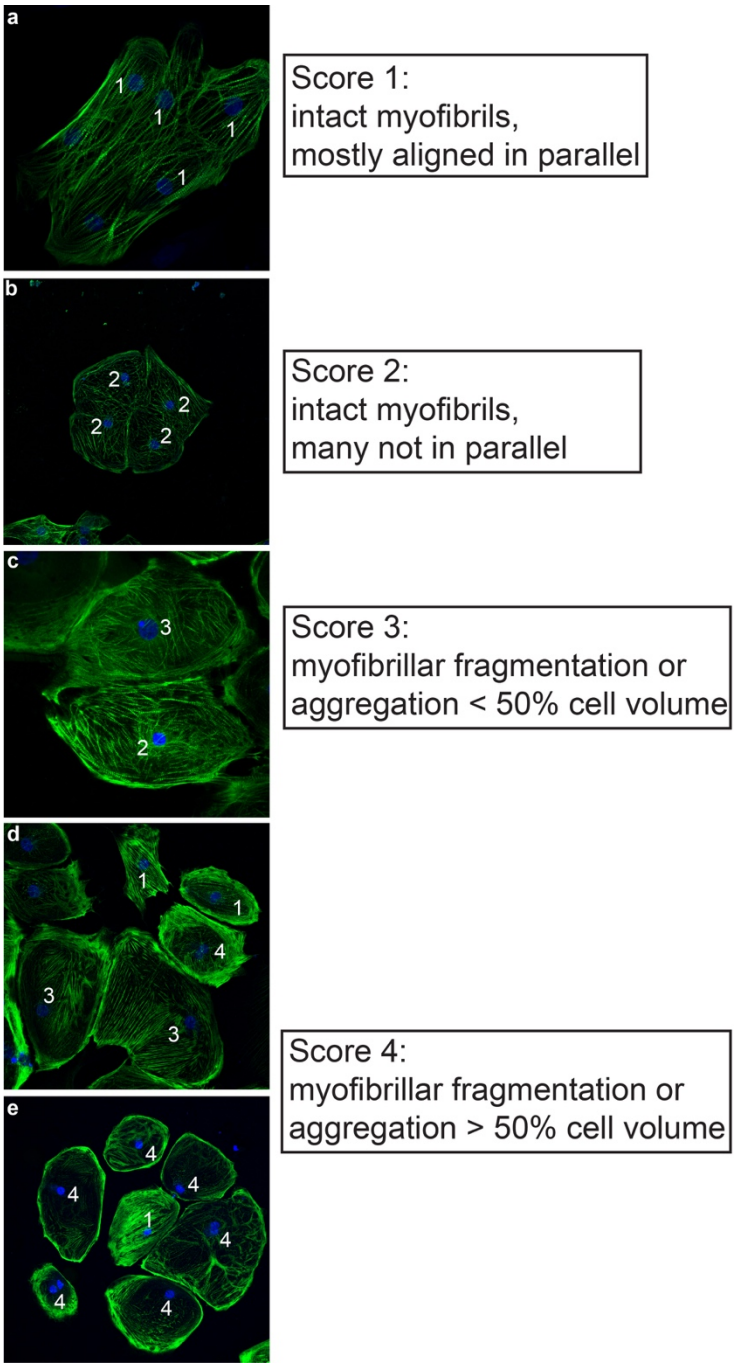
d 1bp insertion for *TBX5*$^{in/del}$



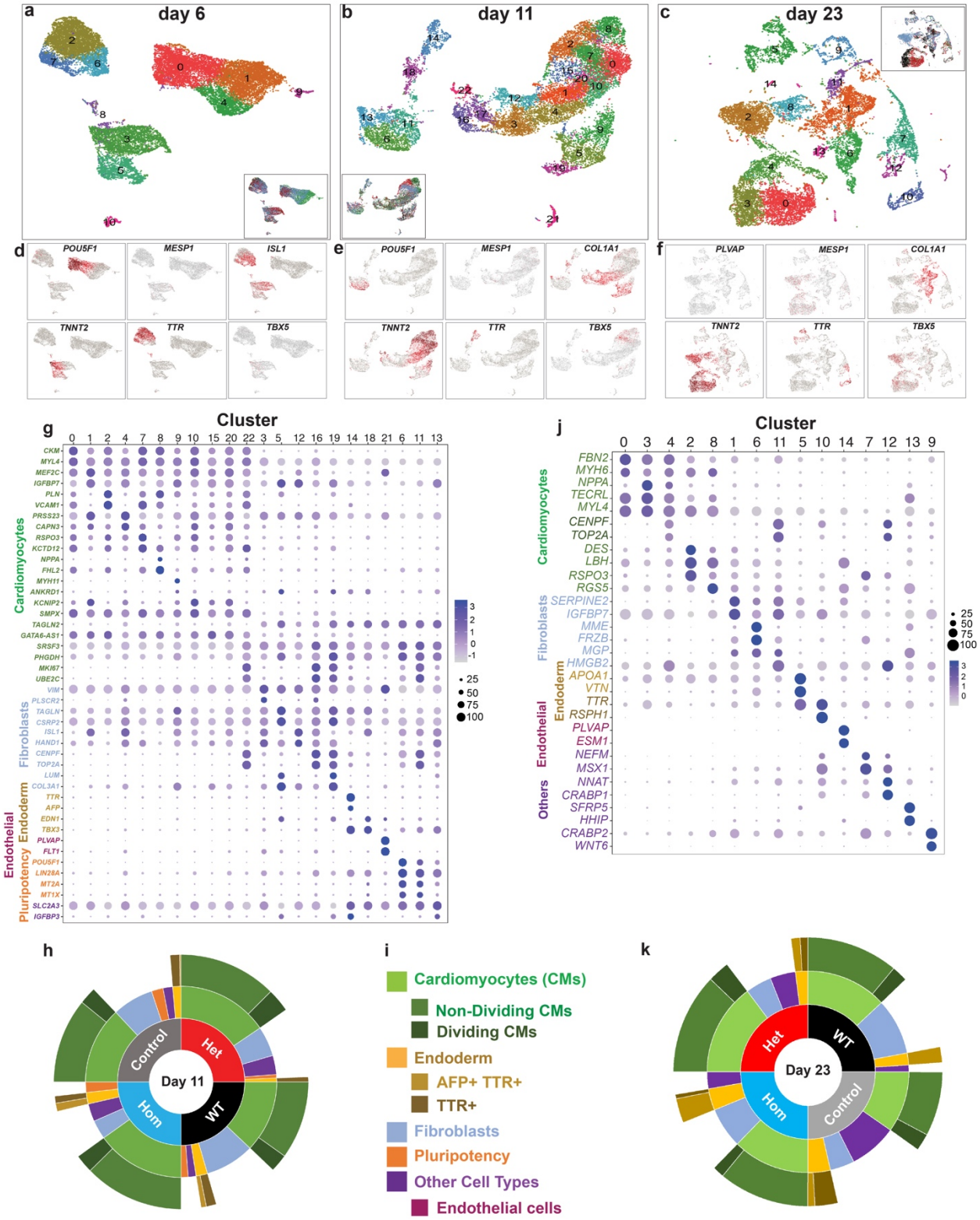
e 8bp deletion for *TBX5*$^{in/del}$



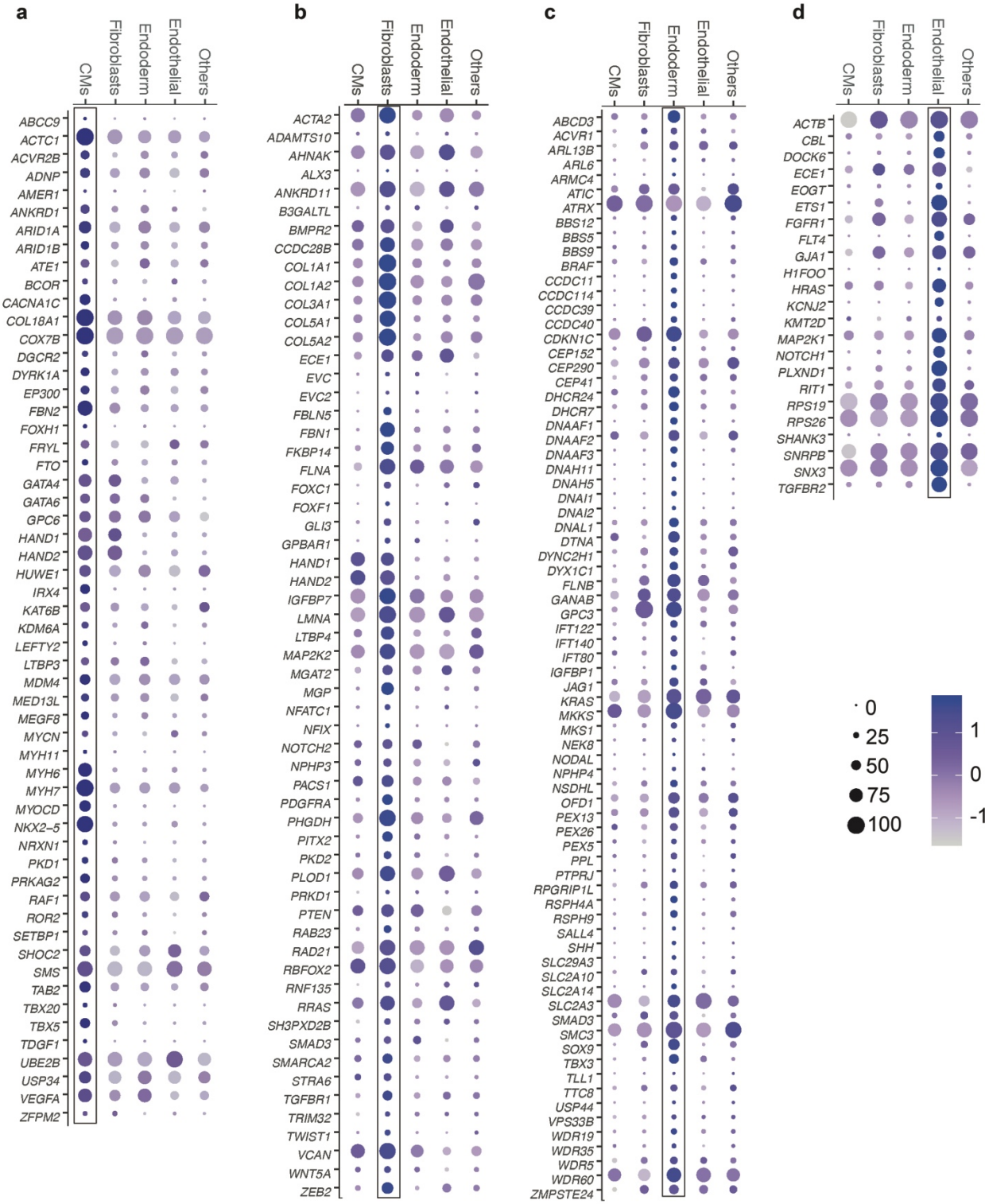
Extended Data 1. Genotype information for targeted WTC11-derived iPS cells mutant for *TBX5*. a, Table shows genotypes of WTC11-derived iPS cell lines that were targeted for *TBX5* at exon 3. Predicted translation for each *TBX5* genotype is indicated. **b**, Sequence of the exon 3 of *TBX5* is shown, along with the sgRNA1 location. The PAM site is boxed in blue. Loss of the NlaIII site at the PAM site was used in initial screening for mutant iPS cell clones by PCR. The encoded wildtype protein sequence includes the start of the T-box domain. **c**, Sequence and chromatogram for the 2bp insertion of the mutant allele for *TBX5*$^{in/+}$ predicts a premature truncation, as indicated by a stop codon (white asterisk in red box) in the frame-shifted protein sequence. **d**, **e**, Sequence and chromatogram for the 1 bp insertion, or 8 bp deletion, respectively, of the mutant allele for *TBX5*$^{in/del}$, along with corresponding protein sequences, are shown.



Extended Data 2. Scoring criteria for myofibrillary disarray. Criteria for scores 1-5 are indicated, and representative images for each score are shown.


a
day 6
b
day 11
c
day 23
d
POU5F1
MESP1
ISL1
TNNT2
TTR
TBX5
e
POU5F1
MESP1
COL1A1
TNNT2
TTR
TBX5
f
PLVAP
MESP1
COL1A1
TNNT2
TTR
TBX5
g
Cluster
Cardiomyocytes
Fibroblasts
Endothelial
Endoderm
Pluripotency
j
Cluster
Cardiomyocytes
Fibroblasts
Endoderm
Endothelial
Others
h
Control
Het
Hom
WT
Day 11
i
Cardiomyocytes (CMs)
Non-Dividing CMs
Dividing CMs
Endoderm
AFP+ TTR+
TTR+
Fibroblasts
Pluripotency
Other Cell Types
Endothelial cells
k
Het
WT
Hom
Control
Day 23

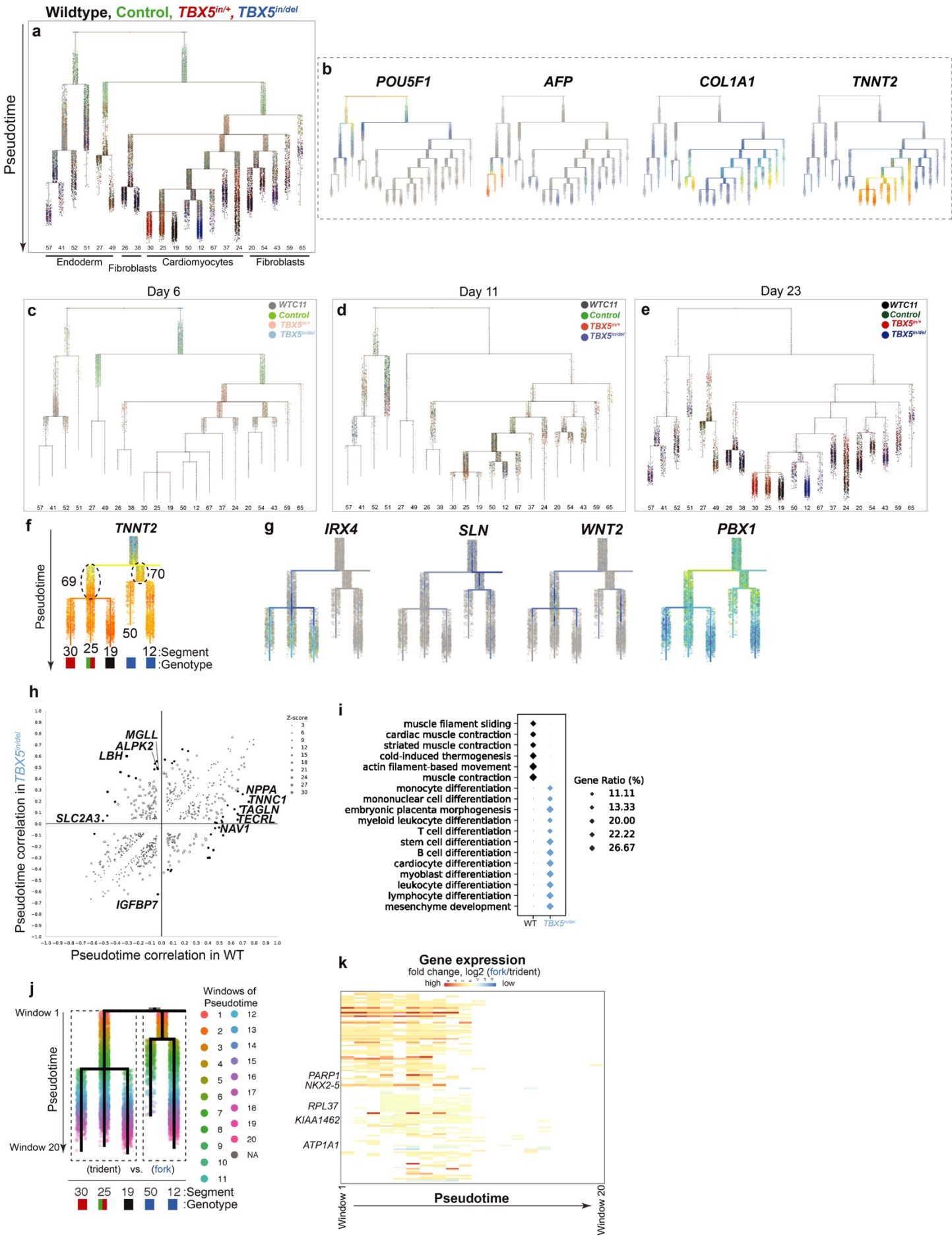
1006 **Extended Data 3. Human cardiac differentiation is sensitive to reduced TBX5 dosage. a-c**,
UMAPs display cells at day 6, day 11 and day 23, respectively, by Louvain clustering. Insets
1008 show UMAPs colored by *TBX5* genotype. **d-f**, Feature plots in UMAP space demonstrate
expression of selected marker genes, which represent major cell types at each timepoint. **g, j**,
1010 Dot plots display two of the top marker genes that are expressed in each cluster at day 11 or
day 23 (adj p-value<0.05 by Wilcoxon Rank Sum test), and define at least five major human cell
1012 types. The size of the dot corresponds to the percentage of cells expressing the gene in the
cluster, and the color intensity represents the level of gene expression in the cluster. **h, i, k**,
1014 Sunburst plots indicate the proportion of at least five major cell types, based on top marker
genes, by *TBX5* genotype at day 11 or day 23.
1016



Extended Data 4. Many CHD-associated genes are enriched among several human cell types. a-d, Dot plots show the expression of congenital heart disease (CHD)-associated genes that are significantly enriched in cardiomyocytes (**a**), fibroblasts (**b**), endodermal cells (**c**) or

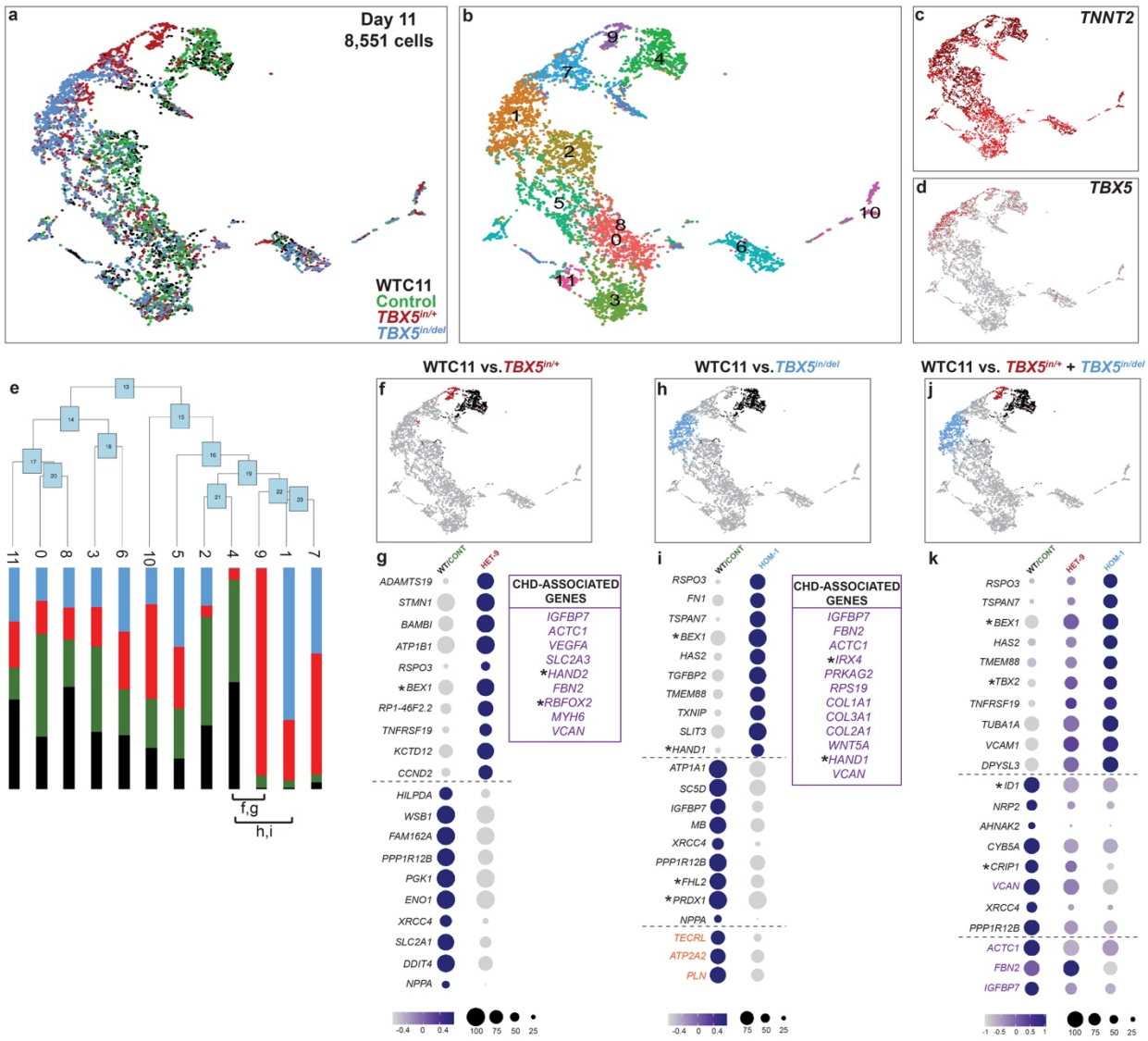
1024 endothelial cells (**d**). Significance was determined by Wilcoxon Rank Sum test (adj p-value<0.05).

1026

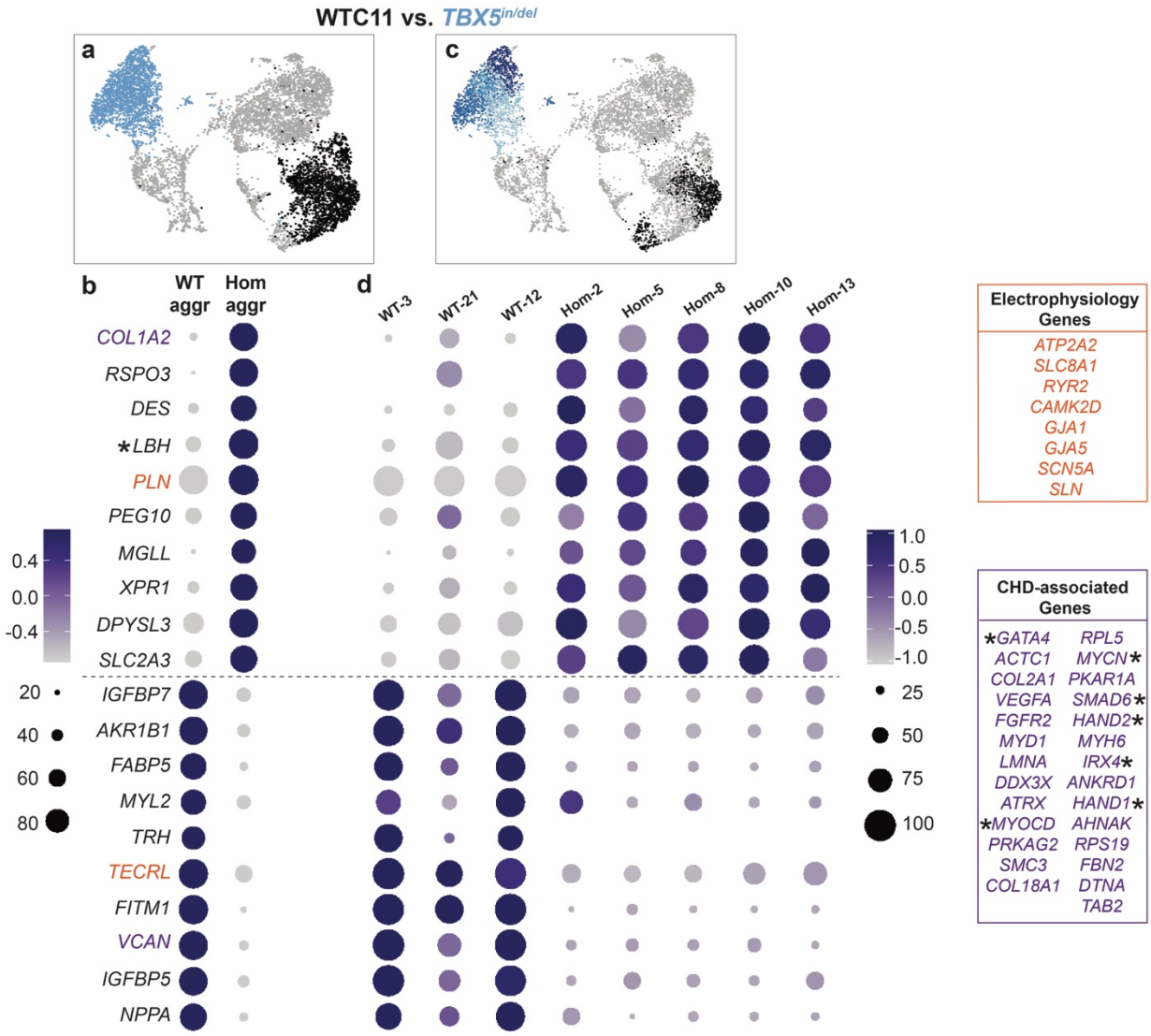

Wildtype, Control, TBX5in/+, TBX5in/del
a
Pseudotime
Endoderm
Fibroblasts
Cardiomyocytes
Fibroblasts
b
POU5F1
AFP
COL1A1
TNNT2
Day 6
Day 11
Day 23
c
d
e
WTC11
Control
TBX5in/+
TBX5in/del
f
TNNT2
Pseudotime
69
70
50
30
25
19
12
:Segment
:Genotype
g
IRX4
SLN
WNT2
PBX1
h
Pseudotime correlation in TBX5in/del
Pseudotime correlation in WT
MGLL
ALPK2
LBH
SLC2A3
NPPA
TNNC1
TAGLN
TECRL
NAV1
IGFBP7
Z-score
i
muscle filament sliding
cardiac muscle contraction
striated muscle contraction
cold-induced thermogenesis
actin filament-based movement
muscle contraction
monocyte differentiation
mononuclear cell differentiation
embryonic placenta morphogenesis
myeloid leukocyte differentiation
T cell differentiation
stem cell differentiation
B cell differentiation
cardiocyte differentiation
myoblast differentiation
leukocyte differentiation
lymphocyte differentiation
mesenchyme development
WT
TBX5in/del
Gene Ratio (%)
11.11
13.33
20.00
22.22
26.67
j
Window 1
Window 20
Pseudotime
(trident) vs. (fork)
Windows of Pseudotime
NA
30
25
19
50
12
:Segment
:Genotype
k
Gene expression
fold change, log2 (fork/trident)
high
low
PARP1
NKX2-5
RPL37
KIAA1462
ATP1A1
Window 1
Pseudotime
Window 20

1028

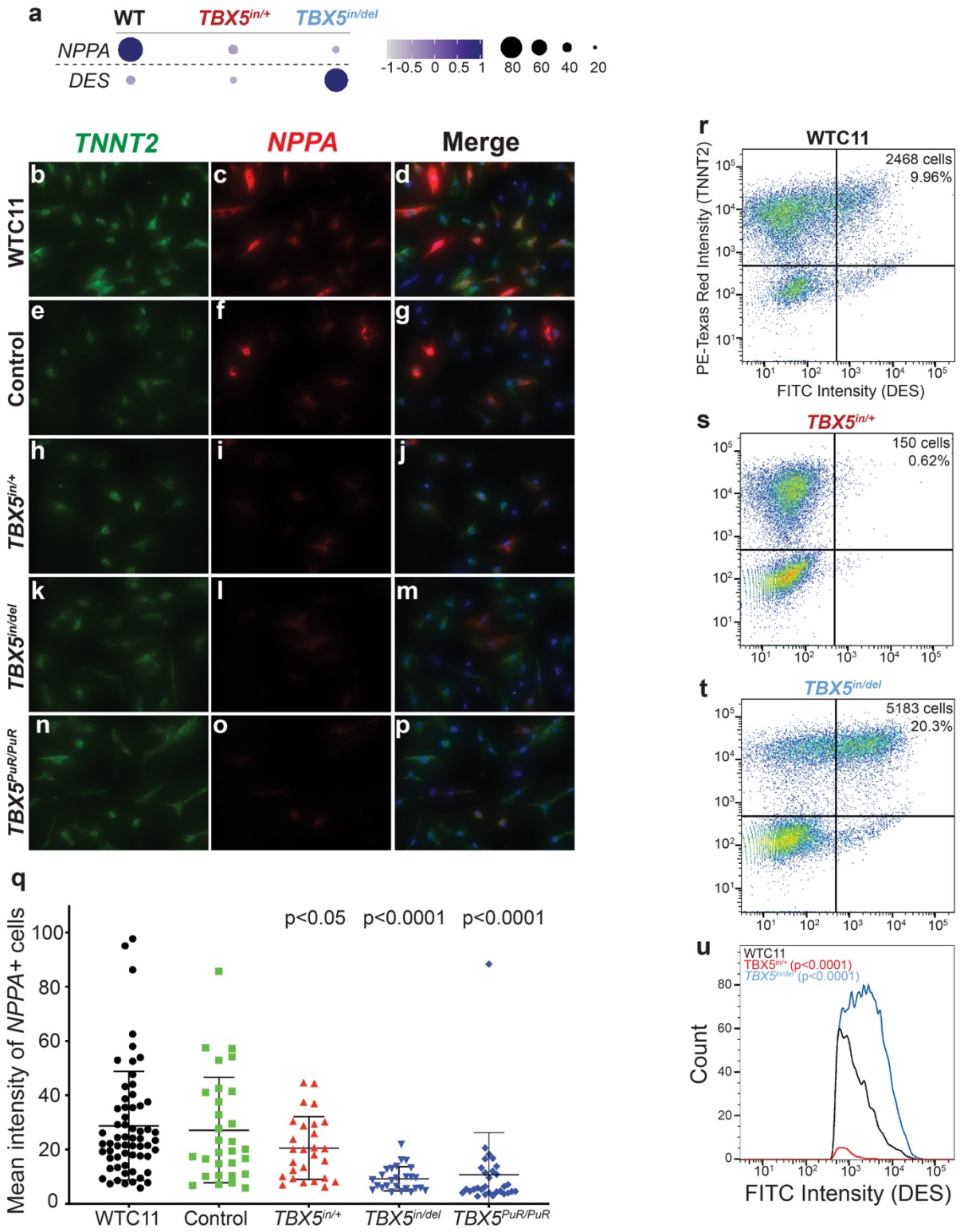
Extended Data 5. Pseudotime analysis of cell trajectory reconstructions by genotype, time point and cell type. a, All cells for all *TBX5* genotypes and time points are shown for directed differentiation to cardiomyocytes in a single plot. Pseudotime is displayed from root (top) to tips (bottom). **b,** Genes, which mark each major cell type (pluripotent cell, endoderm, fibroblast and cardiomyocyte) during the directed differentiation, are shown. **c-e**, Cells from all *TBX5* genotypes are shown by time point. **f,** Differential gene expression between inferred precursors (dashed ovals, segments 69 vs. 70) of the WT or *TBX5*$^{in/del}$ path was evaluated. **g**, Delayed activation of *WNT2*, *PBX1* and *NKX2-5* in the *TBX5*$^{in/del}$ path (to segments 50 and 12) diverged from the path of other genotypes (to segments 20, 25, 19). *IRX4* was absent in the *TBX5*$^{in/del}$ path, and instead, *SLN* was enriched. Significance was determined by Wilcoxon Rank Sum test (adj p-value<0.05). **h**, Scatter plot displays two axes of Spearman rho correlations, which measure a potential relationship between gene expression and pseudotime for WT or *TBX5*$^{in/del}$ cell. Each white circle represents a gene that has a significant rho correlation with pseudotime (p-value<0.05 by two-sided *t* test). Size of the circle represents a Z-score difference for rho correlation of a gene to pseudotime for WT versus *TBX5*$^{in/del}$ path. As the size of the circle increases, there is a larger difference in how a gene is correlated with pseudotime between the two genotypes. Black circles represent genes with a moderate correlation or better with pseudotime in either path (|rho|≥0.4) and (Z-score≥15), when the opposing genotype has no significant correlation to pseudotime (p-value≥0.05 by two-sided *t* test). **i**, Gene ontology (GO) terms for enriched biological processes (p-value<0.05 and q-value<0.05), which are based on genes that positively correlate with pseudotime (|rho|≥0.4 and Z-score≥15 by difference in rho) in the WT- or the *TBX5*$^{in/del}$ path, are shown. Black (WT) or blue (*TBX5*$^{in/del}$) diamonds show gene ratios. **j,** Deduced paths for WT/control/ *TBX5*$^{in/+}$(trident) or *TBX5*$^{in/del}$ (fork) to cardiomyocytes were divided into windows (1-20) along pseudotime for comparison. **k**, Heat map shows fold change for genes in a cluster that includes *NKX2-5,* which was significantly different after correction (adj p-value<0.05 by Bonferroni-Holm test) in windows 2 through 8, along with genes of a similar pattern, including *PARP1, RPL37, KIAA1462,* and *ATP1A1* (adj p-value<0.05 by Bonferroni-Holm test).



Extended Data 6. Evaluation of reduced TBX5 dosage by gene expression in cardiomyocytes at day 11. **a, b**, UMAPs of *TNNT2*$^{+}$ cells from day 11 of differentiation are colored by *TBX5* genotype or by cluster identity, respectively. **c, d**, Feature plots by UMAP display expression of *TNNT2* or *TBX5*. **e,** Phylogenetic tree depicts cluster relatedness. Vertical bars show the proportion of cells by color of each *TBX5* genotype. **f, h,** UMAPs highlight clusters chosen for pair-wise differential gene expression comparisons. **g, i**, Dot plots show top differentially expressed genes between WT/control-enriched and *TBX5*$^{in/+}$-enriched clusters, or WT/control-enriched and *TBX5*$^{in/del}$-enriched clusters, respectively. Electrophysiology-related (orange) and congenital heart disease (CHD)-associated (purple) genes are highlighted. Significance was determined by Wilcoxon Rank Sum test (adj p-value<0.05). **j**, UMAP demonstrates pair-wise comparisons between WT/control-enriched and *TBX5*$^{in/+}$-enriched clusters, and WT/control-enriched and *TBX5*$^{in/+}$-enriched clusters. **k**, Common genes that are differentially expressed in both *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$, in each comparison with WT/control are shown.

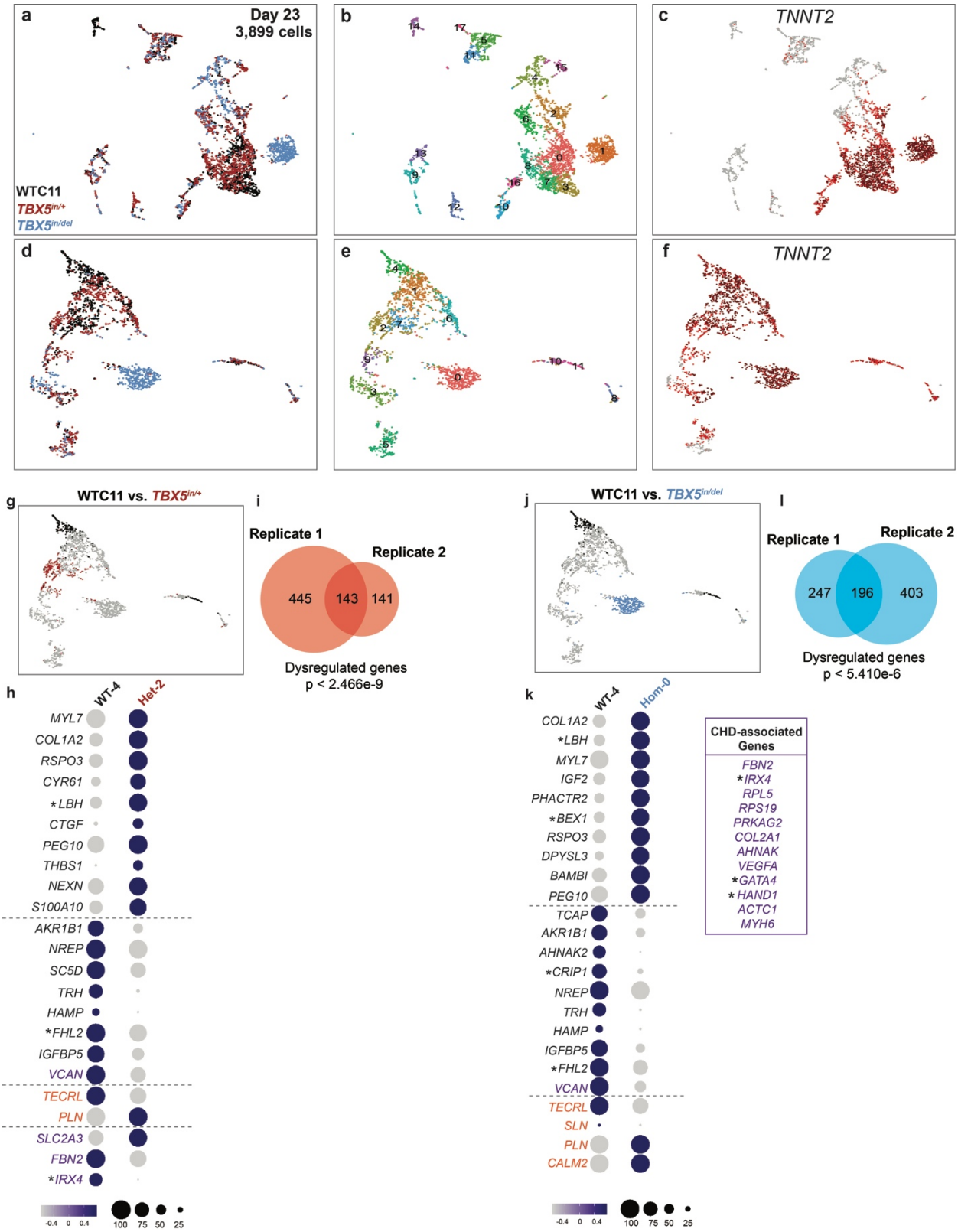


Extended Data 7. Loss of *TBX5* by gene expression in individual cardiomyocytes at day 23. a, UMAP of day 23 cardiomyocytes displays a wildtype-enriched (black) and *TBX5*$^{in/del}$-enriched (blue) cluster, at a low Louvain resolution in Seurat, for comparison between clusters of *TBX5* genotypes. **b**, Dot plots showing top differentially expressed genes between aggregate WT-enriched and *TBX5*$^{in/del}$-enriched clusters. Transcriptional regulators are denoted by asterisks. Electrophysiology-related (orange) and congenital heart disease (CHD)-associated (purple) are highlighted. Significance was determined by Wilcoxon Rank Sum test (adj p-value<0.05). **c**, UMAP shows clusters at a higher resolution, with three wildtype-enriched (black) sub-clusters and five *TBX5*$^{in/del}$-enriched (shades of blue) sub-clusters. **d,** Heterogeneity of gene expression among labeled sub-clusters are shown.



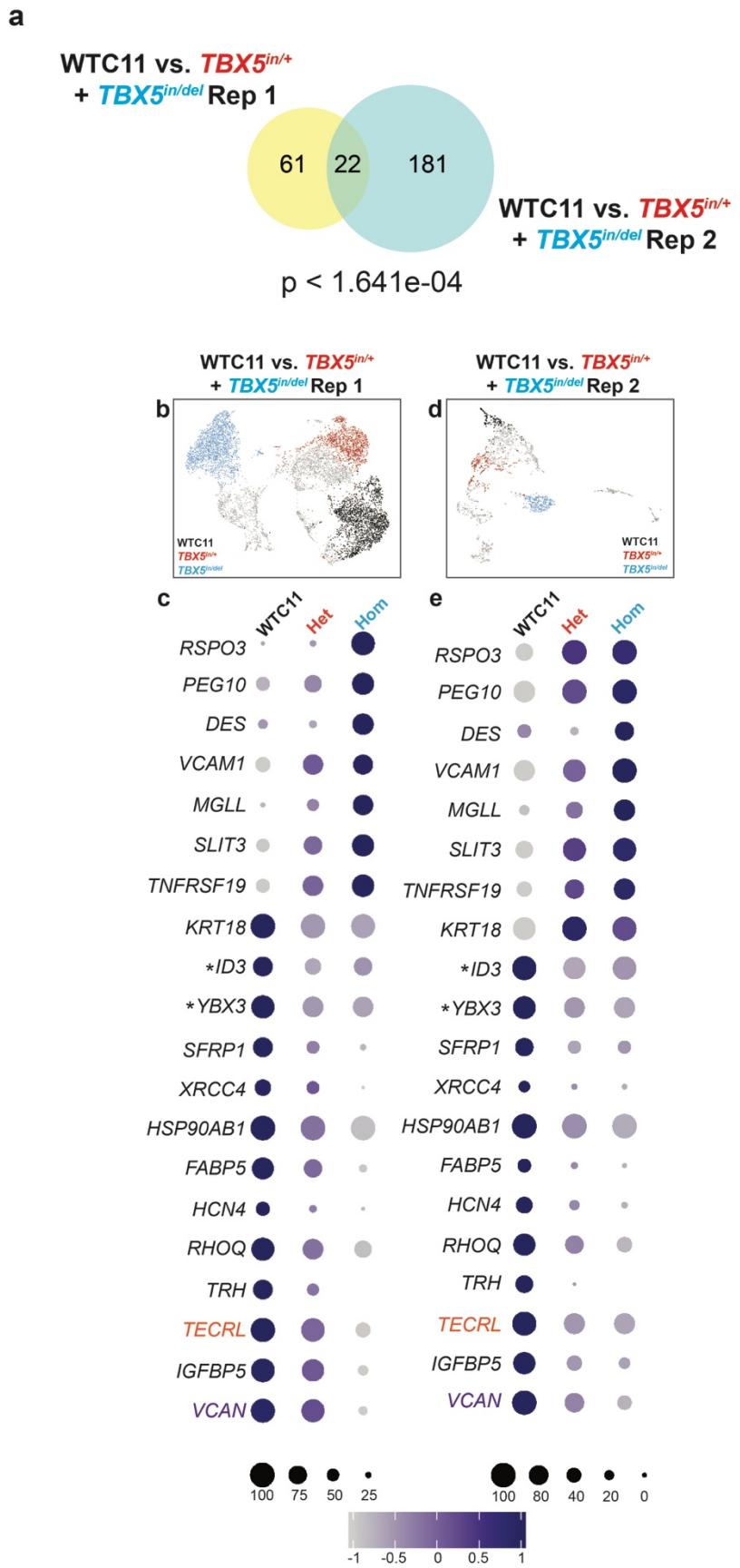
Extended Data 8. Single cell analysis of TBX5-dependent genes by orthogonal assays. a, Dot plots show expression of *NPPA* and *DES* in WTC11 (WT)-, *TBX5*$^{in/+}$- and *TBX5*$^{in/del}$ -enriched clusters at day 23. **b-p,** Fluorescence *in situ* hybridization is visualized for *TNNT2*

(green) or *NPPA* (red) in day 23 cardiomyocytes, from (**b-d**) WTC11, (**e-g**) control, (**h-j**)
1096 *TBX5*$^{in/+}$, (**k-m**) *TBX5*$^{in/del}$ and (**n-p**) *TBX5*$^{PuR/PuR}$ cells. **q,** Graph displays mean intensity of *NPPA* signal of individual double-positive *TNNT2*$^{+}$/*NPPA*$^{+}$ cells by *TBX5* genotype. Significance of p-
1098 values were calculated by unpaired *t* test. **r-t,** Pseudocolor plots showing cTNT^{+} and DES^{+} double positive cells in (**r**) wildtype, (**s**) *TBX5*$^{in/+}$ or (**t**) *TBX5*$^{in/del}$ at day 23. **u,** Histogram of FITC
1100 intensity of DES^{+} cells from wildtype, *TBX5*$^{in/+}$ or *TBX5*$^{in/del}$ cells is shown (****p-value<1E-4 by Chi-Square test).
1102


a
Day 23
3,899 cells
WTC11
TBX5in/+
TBX5in/del
b
c
TNNT2
d
e
f
TNNT2
g
WTC11 vs. TBX5in/+
i
Replicate 1
Replicate 2
445
143
141
Dysregulated genes
p < 2.466e-9
j
WTC11 vs. TBX5in/del
l
Replicate 1
Replicate 2
247
196
403
Dysregulated genes
p < 5.410e-6
h
WT-4
Het-2
MYL7
COL1A2
RSPO3
CYR61
*LBH
CTGF
PEG10
THBS1
NEXN
S100A10
AKR1B1
NREP
SC5D
TRH
HAMP
*FHL2
IGFBP5
VCAN
TECRL
PLN
SLC2A3
FBN2
*IRX4
-0.4 0 0.4
100 75 50 25
k
WT-4
Hom-0
COL1A2
*LBH
MYL7
IGF2
PHACTR2
*BEX1
RSPO3
DPYSL3
BAMBI
PEG10
TCAP
AKR1B1
AHNAK2
*CRIP1
NREP
TRH
HAMP
IGFBP5
*FHL2
VCAN
TECRL
SLN
PLN
CALM2
-0.4 0 0.4
100 75 50 25
CHD-associated Genes
FBN2
*IRX4
RPL5
RPS19
PRKAG2
COL2A1
AHNAK
VEGFA
*GATA4
*HAND1
ACTC1
MYH6

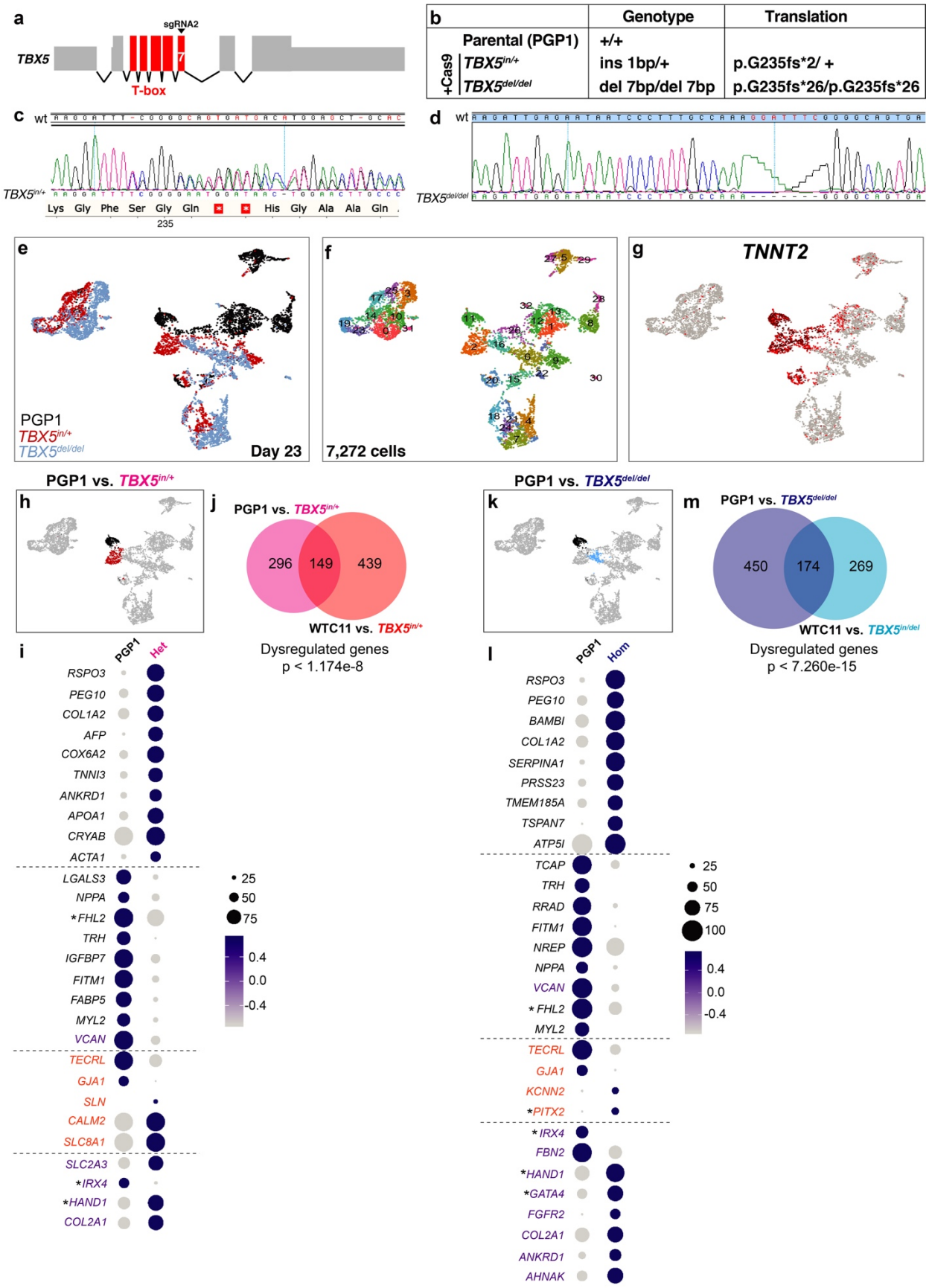
1104

1106 **Extended Data 9. Biological replicate of TBX5 dose-sensitive responses in WTC11-derived cardiomyocytes at day 23. a, b,** Cells at day 23 of an independent differentiation
1108 (biological replicate) were collected for single cell RNAseq analysis. UMAP shows all cells from day 23 colored by *TBX5* genotype or by cluster identity, respectively. **c,** UMAP shows *TNNT2*
1110 expression. **d, e,** UMAP of *TNNT2*$^{+}$ cells from day 23 are colored by *TBX5* genotype or by cluster identity, respectively. **f,** UMAP shows *TNNT2* expression in *TNNT2*$^{+}$ cells. **g, j,** UMAPs
1112 highlight clusters used in pair-wise differential gene tests. **h, k,** Dot plots show top differentially expressed genes between wildtype (WT)-enriched and *TBX5*$^{in/+}$-enriched or *TBX5*$^{in/del}$-enriched
1114 clusters (Supplementary Table 2). Electrophysiology-related (orange) or congenital heart disease (CHD)-associated (purple) genes are indicated. Asterisks denote transcription factors.
1116 Significance was determined by Wilcoxon Rank Sum test (adj p-value<0.05). **i, l,** Venn diagrams show overlap of differentially expressed genes detected in the first and second
1118 biological replicates at day 23. Statistical significance was determined by Hypergeometric test, based on the union of highly variable genes from both datasets (1,688) as the population size.
1120


a
WTC11 vs. TBX5in/+ + TBX5in/del Rep 1
61
22
181
WTC11 vs. TBX5in/+ + TBX5in/del Rep 2
p < 1.641e-04
WTC11 vs. TBX5in/+ + TBX5in/del Rep 1
b
WTC11
TBX5in/+
TBX5in/del
WTC11 vs. TBX5in/+ + TBX5in/del Rep 2
d
WTC11
TBX5in/+
TBX5in/del
c
WTC11
Het
Hom
RSPO3
PEG10
DES
VCAM1
MGLL
SLIT3
TNFRSF19
KRT18
*ID3
*YBX3
SFRP1
XRCC4
HSP90AB1
FABP5
HCN4
RHOQ
TRH
TECRL
IGFBP5
VCAN
e
WTC11
Het
Hom
RSPO3
PEG10
DES
VCAM1
MGLL
SLIT3
TNFRSF19
KRT18
*ID3
*YBX3
SFRP1
XRCC4
HSP90AB1
FABP5
HCN4
RHOQ
TRH
TECRL
IGFBP5
VCAN
100 75 50 25
100 80 40 20 0
-1 -0.5 0 0.5 1

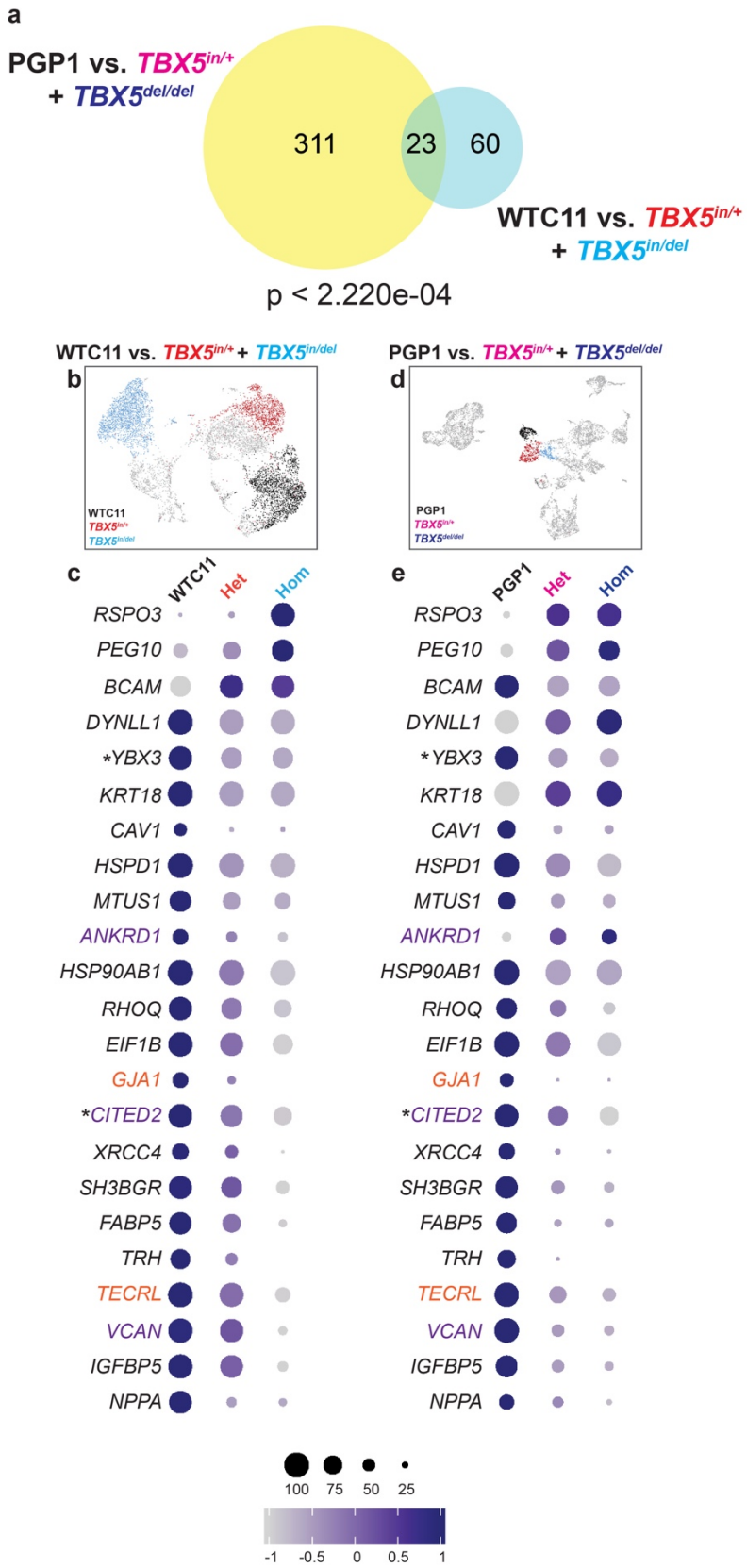
1122

Extended Data 10. Overlap of TBX5 dose-dependent genes between biological replicates.
1124 **a,** Venn diagram shows overlap of differentially expressed genes detected at day 23 in WTC11-
targeted *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells from biological replicate 1 and replicate 2. Statistical
1126 significance was determined by Hypergeometric test, based on the union of highly variable
genes from both datasets (1,688) as the population size. **b, d,** UMAPs demonstrate clusters
1128 chosen for differential gene expression within each experiment for comparison. **c, e,** Dot plots
show top differentially expressed genes for replicates. Significance was determined by Wilcoxon
1130 Rank Sum test (adj p-value<0.05). **f,** Venn diagram shows overlap of differentially expressed
genes between genetic backgrounds of WTC11-targeted *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells and PGP1-
1132 targeted *TBX5*$^{in/+}$ and *TBX5*$^{del/del}$ cells. **g, i,** UMAPs show clusters chosen for differential gene
expression for comparison. **h, j,** Dot plots show top differentially expressed genes in each
1134 genetic background by *TBX5* genotype.


a
sgRNA2
TBX5
T-box
b
Genotype
Translation
Parental (PGP1)
+/+
+Cas9
TBX5in/+
ins 1bp/+
p.G235fs*2/ +
TBX5del/del
del 7bp/del 7bp
p.G235fs*26/p.G235fs*26
c
wt
TBX5in/+
Lys Gly Phe Ser Gly Gln His Gly Ala Ala Gln
235
d
wt
TBX5del/del
e
PGP1
TBX5in/+
TBX5del/del
Day 23
f
7,272 cells
g
TNNT2
PGP1 vs. TBX5in/+
h
j
PGP1 vs. TBX5in/+
296
149
439
WTC11 vs. TBX5in/+
Dysregulated genes
p < 1.174e-8
PGP1 vs. TBX5del/del
k
m
PGP1 vs. TBX5del/del
450
174
269
WTC11 vs. TBX5in/del
Dysregulated genes
p < 7.260e-15
i
PGP1
Het
RSPO3
PEG10
COL1A2
AFP
COX6A2
TNNI3
ANKRD1
APOA1
CRYAB
ACTA1
LGALS3
NPPA
*FHL2
TRH
IGFBP7
FITM1
FABP5
MYL2
VCAN
TECRL
GJA1
SLN
CALM2
SLC8A1
SLC2A3
*IRX4
*HAND1
COL2A1
25
50
75
0.4
0.0
-0.4
l
PGP1
Hom
RSPO3
PEG10
BAMBI
COL1A2
SERPINA1
PRSS23
TMEM185A
TSPAN7
ATP5I
TCAP
TRH
RRAD
FITM1
NREP
NPPA
VCAN
*FHL2
MYL2
TECRL
GJA1
KCNN2
*PITX2
*IRX4
FBN2
*HAND1
*GATA4
FGFR2
COL2A1
ANKRD1
AHNAK
25
50
75
100
0.4
0.0
-0.4

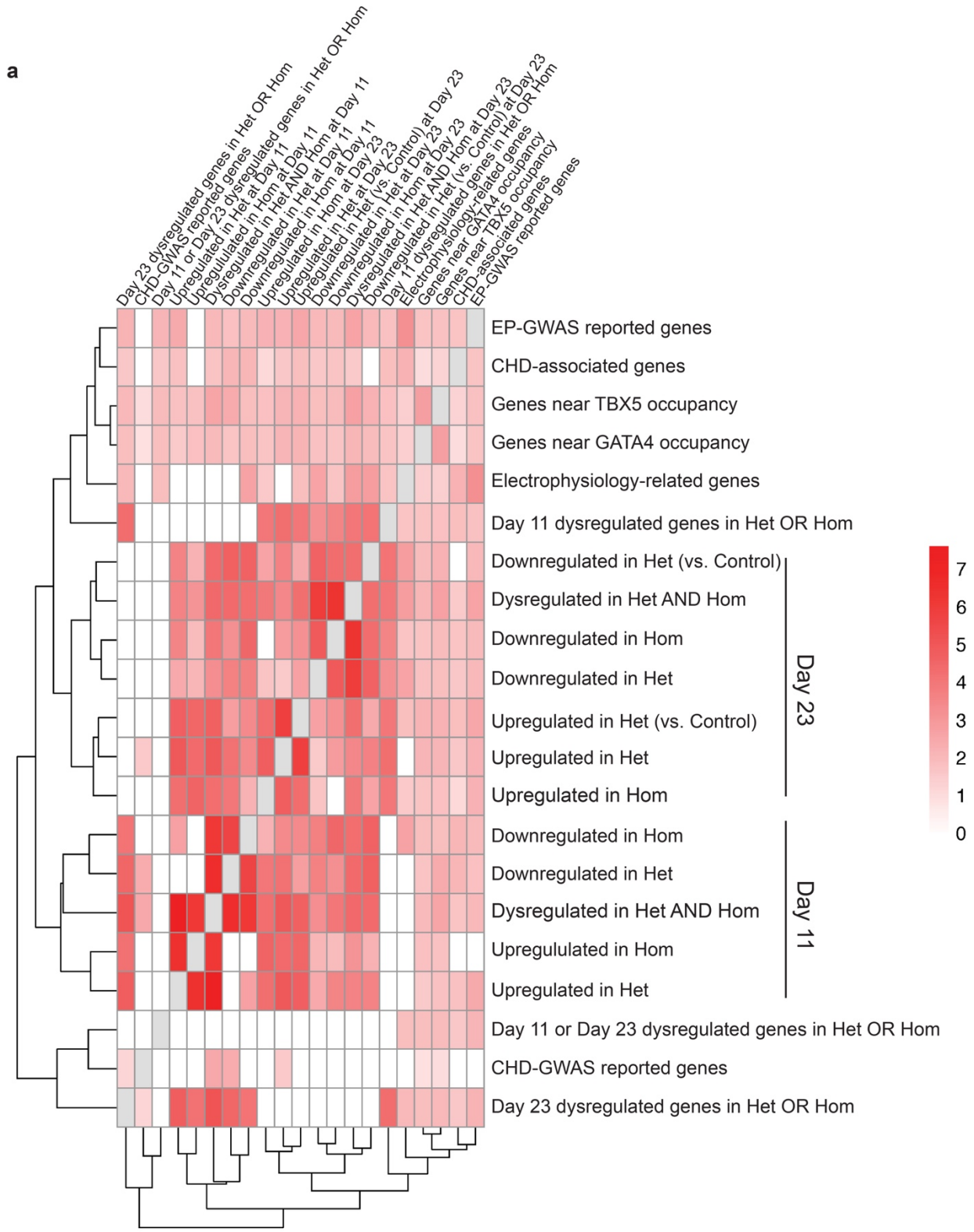
1136

1138 **Extended Data 11. Assessment of an alternative genetic background for TBX5 dose-sensitive gene expression. a,** Diagram of human *TBX5* gene is shown, with exons in red. The
1140 guide sgRNA2 was used to target exon 7, which encodes a portion of the T-box domain, of *TBX5* in PGP1 iPS cells. **b,** Table specifies the nature of each *TBX5* mutation and the predicted
1142 translation of TBX5 for PGP1-derived $TBX5^{in/+}$ or $TBX5^{del/del}$ cells. **c, d,** Chromatogram indicates a 1 base pair insertion of the mutant *TBX5* allele for PGP1-derived $TBX5^{in/+}$ cells (**c**) or a 7 base
1144 pair deletion of *TBX5* in $TBX5^{del/del}$ cells (**d**). **e-g,** UMAPs of all cells from day 23 by *TBX5* genotype (**e**), cluster identity (**f**), or by *TNNT2* expression (**g**). **h, k,** UMAPs mark *TBX5*
1146 genotype-enriched clusters by color that were selected for differential gene expression test. **i, l,** Dot plots show top differentially expressed genes in (**i**) $TBX5^{in/+}$- or (**l**) $TBX5^{del/del}$-enriched
1148 clusters. Significance was determined by Wilcoxon Rank Sum test (adj p-value<0.05) (Supplementary Table 2). **j**, Venn diagram shows overlap of TBX5-dependent gene sets
1150 between differences of WTC11 and WTC11-derived $TBX5^{in/+}$ clusters and differences of PGP1 and PGP1-derived $TBX5^{in/+}$ clusters. **m,** Venn diagram displays overlap of TBX5-dependent
1152 gene sets between differences of WTC11 and WTC11-derived $TBX5^{in/del}$ and differences of PGP1 and PGP1-derived $TBX5^{del/del}$. Statistical significance was determined by Hypergeometric
1154 test, based on the union of highly variable genes from both datasets (2,566) as the population size.
1156

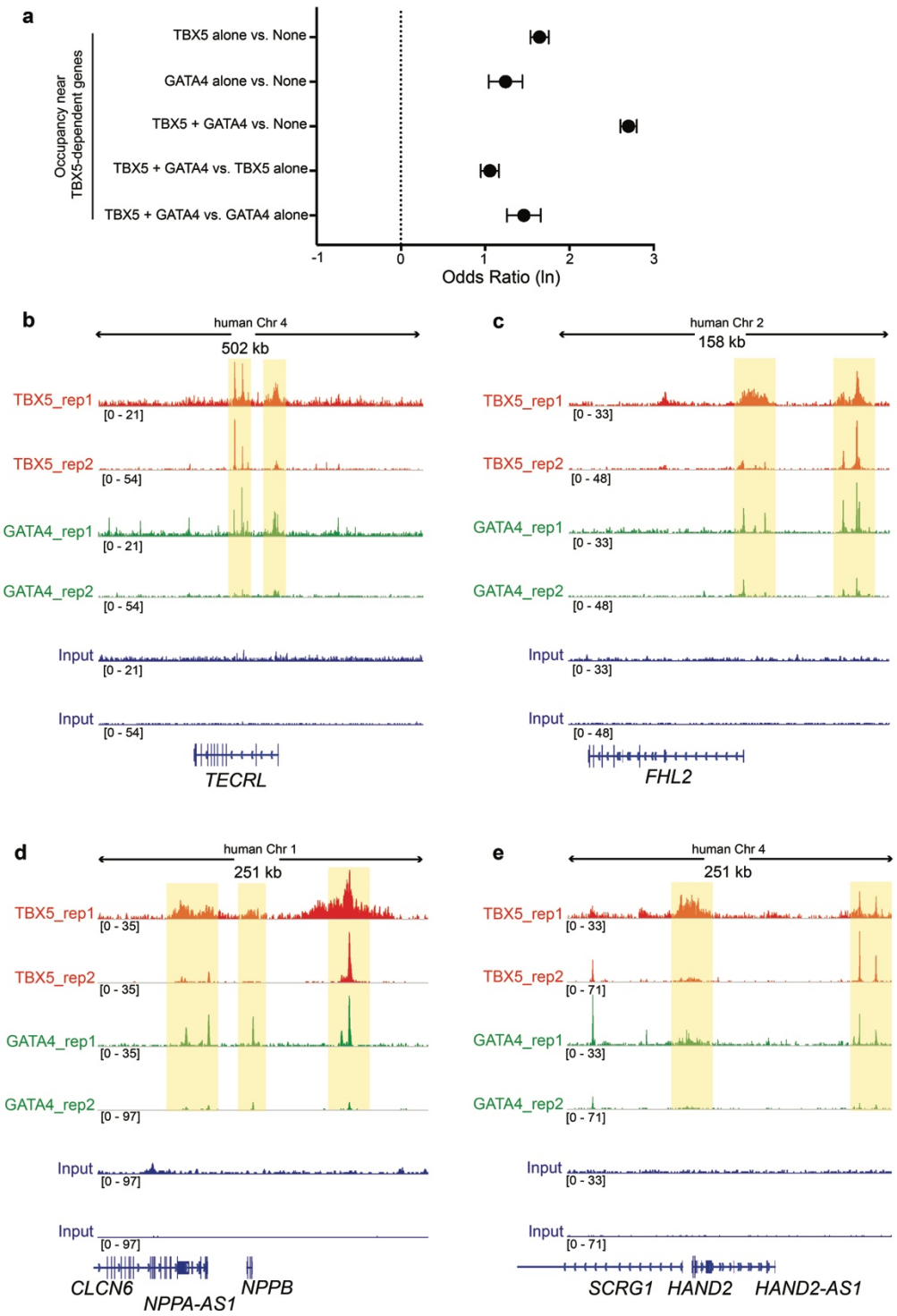

a
PGP1 vs. TBX5in/+ + TBX5del/del
311
23
60
WTC11 vs. TBX5in/+ + TBX5in/del
p < 2.220e-04
WTC11 vs. TBX5in/+ + TBX5in/del
b
WTC11
TBX5in/+
TBX5in/del
PGP1 vs. TBX5in/+ + TBX5del/del
d
PGP1
TBX5in/+
TBX5del/del
c
WTC11
Het
Hom
RSPO3
PEG10
BCAM
DYNLL1
*YBX3
KRT18
CAV1
HSPD1
MTUS1
ANKRD1
HSP90AB1
RHOQ
EIF1B
GJA1
*CITED2
XRCC4
SH3BGR
FABP5
TRH
TECRL
VCAN
IGFBP5
NPPA
e
PGP1
Het
Hom
RSPO3
PEG10
BCAM
DYNLL1
*YBX3
KRT18
CAV1
HSPD1
MTUS1
ANKRD1
HSP90AB1
RHOQ
EIF1B
GJA1
*CITED2
XRCC4
SH3BGR
FABP5
TRH
TECRL
VCAN
IGFBP5
NPPA
100
75
50
25
-1
-0.5
0
0.5
1

1158

1160 **Extended Data 12. Overlap of TBX5 dose-dependent genes between genetic backgrounds. a,** Venn diagram shows overlap of differentially expressed genes between
1162 genetic backgrounds of WTC11-targeted *TBX5*$^{in/+}$ and *TBX5*$^{in/del}$ cells and PGP1-targeted *TBX5*$^{in/+}$ and *TBX5*$^{del/del}$ cells. Statistical significance was determined by Hypergeometric test,
1164 based on the union of highly variable genes from both datasets (2,566) as the population size. **b, d,** UMAPs show clusters chosen for differential gene expression for comparison. **c, e,** Dot
1166 plots show top differentially expressed genes in each genetic background by *TBX5* genotype (adj p-value<0.05 by Wilcoxon Rank Sum test).
1168

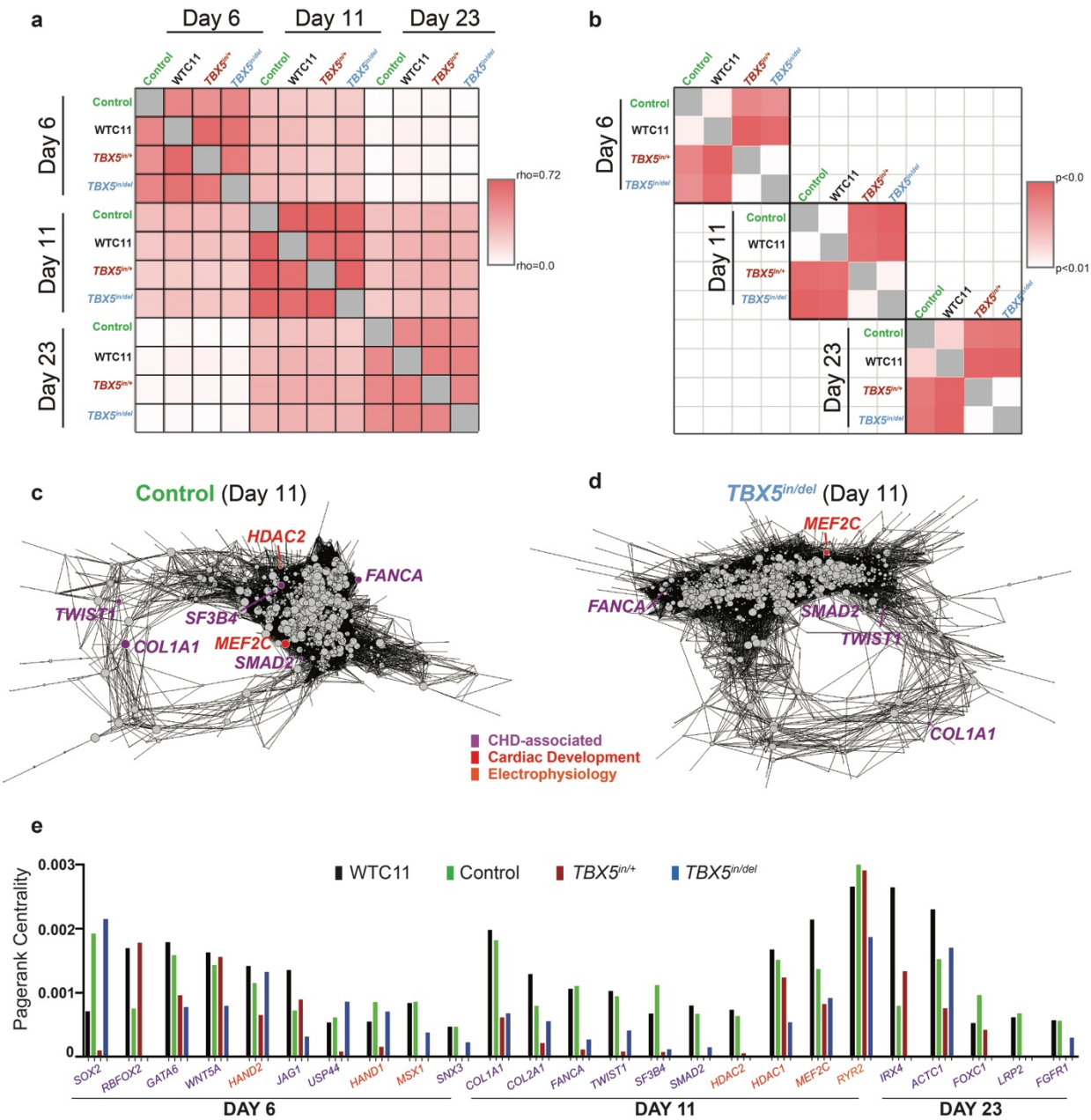

a
Day 23 dysregulated genes in Het OR Hom
CHD-GWAS reported genes
Day 11 or Day 23 dysregulated genes in Het OR Hom
Upregulated in Het at Day 11
Dysregulated in Hom at Day 11
Downregulated in Het AND Hom at Day 11
Upregulated in Hom at Day 11
Upregulated in Hom at Day 23
Downregulated in Hom at Day 23
Dysregulated in Hom at Day 23
Downregulated in Hom at Day 11
Upregulated in Hom at Day 11
Downregulated in Hom at Day 23
Upregulated in Hom at Day 23
Downregulated in Hom at Day 23
Dysregulated in Hom at Day 23
Downregulated in Het (vs. Control) at Day 23
Day 11 dysregulated genes in Hom at Day 23
Dysregulated in Hom at Day 23
Downregulated in Hom at Day 23
Dysregulated in Hom at Day 23
Downregulated in Hom at Day 11
Dysregulated in Hom at Day 11
Downregulated in Hom at Day 11
Upregulated in Hom at Day 11
Up

1174 **Extended Data 13. Associations between disease candidates, GWAS reported genes, TBX5 or GATA4 occupancy, or TBX5-dependent genes. a,** Heat map indicates odds ratios
1176 (FDR<0.05) between several variables, including CHD-associated genes, EP-related genes, CHD-related GWAS reported genes, arrhythmia-related GWAS (EP-GWAS) reported genes,
1178 TBX5 or GATA4 occupancy, and TBX5-dependent gene sets at day 11 or day 23 (Supplementary Table 4). Statistical significance was determined by Benjamini-Hochberg
1180 multiple testing.

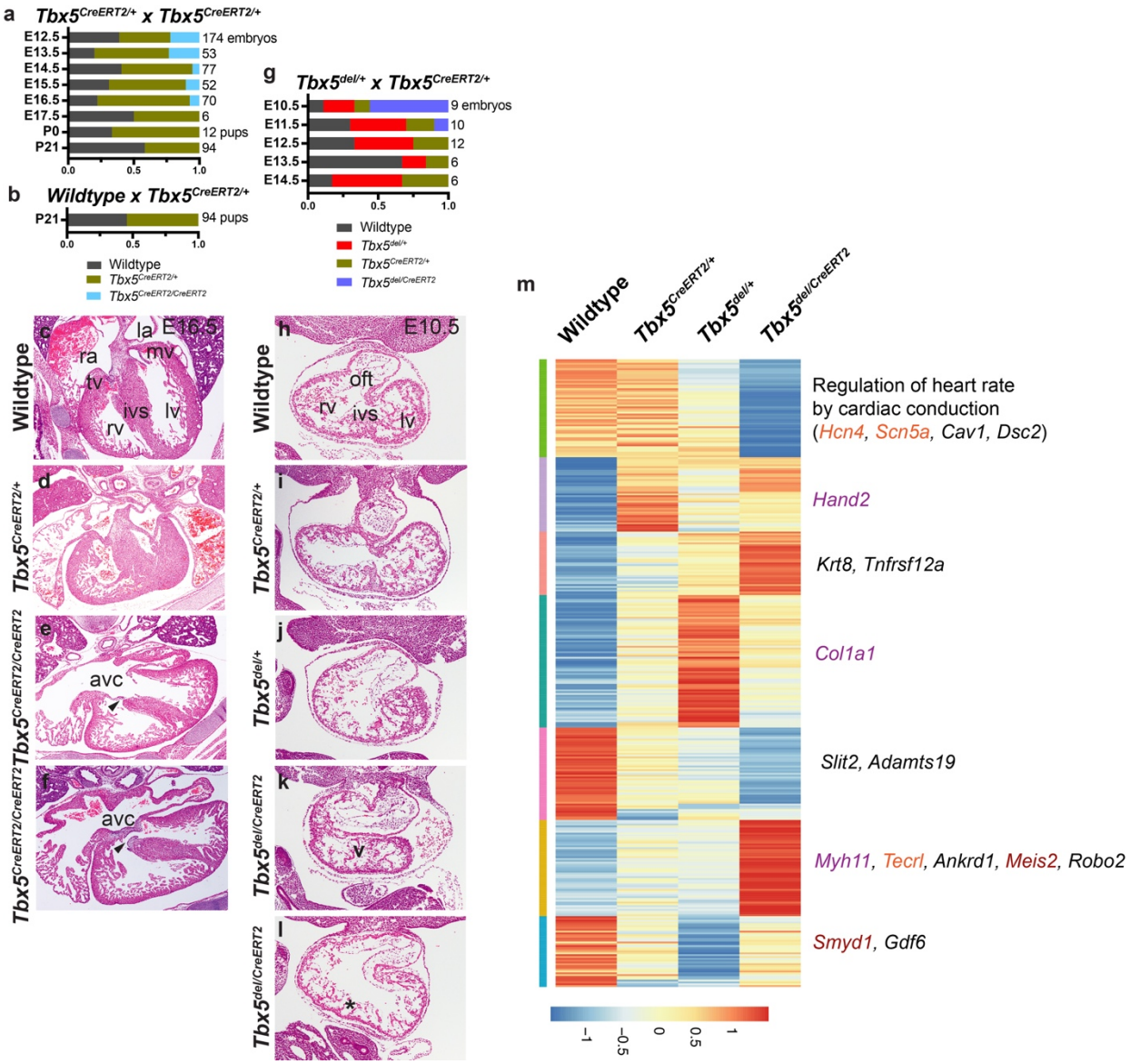

a
Occupancy near TBX5-dependent genes
TBX5 alone vs. None
GATA4 alone vs. None
TBX5 + GATA4 vs. None
TBX5 + GATA4 vs. TBX5 alone
TBX5 + GATA4 vs. GATA4 alone
-1
0
1
2
3
Odds Ratio (ln)
b
human Chr 4
502 kb
TBX5_rep1
[0 - 21]
TBX5_rep2
[0 - 54]
GATA4_rep1
[0 - 21]
GATA4_rep2
[0 - 54]
Input
[0 - 21]
Input
[0 - 54]
TECRL
c
human Chr 2
158 kb
TBX5_rep1
[0 - 33]
TBX5_rep2
[0 - 48]
GATA4_rep1
[0 - 33]
GATA4_rep2
[0 - 48]
Input
[0 - 33]
Input
[0 - 48]
FHL2
d
human Chr 1
251 kb
TBX5_rep1
[0 - 35]
TBX5_rep2
[0 - 35]
GATA4_rep1
[0 - 35]
GATA4_rep2
[0 - 97]
Input
[0 - 97]
Input
[0 - 97]
CLCN6
NPPA-AS1
NPPB
e
human Chr 4
251 kb
TBX5_rep1
[0 - 33]
TBX5_rep2
[0 - 71]
GATA4_rep1
[0 - 33]
GATA4_rep2
[0 - 71]
Input
[0 - 33]
Input
[0 - 71]
SCRG1
HAND2
HAND2-AS1

1182

Extended Data 14. Co-occupancy of TBX5 or GATA4 near human TBX5-dependent genes.
1184 **a,** Odds ratios (FDR<0.05) of TBX5, GATA4 or TBX5 and GATA4 occupancy near all human TBX5-dependent genes. Statistical significance was determined by Benjamini-Hochberg
1186 multiple testing (Supplementary Table 5). **b,** Browser tracks of TBX5 and GATA4 occupancy from iPS cell-derived cardiomyocytes are shown for loci of TBX5-dependent genes *TECRL,*
1188 *FHL2, NPPA/NPPB* or *HAND2.*



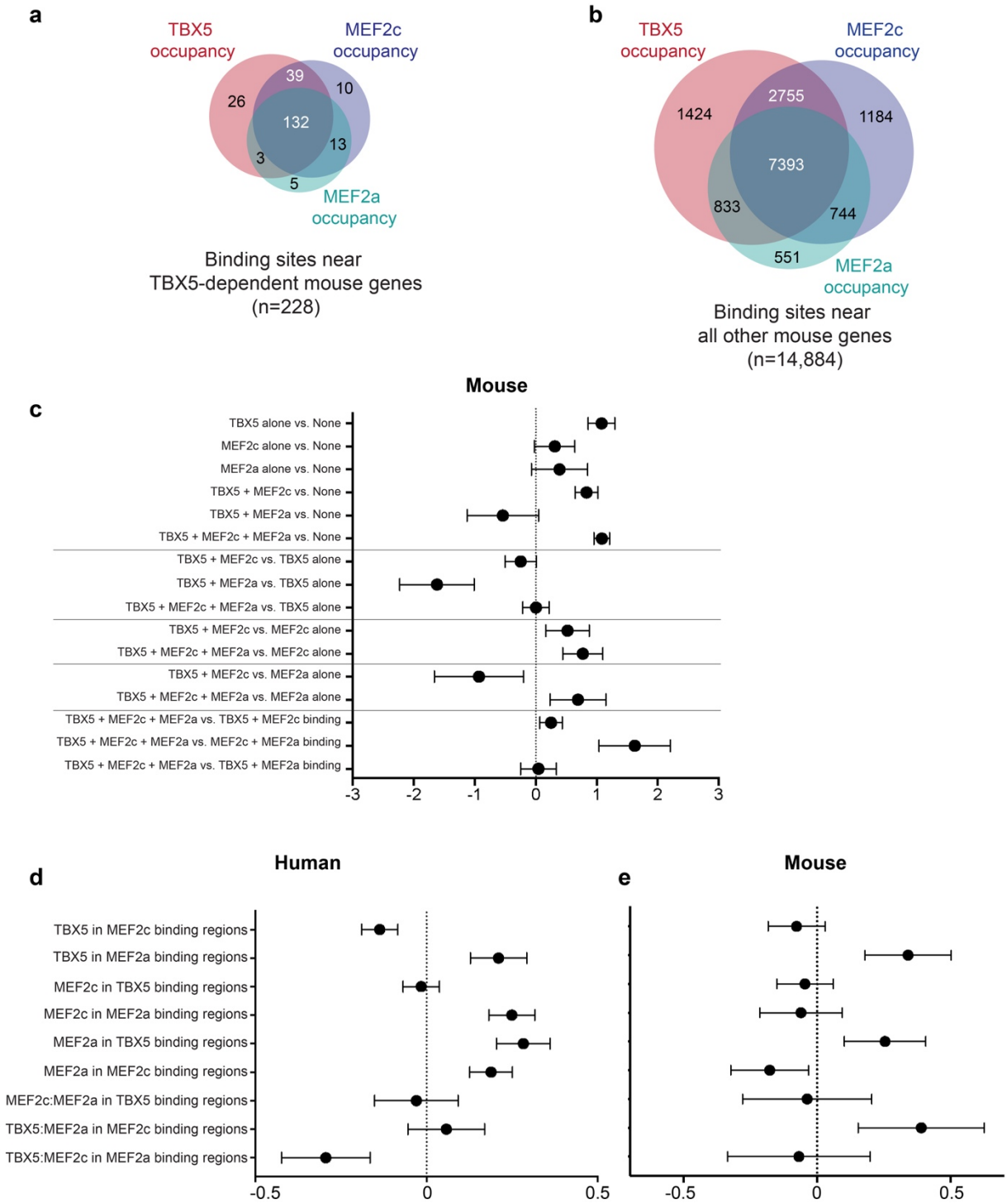
Extended Data 15. Analysis of TBX5 dosage-sensitive gene regulatory networks. **a,** Correlation plot (Pearson correlations of pagerank centralities) of networks by genotypes and time points are shown. Note that networks display highest similarity (red) within a time point. An inter-stage dissimilarity (white) grows proportionally to the time difference (i.e. Day 23 is less similar to Day 6 than Day 11). Therefore, comparisons for genotype differences were made within differentiation stages. **b,** Network similarity among genotypes within each time point is shown (Wilcoxon Rank Sum test of pagerank centralities for nodes from selected time point comparisons). **c, d,** Network diagrams of day 11 cardiomyocytes for control (**c**) or *TBX5*$^{in/del}$ (**d**) are shown. **e**, Quantification of pagerank centrality for significantly altered (top 5% cutoff) nodes of CHD-associated or heart development genes at specific time points are shown.



Extended Data 16. An allelic series of *Tbx5* mutants in mouse. a, Embryos or pups from matings of *Tbx5*$^{CreERT2/+}$ X *Tbx5*$^{CreERT2/+}$ were genotyped. Expected Mendelian ratios were not observed for *Tbx5*$^{CreERT2/CreERT2}$ after E13.5. **b,** Expected Mendelian ratios were observed for pups at weaning from matings of *Tbx5*$^{CreERT2/+}$ X wildtype. **c-f,** Transverse sections of hearts at embryonic day 16.5 (E16.5) from each genotype are shown. Right atrium (ra), left atrium (la), right ventricle (rv), left ventricle (lv), tricuspid valve (tv), mitral valve (mv) and interventricular septum (ivs) are labeled. Note atrioventricular canal defects (avc), which includes membranous VSDs (arrowheads), in *Tbx5*$^{CreERT2/CreERT2}$ hearts (**e, f**). **g,** Expected Mendelian ratios from matings of *Tbx5*$^{CreERT2/+}$ X *Tbx5*$^{del/+}$ were not observed for *Tbx5*$^{del/CreERT2}$ after E10.5. **h-l,** Transverse sections of individual hearts at embryonic E10.5 from each genotype are shown. Note a single ventricle (v) and lack of invagination (*) at the site of the nascent interventricular groove. **m,** Heat map by genotype of average gene expression data from 4 individual whole hearts at E10.5 (Supplementary Table 7). Significant GO term (adj p-value <0.05) for one cluster is shown, including underlying genes. TBX5-dependent mouse genes that are also

1220 TBX5-sensitive in human cells are shown. CHD-associated (purple), electrophysiology-related (orange), or heart development (red) genes are highlighted.

1222



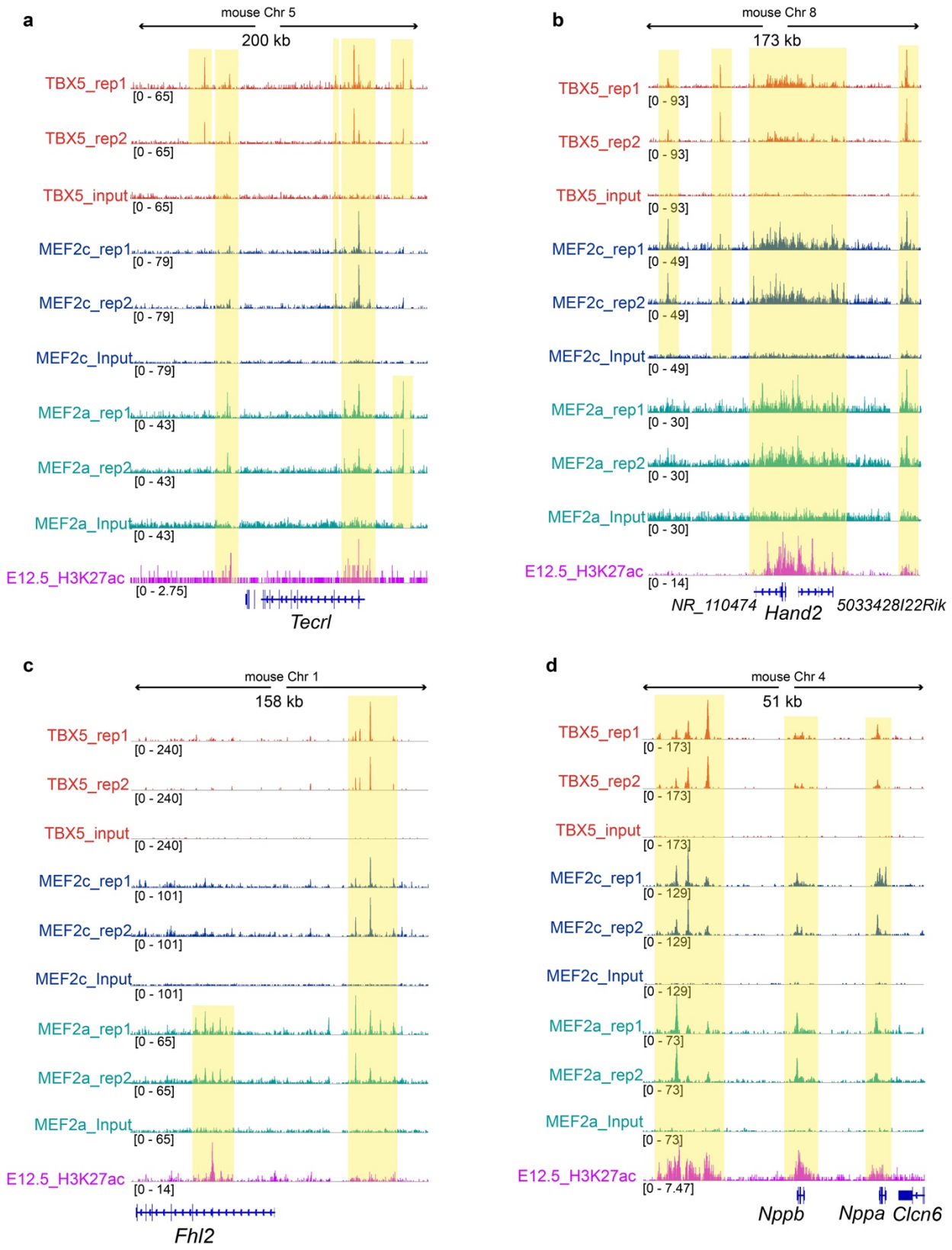
1224

1226

1228 **Extended Figure 17. Analyses of TBX5, MEF2c and MEF2a co-occupancy. a, b,** Venn diagrams display the overlap of TBX5, MEF2a or MEF2c occupancy near mouse orthologs of
1230 human TBX5-dependent or -independent genes, respectively. **c,** Odds ratios (FDR<0.05) of combinations of TBX5, MEF2a or MEF2c occupancy near TBX5-dependent mouse genes

1232 (Supplementary Table 5). **d, e,** Odds ratios (FDR<0.05) of combinations of MEF2a or MEF2c occupancy within 1kb of TBX5 binding sites near TBX5-dependent or all other genes in human
1234 (**d)** or mouse (**e**) (Supplementary Table 8).

1236


a
mouse Chr 5
200 kb
TBX5_rep1
[0 - 65]
TBX5_rep2
[0 - 65]
TBX5_input
[0 - 65]
MEF2c_rep1
[0 - 79]
MEF2c_rep2
[0 - 79]
MEF2c_Input
[0 - 79]
MEF2a_rep1
[0 - 43]
MEF2a_rep2
[0 - 43]
MEF2a_Input
[0 - 43]
E12.5_H3K27ac
[0 - 2.75]
Tecrl
b
mouse Chr 8
173 kb
TBX5_rep1
[0 - 93]
TBX5_rep2
[0 - 93]
TBX5_input
[0 - 93]
MEF2c_rep1
[0 - 49]
MEF2c_rep2
[0 - 49]
MEF2c_Input
[0 - 49]
MEF2a_rep1
[0 - 30]
MEF2a_rep2
[0 - 30]
MEF2a_Input
[0 - 30]
E12.5_H3K27ac
[0 - 14]
NR_110474
Hand2
5033428I22Rik
c
mouse Chr 1
158 kb
TBX5_rep1
[0 - 240]
TBX5_rep2
[0 - 240]
TBX5_input
[0 - 240]
MEF2c_rep1
[0 - 101]
MEF2c_rep2
[0 - 101]
MEF2c_Input
[0 - 101]
MEF2a_rep1
[0 - 65]
MEF2a_rep2
[0 - 65]
MEF2a_Input
[0 - 65]
E12.5_H3K27ac
[0 - 14]
Fhl2
d
mouse Chr 4
51 kb
TBX5_rep1
[0 - 173]
TBX5_rep2
[0 - 173]
TBX5_input
[0 - 173]
MEF2c_rep1
[0 - 129]
MEF2c_rep2
[0 - 129]
MEF2c_Input
[0 - 129]
MEF2a_rep1
[0 - 73]
MEF2a_rep2
[0 - 73]
MEF2a_Input
[0 - 73]
E12.5_H3K27ac
[0 - 7.47]
Nppb
Nppa
Clcn6

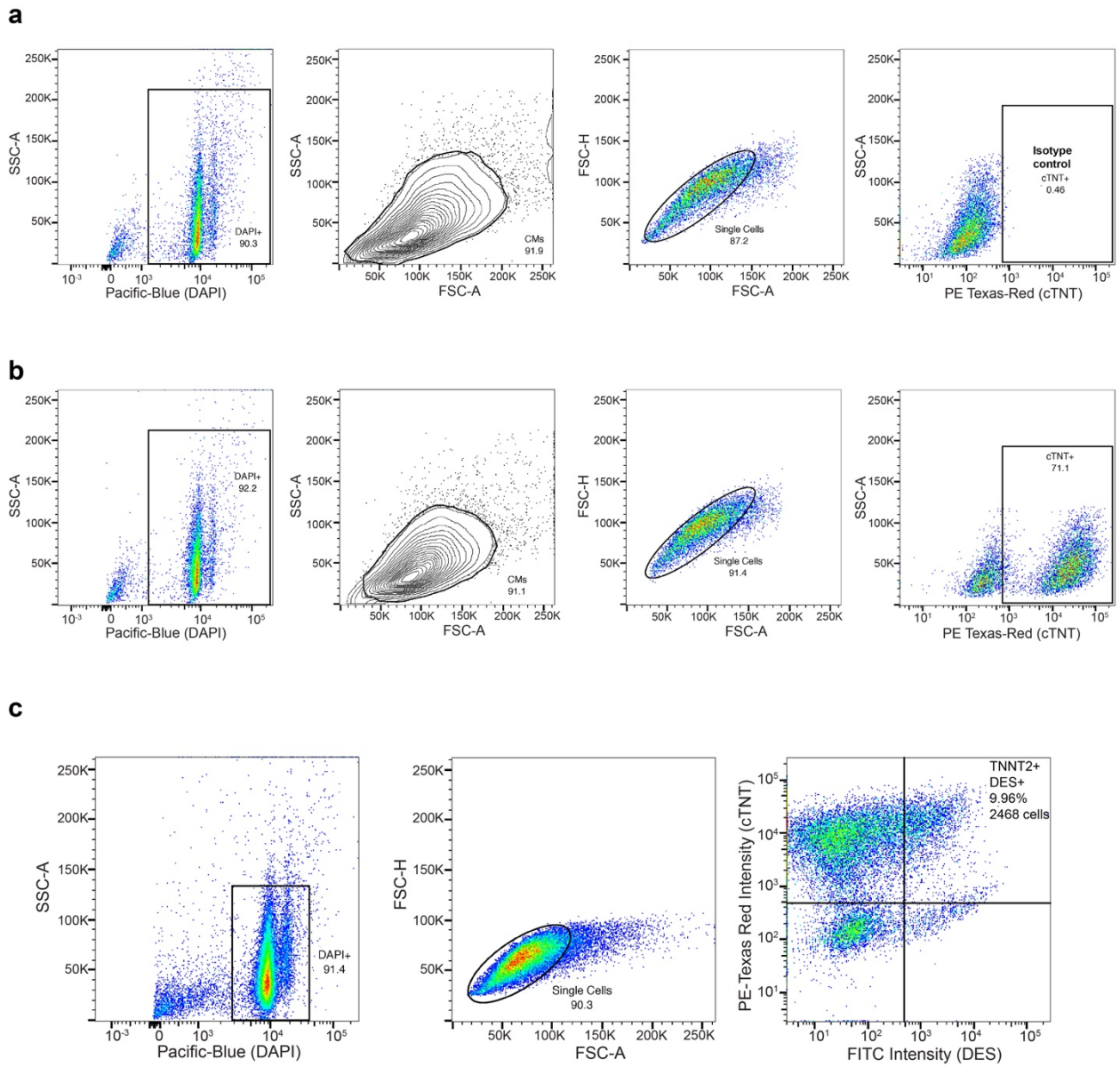
1238

Extended Data 18. Co-occupancy of TBX5, MEF2a and MEF2c near TBX5-dependent genes. a-d, Browser tracks for ChIP-seq data from E12.5 hearts of TBX5, MEF2c, MEF2a and H3K27ac near conserved TBX5-dependent genes, *Hand2*, *Tecrl, Fhl2* and *Nppa/Nppb*. Yellow bands of shading indicate co-occupancy that is present in replicates but not in input.

a
All genes downregulated in mouse
All genes upregulated in mouse
All mouse genes upregulated in TBX5CreERT2/del
Mouse genes upregulated in TBX5CreERT2/+ < TBX5del/+ < TBX5del/CreERT2
All mouse genes dysregulated in TBX5CreERT2/+
Mouse genes dysregulated in TBX5del/+
Human genes downregulated in Hom at Day 23
Human genes downregulated in Hom at Day 11
Human genes downregulated in Het (vs. Control) at Day 23
Human genes downregulated in Het AND Hom at Day 23
Human genes downregulated in Hom at Day 23
Human cardiac development-related genes
Human EP-related genes
Human CHD-associated genes
Tbx5 binding site within 100kb of two nearest genes
Mef2a binding site within 100kb of two nearest genes
Mef2c binding site within 100kb of two nearest genes
All human genes dysregulated at Day 23
Human genes downregulated at Days 11 or 23
Human genes upregulated at Days 11 or 23
Human genes upregulated in Hom at Day 23
Human genes upregulated in Het at Day 23
Human genes upregulated in Het (vs. Control) at Day 23
Human genes upregulated in Hom at Day 11
Human genes upregulated in Hom at Day 23
Human genes upregulated in Hom at Day 11
Human genes upregulated in Het at Day 11
Human genes upregulated in Hom at Day 23
Human genes upregulated in Hom

1252 or day 23 or TBX5-dependent mouse gene sets (Supplementary Table 6). Statistical significance was determined by Benjamini-Hochberg multiple testing.

1254



Extended Data 20. Gating strategies for flow cytometry. a, b, Gates were determined in each of the isotype controls per genotype and applied to genotype-matched samples. First, gates were drawn around DAPI$^+$ cells in SSC-A vs. Pacific-Blue plots. Next, SSC-A vs. FSC-A plots were used to exclude debris. Single cells were then isolated by gating of FSC-H vs. FSC-A plots. cTNT$^+$ gates were drawn in SSC-A vs. FITC plots. **c,** For cTNT$^+$/DES$^+$ cells, PE-Texas Red vs. FITC plots were used.

Supplementary Tables.

1272

Supplementary Table 1. Curated gene lists, which are used in this study, include
1274 electrophysiology (EP)-related genes, human congenital heart disease (CHD)-associated genes, mouse CHD-associated genes and cardiac development-related genes.

1276

Supplementary Table 2. Lists of differential genes from comparisons between *TNNT2*$^{+}$
1278 clusters, at day 11 or day 23, or by biological replicate or genetic background at day 23.

1280 **Supplementary Table 3.** TBX5-sensitive gene regulatory network analyses, by pagerank or degree, or by correlation with *TBX5* or *MEF2C* expression.

1282

Supplementary Table 4. Data for odds ratios for correlation of human TBX5-dependent genes
1284 near TBX5 or GATA4 occupancy, congenital heart disease (CHD)-associated GWAS, electrophysiology (EP)-related GWAS, CHD-associated genes, or EP-related genes.

1286

Supplementary Table 5. Odds ratios for co-occupancy data near human TBX5-dependent
1288 genes for TBX5 and GATA4, or for co-occupancy of TBX5, MEF2c and MEF2a near human or mouse TBX5-dependent genes.

1290

Supplementary Table 6. Data for odds ratios for correlation of human or mouse TBX5-
1292 dependent genes near TBX5, MEF2c or MEF2a occupancy, congenital heart disease (CHD)-associated GWAS, electrophysiology (EP)-related GWAS, CHD-associated genes, or EP-
1294 related genes.

1296 **Supplementary Table 7.** Bulk RNAseq data of mouse embryonic hearts at E10.5 for an allelic series of *TBX5* mutants.

1298

Supplementary Table 8. Odds ratio for transcription factor (TF) binding of TBX5, MEF2c, or
1300 MEF2a within 1kb of the other TFs in the trio, near human or mouse TBX5-depenedent genes.

1302

1304